The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

This electronic document was made available from www.rand.org as a public service of the RAND Corporation.

Skip all front matter: Jump to Page 1

Support RAND

- Purchase this document
- Browse Reports & Bookstore
- Make a charitable contribution

For More Information

- Visit RAND at www.rand.org
- Explore the RAND National Defense Research Institute
- View document details

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND electronic documents to a non-RAND website is prohibited. RAND electronic documents are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see RAND Permissions.
This report is part of the RAND Corporation tool series. RAND tools may include models, databases, calculators, computer code, GIS mapping tools, practitioner guidelines, web applications, and various toolkits. All RAND tools undergo rigorous peer review to ensure both high data standards and appropriate methodology in keeping with RAND’s commitment to quality and objectivity.
RAND
Suicide Prevention Program Evaluation
TOOLKIT

Joie D. Acosta, Rajeev Ramchand,
Amariah Becker, Alexandria Felton,
Aaron Kofner
The research described in this report was prepared for the Office of the Secretary of Defense (OSD). The research was conducted within the RAND National Defense Research Institute, a federally funded research and development center sponsored by OSD, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community under Contract W74V8H-06-C-0002.

Library of Congress Cataloging-in-Publication Data is available for this publication.

ISBN: 978-0-8330-8072-1

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

Cover images: iStockphoto/Thinkstock

© Copyright 2013 RAND Corporation

Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Copies may not be duplicated for commercial purposes. Unauthorized posting of RAND documents to a non-RAND website is prohibited. RAND documents are protected under copyright law. For information on reprint and linking permissions, please visit the RAND permissions page (http://www.rand.org/publications/permissions.html).

Published 2013 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org
Preface

Evaluations are critical for assessing the impact of U.S. Department of Defense (DoD) investments in suicide prevention and can be used as the basis for decisions about whether to sustain or scale up existing efforts. The Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury asked the RAND Corporation to draw from the scientific literature and create a toolkit to guide future evaluations of DoD-sponsored suicide prevention programs (SPPs). The overall goal of the toolkit is to help those responsible for SPPs determine whether their programs produce beneficial effects and, ultimately, to guide the responsible allocation of scarce resources.

The toolkit is accompanied by a companion document (available at http://www.rand.org/pubs/research_reports/RR283.html) that provides additional background for those who are interested in learning about the toolkit’s development. To develop the toolkit, RAND researchers examined the peer-reviewed literature and other evaluation toolkits and elicited feedback from program staff responsible for implementing DoD-sponsored SPPs.

The contents of this toolkit will be of particular interest to DoD SPP managers and suicide prevention coordinators and should also be useful to policymakers in other sectors who sponsor or manage SPPs more generally.

This research was sponsored by the Office of the Assistant Secretary of Defense for Health Affairs and the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury and conducted within the Forces and Resources Policy Center of the National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community.

For more information on the Forces and Resources Policy Center, see http://www.rand.org/nsrd/ndri/centers/frp.html or contact the director (contact information is on the web page).
## Contents

Preface ................................................................. iii  
Tools ........................................................................ vii  
Figures ...................................................................... ix

### CHAPTER ONE

**Introduction and Overview** ........................................ 1

- Intended Audience ................................................. 1  
- Toolkit Goals and Specific Aims ................................. 1  
- Evaluation Challenges Addressed ............................... 2  
- How the Toolkit Was Developed ............................... 2  
- User’s Guide ....................................................... 4  
- Start Using the Toolkit ......................................... 5  
- Summary ............................................................ 7

### CHAPTER TWO

**Identify Your Program’s Core Components and Build a Logic Model** ........................................... 9

- Specify Core Components ..................................... 9  
- Use Core Components to Build a Logic Model .......... 38  
- Assess the Quality of Your Program’s Logic Model .... 41  
- Summary ............................................................ 46

### CHAPTER THREE

**Design an Evaluation for Your Program** ................................................................. 47

- Identify Your Program Type ................................... 47  
- Learn How Similar Programs Have Been Evaluated .... 50  
- Learn the Types of Evaluation Designs ..................... 51  
- Identify Issues Pertinent to Your Evaluation Design ... 52  
- Select an Evaluation Design .................................. 59  
- Assess the Quality of the Design ............................ 61  
- Summary ............................................................ 61
# Tools

## Worksheets

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Identifying Core Components</td>
<td>10</td>
</tr>
<tr>
<td>3.1. Issues to Consider for My Program</td>
<td>57</td>
</tr>
<tr>
<td>6.1. Assessing Participation in Your Program’s Evaluation</td>
<td>103</td>
</tr>
<tr>
<td>6.2. Review Program Outcomes, with Example</td>
<td>105</td>
</tr>
<tr>
<td>6.3. Review Program Outcomes</td>
<td>106</td>
</tr>
<tr>
<td>6.4. Program Improvement Plan</td>
<td>112</td>
</tr>
</tbody>
</table>

## Templates

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Blank Program Logic Model Template</td>
<td>39</td>
</tr>
<tr>
<td>2.2. Sample Program Logic Model</td>
<td>40</td>
</tr>
<tr>
<td>3.1. Evaluation Planner</td>
<td>60</td>
</tr>
</tbody>
</table>

## Checklists

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Is This Toolkit Right for My Program?</td>
<td>6</td>
</tr>
<tr>
<td>2.1. Is Your Logic Model Complete and Appropriately Detailed?</td>
<td>42</td>
</tr>
<tr>
<td>2.2. Are the Core Components of Your Logic Model Appropriately Aligned?</td>
<td>43</td>
</tr>
<tr>
<td>3.1. Which Program Type Best Describes Your SPP?</td>
<td>51</td>
</tr>
<tr>
<td>3.2. Does Your Evaluation Plan Consider the Following Factors?</td>
<td>61</td>
</tr>
<tr>
<td>4.1. To What Extent Do the Measures Selected Align with Your Program’s Target Population, Activities, and Outcomes?</td>
<td>74</td>
</tr>
<tr>
<td>6.1. What CQI Actions Are Needed to Improve the Program?</td>
<td>110</td>
</tr>
</tbody>
</table>
Tables

3.1. Program Types, by Response to Worksheet 2.1 ........................................ 50
3.2. Types of Evaluation Designs ..................................................................... 53
3.3. Statistically Significant Effects of Suicide Prevention Programs, by Program Type ...... 54
4.1. Sample Process Measures ........................................................................ 65
4.2. Sample Outcome Measures ...................................................................... 69
5.1. Types of Analyses Addressed in Each Data Analysis Primer ....................... 80
6.1. Results-Based Scenarios and Associated Strategies for Program Improvement ........ 109
Figures

5.1. Sample Attendance Sheet with Unique Identifiers ........................................... 77
5.2. Survey Cover Sheet with Unique Identifier .................................................. 77
5.3. Formatting Ranges in Excel ............................................................................ 78
In this chapter, we review the intended audience and this toolkit’s goals and specific aims, as well as identify the evaluation challenges that this toolkit intends to address. We conclude with a brief user’s guide that previews the toolkit’s content and offers tips for its use and navigation.

**Intended Audience**

This toolkit is designed primarily to be used by individuals who are implementing suicide prevention programs (SPPs), including U.S. Department of Defense SPP coordinators and Veterans Health Administration suicide prevention program directors. Additionally, directors of community-based SPPs and state and local health department employees in charge of SPPs may find the toolkit useful. Individuals responsible for assessing or evaluating these SPPs, as well as those with an oversight role, may also be interested in the contents of this toolkit.

**Toolkit Goals and Specific Aims**

The overall goal of this toolkit is to help those responsible for SPPs detect whether their programs produce beneficial effects and, ultimately, to guide the responsible allocation of scarce resources. To accomplish this goal, this toolkit aims to

1. Inform you about the latest evaluation research from similar SPPs
2. Help you design an evaluation that is appropriate based on the type of program and available resources and expertise
3. Support the selection of measures for new evaluations and augment or enhance ongoing evaluations
4. Offer basic guidance on how to analyze evaluation data and then use these data to improve the effectiveness of SPPs.

Each chapter of the toolkit contributes to these specific aims, which can be achieved by working through the entire toolkit. Chapters Three and Four, respectively, summarize the latest evaluation research and offer guidance on specific evaluation designs and measures. To design an evaluation for your program, you will need to first identify your program’s core components (Chapter Two). You will then need to weigh the benefits and costs of specific evaluation designs so that you can select an evaluation design that is appropriate for your program’s resources and expertise (Chapter Three). Finally, you will select your evaluation measures (Chapter Four). Sample process and evaluation measures are included in Chapter Four. Chapters Five and Six, respectively, offer guidance on analyzing the evaluation data and using the data for program improvement.

Evaluation Challenges Addressed

Evaluating suicide prevention programs has been a challenge for many program staff. This toolkit will help address two common challenges to evaluating these types of programs.

Challenge 1: Suicide is a rare event, and data on suicides often lag by several years. In addition, because of low base rates, it can be difficult to identify meaningful program effects on suicide rates.

The toolkit provides options for evaluating intermediate outcomes associated with preventing suicides, as well as available suicide surveillance data.

Challenge 2: Many programs have multiple components, making it difficult to discern the components or characteristics that are responsible for any observed effect.

The toolkit supports the development of a program logic model that ties specific program activities to specific intermediate outcomes. Understanding which intermediate outcomes have been achieved can help program staff better understand the drivers of any changes in long-term outcomes, such as suicide rates.

How the Toolkit Was Developed

To create this toolkit, we conducted a systematic review of the program evaluation literature, including evaluation studies of SPPs, and existing evaluation toolkits. The literature review was used to identify the types of evaluation approaches and measures used elsewhere. The review of existing toolkits informed the outline of this toolkit’s components and our decisions about the types of tools that should be included (e.g., checklists, worksheets).
Chapter One: Welcome

Objective:
To briefly summarize how the toolkit was developed. For details, see the companion report, Development and Pilot Test of the RAND Suicide Prevention Program Evaluation Toolkit.

Literature Review
Between January and February 2012, we conducted a search of the peer-reviewed literature to identify evaluation approaches and process and outcome evaluation measures used in studies of SPPs. We did not limit our search to a specific range of years but, rather, searched all years of available data. For each evaluation study (n = 166), we abstracted two sets of information: evaluation data and measure data. Evaluation data included a description of each suicide prevention/reduction program, details about the evaluation design, and a synopsis of the study’s findings. Measure data described how study outcomes were assessed and included details about measure administration, scoring, and reliability. Each piece of abstracted information represents a characteristic or quality of a program or measure that we found useful when constructing this toolkit. We also identified articles that did not include such an evaluation but were still relevant to the development of this toolkit (e.g., discussions about evaluation methodology) and the objectives outlined in Chapters Two through Six.

Review of Existing Evaluation Toolkits
We reviewed existing evaluation toolkits to develop an outline for this toolkit. We identified several types of tools during the review, including sample measures, checklists with yes/no questions, worksheets with open-ended questions, and designs of prior evaluation studies. In developing this toolkit, we relied, in particular, on the Getting To Outcomes® (GTO) approach because GTO is the only evidenced-based model and intervention proven to increase programs’ ability to conduct self-evaluations (see Acosta and Chinman, 2011, and two forthcoming articles by Chinman et al.).

A detailed description of the methods can be found in the companion report, Development and Pilot Test of the RAND Suicide Prevention Program Evaluation Toolkit, available at http://www.rand.org/pubs/research_reports/RR283.html. A list of the literature reviewed can be found in the references section of this document.
User’s Guide

Overview of Content
The remainder of this toolkit walks the user through a series of worksheets and checklists. These worksheets and checklists will help you create a detailed and complete logic model that serves as the foundation for the program evaluation design (Chapter Two) and a plan for your program evaluation (Chapter Three). The evaluation plan will also specify process and outcome measures that can be used to evaluate your program (Chapter Four).

Once you have collected evaluation data, the toolkit will take you through a series of basic analyses of your data (Chapter Five) and help you interpret the findings so that they can be used for continuous quality improvement (Chapter Six).

A detailed description of how the toolkit was developed is available in the companion report.

Potential Benefits
This toolkit will help you achieve the following objectives:

- Develop a complete and detailed logic model that summarizes your program.
- Focus on available resources and important needs in your community.
- Use current research to augment and enhance your process and outcome evaluation plans.
- Collect and apply evaluation data to improve the implementation and the effectiveness of your program.

The toolkit is designed to guide users through a series of sequential steps in designing and implementing a program evaluation. Repeating the process on a regular basis will help program staff continually improve their SPP and contribute to the program’s ability to achieve its intended outcomes, which can ultimately result in fewer suicides in your community.
Tips for Navigating the Toolkit

This document contains several types of tools, which are marked with corresponding signposts. Worksheets are also available electronically at http://www.rand.org/pubs/tools/TL111.html.

- **Worksheets**
  - Ask you to answer key questions

- **Checklists**
  - Help direct you through the toolkit
  - Provide guidelines to review your own work

- **Templates**
  - Blank forms for use in planning
  - Complete planning forms to use as an example

- **Tables**
  - Summarize relevant research

Start Using the Toolkit

**Checklist 1.1** will help you decide whether this toolkit is appropriate for your SPP. If the toolkit is right for you, it is time to start using it! Be sure to use the toolkit sequentially. Worksheets are intended to inform other worksheets in later sections of the toolkit. Starting in the middle of the toolkit (e.g., in **Chapter Four**) may require referring back to earlier chapters. Therefore, we strongly encourage users to go through the toolkit sequentially.
Checklist 1.1
Is This Toolkit Right for My Program?

1. Are you planning to implement an SPP or are you currently implementing an SPP?

☐ Yes — This toolkit is right for you!

☐ No — This toolkit is intended primarily for individuals who are currently implementing or planning to implement an SPP or who have selected a specific SPP to implement. Individuals evaluating an SPP or with oversight of an SPP may also be interested in the content. However, you must have selected an SPP to utilize this toolkit. If you are looking for an SPP, consider accessing a registry of evidence-based SPPs, such as the Substance Abuse and Mental Health Services Administration’s National Registry of Evidence-Based Programs and Practices (http://www.nrepp.samhsa.gov), your regional suicide prevention network, or the Suicide Prevention Resource Center’s Best Practices Registry (http://www.sprc.org/bpr/section-i-evidence-based-programs).

2. Are you interested in beginning an evaluation of your SPP or enhancing your existing evaluation efforts?

☐ Yes — This toolkit is right for you!

☐ No — Proceed to Question 3.

3. Are you looking for evaluation measures for your SPP?

☐ Yes — This toolkit is right for you! Evaluation measures are available in Chapter Four. However, we strongly recommend working through the toolkit sequentially.

☐ No — Proceed to Question 4.

4. Have you already collected evaluation data on your SPP that you need help analyzing?

☐ Yes — This toolkit is right for you! Guidance on evaluation data analysis is available in Chapter Five. However, we strongly recommend working through the toolkit sequentially.

☐ No — Proceed to Question 5.
Chapter One: Welcome

Checklist 1.1
Is This Toolkit Right for My Program?

5. Do you need help interpreting evaluation data that you have already collected or applying evaluation data to improve your program?

☐ Yes — This toolkit is right for you! Guidance on interpreting evaluation data is available in Chapter Six. However, we strongly recommend working through the toolkit sequentially.

☐ No — Proceed to Question 6.

6. Are you interested in developing a logic model for your SPP?¹

☐ Yes — This toolkit is right for you!

☐ No — It is likely that the types of information provided in here are not applicable to your SPP.

Summary

This chapter described the purpose and content in the toolkit in order to help you decide whether this toolkit is appropriate for use with your program and provided a brief summary of how the toolkit was developed. Now that you have read this chapter and completed Checklist 1.1, you should know whether this toolkit is right for you and your SPP. If it is, proceed to Chapter Two, which will help you identify the core components of your SPP and develop a logic model for your program that can be used to guide your evaluation design.

¹ A logic model is a graphical depiction of the rationale and expectations of a program (Leviton et al., 2010). A logic model clarifies the causal relationships among program resources, activities, and outcomes (McLaughlin and Jordan, 1999; Wholey, Hatry, and Newcomer, 2010).
Chapter Two

Identify Your Program’s Core Components and Build a Logic Model

In this chapter, we will help you identify the core components of your suicide prevention program by considering the available resources, program activities, intended audience, intended outcomes, methods of assessment, and community needs being filled. Then, we will help you organize these elements into a logic model to clearly visualize the relationships and dependencies between components. Finally, we will help you review the logic model, assessing whether it is complete and reasonable.

Specify Core Components

Reason for Assessing Core Components
To conduct an effective evaluation of your SPP, it is important to have a clear understanding of the components that make up the program, as well as the interplay between these elements.

How to Assess Core Components
Your SPP’s core components are the resources available to the program; the activities, target population, and intended outcomes of the program; any current evaluation activities being conducted; and the need being addressed by the program. Ensuring that these components are well specified will facilitate the development of your evaluation plan. Indicate the core components of your SPP in Worksheet 2.1.
Fill in responses to each of the questions below to help build your program’s logic model.

**Resources**

Resources are the investments or “inputs” for the program. The performance of program activities and the fulfillment of program goals depend on these resources. A successful program needs adequate resources to fill program needs, as well as prudent allocation of these resources to avoid wasting time and money. Below is a list of the most common resources.

1. Specify which of these resources are available to the program:

   ☐ Space

   Describe:

   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ Staff (e.g., expertise, availability)

Describe:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Materials or technologies (e.g., videos, DVD players, computers)

Describe:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ Funding
Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Other
Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Next, we’ll need to transform these descriptions into well-specified bullets that describe program resources. Below are examples of well-specified and poorly specified lists of resources.

<table>
<thead>
<tr>
<th>Poorly Specified Examples</th>
<th>Well-Specified Examples</th>
</tr>
</thead>
</table>
| • Hotline staff           | • Three psychologists on staff trained to deal with callers in crisis  
                          | • A large volunteer base from the local suicide prevention coalition that can be leveraged as needed |
| • Money                   | • $50,000 funding to cover 1.5 full-time-equivalent program staff |
| • Equipment needed to run the program | • Donated space to run the program at the local department of health  
                          | • Hotline equipment (phones, phone lines, answering service, etc.) |
| • Relationships with key partners | • A memorandum of understanding with the sheriff’s department to conduct house calls when hotline staff identify a caller as high-risk |

Describe these resources using a list of well-specified bullets:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Program Activities
Activities are the actions and efforts that make up the program and are employed to reach the program's goals. Activities include the tools, services, and products that the program provides to the target population. Some general categories of SPP activities are listed below.

2. Specify which of the following activities your program engages in:

☐ Operate a crisis hotline
Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Train individuals to cope with stress or self-refer if they are at risk for suicide
Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ Train individuals (e.g., peers, clergy) to recognize individuals at risk for suicide

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Train mental health providers to assess suicidal risk or treat patients at risk for suicide

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
### Worksheet 2.1
Identifying Core Components

- Screen individuals for suicide risk or mental health problems

Describe:

- Provide mental health treatment

Describe:
Worksheet 2.1
Identifying Core Components

☐ Enforce policies or procedures that restrict access to lethal means (e.g., guns, lethal medications) or create a safer physical environment for individuals at risk for suicide

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Respond appropriately in the aftermath of a suicide

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

- Provide information on signs and symptoms associated with suicide or resources available to prevent suicide
  
  Describe:

- Other
  
  Describe:
Next, we’ll need to transform these descriptions into well-specified bullets that describe program activities. Below are examples of well-specified and poorly specified program activities.

<table>
<thead>
<tr>
<th>Poorly Specified Examples</th>
<th>Well-Specified Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Train hotline operators</td>
<td>• Provide annual ASIST (Applied Suicide Intervention Skills Training) sessions to all hotline staff</td>
</tr>
<tr>
<td>• Operate a crisis hotline</td>
<td>• Operate a 24-hour hotline for individuals in the metropolitan areas of Baltimore, Maryland (zip codes 21201, 21202, 21203, 21205, 21206, 21209, and 21210)</td>
</tr>
<tr>
<td>• Screen callers for mental health problems</td>
<td>• Screen callers for mental health problems using the PHQ-9 (the nine-item Patient Health Questionnaire scale)</td>
</tr>
</tbody>
</table>

Describe these activities using a list of well-specified bullets:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
**Program’s Target Population**

The target population is the group intended to utilize the program. This may consist of the entire population of a given region or a restricted subgroup. Be as specific as possible when describing the program’s target population.

3. Specify which of these subgroup(s) the program targets:

- [ ] Age group (e.g., adolescents, elderly)
  
  Describe:
  
  ______________________________________________________________
  ______________________________________________________________
  ______________________________________________________________
  ______________________________________________________________
  ______________________________________________________________
  ______________________________________________________________

- [ ] Gender
  
  Describe:
  
  ______________________________________________________________
  ______________________________________________________________
  ______________________________________________________________
  ______________________________________________________________
  ______________________________________________________________
### Worksheet 2.1
Identifying Core Components

- **Profession (e.g., nurses, teachers, health care providers)**

  Describe:
  
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________

- **Region (e.g., residents of Los Angeles County)**

  Describe:
  
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________

- **Military/civilian**

  Describe:
  
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
  __________________________________________________________
People with risk factors (e.g., borderline personality disorder, depression)

Describe:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Other

Describe:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Next, we’ll need to transform these descriptions into a single well-specified bullet that describes the program’s target population. Below are examples of well-specified and poorly specified target populations.

<table>
<thead>
<tr>
<th>Poorly Specified Examples</th>
<th>Well-Specified Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>• High school students</td>
<td>• 14- to 18-year-old students in all Kansas City public schools</td>
</tr>
<tr>
<td>• Noncommissioned officers</td>
<td>• All noncommissioned officers stationed at Fort Hood, Texas</td>
</tr>
</tbody>
</table>
Worksheet 2.1
Identifying Core Components

Describe the target population using a single well-specified bullet. If there are multiple target populations, feel free to include multiple bullets:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Program Outputs

Program outputs are the amount, quality, or volume of goods or services provided by the program. These can include the number of individuals trained, number of individuals served, or the number of materials developed as part of the program (e.g., program manual). These outputs represent the quantity or volume of program activities.

4. What are the outputs provided by the program? Specify:

☐ Were any individuals trained as part of the program?

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ Were any individuals provided with services as part of the program?
Describe:

☐ Were any materials developed as part of the program?
Describe:

Next, we’ll need to transform these descriptions into well-specified bullets that describe program outputs. Below is an example of a well-specified and a poorly specified program output.

<table>
<thead>
<tr>
<th>Poorly Specified Example</th>
<th>Well-Specified Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Program served active-duty service members</td>
<td>• Program served 95 active-duty service members at high risk for suicide</td>
</tr>
</tbody>
</table>
Describe the program outputs using well-specified bullets:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

**Intended Outcomes**

Outcomes are changes in the target population expected as a result of engaging in the program activities. These outcomes may include changes in knowledge, attitude, skills, or behavior and should be directly related to the needs being filled. Effective outcomes follow the SMART method (Lesesne et al., 2011). That is, they should be

- **Specific**: Describe precisely what is expected to change and for whom.
- **Measurable**: There must be a way to determine the presence or extent of change.
- **Achievable**: Outcomes must be feasible for the target population (e.g., based on prior empirical expectations for change).
- **Realistic**: Outcomes should be able to be accomplished with the available resources.
- **Time-bound**: Describe the time frame in which the change is expected to occur.
Worksheet 2.1
Identifying Core Components

5. What are the intended outcomes of the program? Specify:

☐ What is expected to change (e.g., specific knowledge about the signs and symptoms associated with suicide, attitudes toward help seeking, help-seeking behaviors)?

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ What target population is expected to change (use the well-specified bullet from the “Target Population” section of this worksheet)?

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ In what time frame do you expect the change to occur (e.g., increase in knowledge may occur by the end of the final session of a three-session training program)?

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ How much do you expect the intended outcome to change?

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Outcomes can be grouped into short-term and long-term outcomes. “Short-term outcomes should be attainable within 1 to 3 years, while longer-term outcomes should be achievable within a 4 to 6 year timeframe” (W. K. Kellogg Foundation, 2000). However, if your program is only three months long, your short-term outcomes may occur in the one- to three-month time frame, and your long-term outcomes may occur in the six-month to one-year time frame. Anchor your short- and long-term outcomes to your program’s length. Next, we’ll need to transform these descriptions into well-specified bullets that describe intended program outcomes and group them as short-term and long-term outcomes. Below is an example of a well-specified and a poorly specified program outcome.

<table>
<thead>
<tr>
<th>Poorly Specified Example</th>
<th>Well-Specified Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Increase knowledge about suicide</td>
<td>• After Springfield High School students participate in the program for two months, their recognition of suicide risk factors will increase by 20 percent</td>
</tr>
</tbody>
</table>

Describe the intended short-term outcomes using well-specified bullets:
Worksheet 2.1
Identifying Core Components

Describe the intended long-term outcomes using well-specified bullets:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Current Evaluation Activities

You may have already thought about evaluating your program or begun collecting data to inform an evaluation of your program. These activities may help inform your evaluation design. If you are not currently conducting any evaluation activities, leave this section blank and skip to next section (“Needs Being Addressed by the Program”).

6. Summarize any activities you are conducting to inform an evaluation of the program:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Next, we’ll need to transform these descriptions into well-specified bullets that describe your program’s current evaluation activities. Below are examples of well-specified and poorly specified evaluation activities.

<table>
<thead>
<tr>
<th>Poorly Specified Example</th>
<th>Well-Specified Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Track suicidal ideations and attendance</td>
<td>• Assess changes in the frequency of suicidal ideation among program participants using the Suicide Ideation Questionnaire</td>
</tr>
<tr>
<td></td>
<td>• Track each participant’s attendance at the eight program sessions</td>
</tr>
</tbody>
</table>

Describe the activities using well-specified bullets:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Needs Being Addressed by the Program

Needs represent a problem or deficiency in the community that the program intends to remedy. The long-term goal of the program is to fill one or more of these needs. Though broad, these needs are more specific than the ultimate goal of preventing or reducing suicide. Basing needs on objective data (e.g., responding to increases in suicide rates), rather than perceptions or opinions, ensures that your SPP is addressing an actual need in your community.
Worksheet 2.1
Identifying Core Components

7. Specify which of these needs the program addresses:

☐ Increasing rates of suicide among the target population

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Lack of knowledge/awareness about the signs or symptoms of suicide and risk factors

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ Lack of knowledge/awareness about resources and help-seeking

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Detrimental attitudes about suicide

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ Detrimental attitudes toward mental health or help-seeking behaviors
Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Lack of competence in helping suicidal individuals (e.g., family, patients, friends, students).
*Competence* refers to an understanding of how to approach and talk with someone who might be suicidal.
Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ Lack of adequate care/support for at-risk individuals

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Barriers to accessing available resources

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 2.1
Identifying Core Components

☐ Inability to identify individuals at risk for suicide

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

☐ Ease of access to lethal means (e.g., guns, drugs)

Describe:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Next, we’ll need to transform these descriptions into well-specified bullets that describe the need being addressed by the program. Below are examples of well-specified and poorly specified needs.

<table>
<thead>
<tr>
<th>Poorly Specified Examples</th>
<th>Well-Specified Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>• High rates of suicide</td>
<td>• Suicide rates among ninth-grade students at Morris High School have doubled since 2008.</td>
</tr>
<tr>
<td>• Lack of knowledge about signs and symptoms of suicide</td>
<td>• The 2011 Morris High School Student Survey found that 90 percent of ninth-grade students did not know where to refer peers experiencing an emotional crisis or threatening suicide.</td>
</tr>
</tbody>
</table>
Worksheet 2.1
Identifying Core Components

Describe the needs being addressed using well-specified bullets:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Use Core Components to Build a Logic Model

Logic models facilitate the visualization of relationships between the core components of a program. Read from left to right, these models depict the logical flow of inputs (resources) to actions (activities) to outputs (outcomes) to assessment (evaluation) to goals (needs). An “if . . . then” relation connects adjacent components: If the resources are provided, then the activities are conducted; if the activities are conducted, then the targeted population can participate; if the targeted population participates, then the outcomes can occur.

Reason for Building a Logic Model

The organization of a logical model provides the following advantages:

- It simplifies and summarizes the program’s core components.
- It depicts the connection between concrete resources/activities and abstract goals.
- It allows programs to assess how well the plan aligns with the actual implementation.

Now that you have identified and specified the core components of the program, you can use the responses from Worksheet 2.1 to construct a program logic model using Template 2.1. The bullets in each summary section of Worksheet 2.1 can be transferred directly into the corresponding section of the logic model in Template 2.1. For example, if you indicated that one of the needs addressed by the program was “Suicide rates among ninth-grade students at Morris High School have doubled since 2008,” you would enter this text directly into the “Needs” box in the logic model. A sample completed logic model for a fictional program is shown in Template 2.2.
## Template 2.1
### Blank Program Logic Model Template

<table>
<thead>
<tr>
<th>Resources</th>
<th>Activities</th>
<th>Outputs</th>
<th>Outcomes</th>
<th>Evaluation</th>
<th>Needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>The capacities and resources available to the program</td>
<td>The program activities conducted with the resources</td>
<td>The direct products of the program activities</td>
<td>What your program expects to change as a result of the activities</td>
<td>How your program addresses outcomes</td>
<td>Needs in the community that the outcomes address</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity 1:</th>
<th>Outcome 1:</th>
<th>Short Term:</th>
<th>Measure 1:</th>
<th>Need 1:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Long Term:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 2:</td>
<td>Outcome 2:</td>
<td>Short Term:</td>
<td>Measure 2:</td>
<td>Need 2:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long Term:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 3:</td>
<td>Outcome 3:</td>
<td>Short Term:</td>
<td>Measure 3:</td>
<td>Need 3:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long Term:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 4:</td>
<td>Outcome 4:</td>
<td>Short Term:</td>
<td>Measure 4:</td>
<td>Need 4:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long Term:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 5:</td>
<td>Outcome 5:</td>
<td>Short Term:</td>
<td>Measure 5:</td>
<td>Need 5:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long Term:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Template 2.2
#### Sample Program Logic Model

<table>
<thead>
<tr>
<th>Resources</th>
<th>Activities</th>
<th>Outputs</th>
<th>Outcomes</th>
<th>Evaluation</th>
<th>Needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>The capacities and resources available to the program</td>
<td>The program activities conducted with the resources</td>
<td>The direct products of the program activities</td>
<td>What your program expects to change as a result of the activities</td>
<td>How your program addresses outcomes</td>
<td>Needs in the community that the outcomes address</td>
</tr>
<tr>
<td>Three psychologists on staff trained annually in ASIST</td>
<td>Operate a 24-hour hotline for individuals in Baltimore</td>
<td>616 in-crisis callers served in the past year</td>
<td>Short term: None</td>
<td>Assess changes in suicide rates before and after hotline initiation using hospital data</td>
<td>Mayor's Task Force Report indicated Baltimore has high rates of suicide compared with U.S. average</td>
</tr>
<tr>
<td>A large volunteer base from the local suicide prevention coalition that can be leveraged as needed</td>
<td>Screen callers for mental health problems using the PHQ-9</td>
<td>550 in-crisis callers screened for mental health problems in the past year</td>
<td>Short term: Increase the identification of callers at risk for suicide by 15% in the first 6 months</td>
<td>Assess changes in the number of callers screened for risk of suicide</td>
<td>An assessment conducted by the state health department identified a lack of support for in-crisis Baltimore residents</td>
</tr>
<tr>
<td>Enough funding to run the program for one year ($50,000)</td>
<td></td>
<td></td>
<td>Long term: Decrease in suicide rate in Baltimore by 20% over the first 3 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donated space to run the program at the local department of health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hotline equipment (phones, phone lines, answering service, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An MOU with the sheriff’s department to conduct house calls when hotline staff identify a caller as high-risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Program Name:** Fictional Crisis Hotline  
**Date:** 6/1/13  
**Target Population:** In-crisis residents in the metropolitan areas of Baltimore, Maryland (zip codes 21201, 21202, and 21210)
Assess the Quality of Your Program’s Logic Model

Reasons to Assess Quality
An effective logic model provides an accurate and comprehensive representation of the program. To be effective, a logic model must be complete and detailed. Because the logic model serves as the foundation for your program’s evaluation, it is essential that you take the time to develop a complete and detailed logic model. If a program cannot articulate its core components in a logic model, it may be an indicator that the program’s activities are not well specified or concrete enough to merit an evaluation at this time.

In addition, the relationships between the program’s core components must be reasonably clear: If the target population participates, then the outputs will be completed; if the outputs are completed, then the outcomes could be possible; if the outcomes are possible, then they can be evaluated through assessment; if the evaluation is completed, then we will know whether the needs of the community were successfully addressed. To ensure that your program’s evaluation is reasonable and appropriate, you must ensure that the core components of your logic model align with one another.

How to Assess Quality
Use the following checklists to review the logic model you created using Template 2.1. Check off each item that is true for your logic model (W. K. Kellogg Foundation, 2000).
Checklist 2.1
Is Your Logic Model Complete and Appropriately Detailed?

☐ All significant resources contributing to the program are listed.¹

☐ All significant activities contributing to the program are listed.²

☐ All people intended to use the program are represented in the listed target population(s).

☐ The listed target population specifies the relevant age group, gender, rank, etc.

☐ The listed target population provides enough detail that it is clear which subgroups are intentionally excluded from participation in the program.

☐ Outcomes are SMART (specific, measurable, achievable, reasonable, and time-bound).

☐ All significant evaluation activities used to assess the program are listed.

☐ Any significant community needs addressed by the program are listed.

If you did not check off all of these items, revisit the corresponding section in Template 2.1 to further specify your program’s core components. For example, if you did not check off the first item, “All significant resources contributing to the program are listed,” revisit the “Resources” section of Worksheet 2.1.

¹ Significant resources are the resources (e.g., money, staff, equipment and supplies) that are essential to and contribute the majority of support to the program activities. Depending on the size of your program, use your best judgment to identify these resources.

² Significant activities are the defining activities that make up the essence or core of your program.
Chapter Two: Core Components and Logic Model

Checklist 2.2
Are the Core Components of Your Logic Model Appropriately Aligned?

1. **Resources are adequate and suitable to conduct program activities.**
   - Yes → This means that you have appropriate and adequate staff, space, and materials to fully implement your program. For example, if you need a trained mental health professional for your program but do not have access to one, this means that you do not have adequate resources.
   - No → Without adequate resources, your program may not have the desired outcomes. Before implementing your program, you need to secure all necessary resources.

2. **The program activities involve the targeted population.**
   - Yes → This means your program activities are reaching the target population. For example, if your program is targeting service members at high risk for suicide, you will need a way to identify and recruit those individuals to your program.
   - No → If your program activities do not reach your target population, your program will not have the desired outcomes. You may need to reconsider whether you have identified the appropriate target population or augment your program with additional activities that reach the intended population.

3. **Outputs correspond to the program activities listed.**
   - Yes → This means that you have captured the level of effort from each of your program activities. Each of your activities should have an output. All outputs should be tied to at least one or more program activities.
   - No → If you have an output listed that is not linked to a program activity, you may be missing a program activity. If you have a program activity without an output, add an output for that activity. All program activities should have at least one output, and, similarly, all outputs should be tied to one or more program activities.
4. **Outcomes correspond to program activities listed.**

☐ Yes → This means that you have identified which of the activities contribute to each of your program outcomes. You do not want to overreach in identifying outcomes that are grander or more comprehensive than the program activities. Ensuring that your activities and outcomes are closely tied is what allows your program evaluation to accurately assess the impact of the program activities.

☐ No → If the outcomes are not directly tied to program activities, you may need to reconsider whether you have identified the appropriate outcomes. For example, if you are implementing a service member training on coping skills, it is unlikely to improve service members’ ability to intervene with peers who may be at risk for suicide. Before beginning a program evaluation, be sure the program activities and outcomes are closely aligned.

5. **Outcomes fall reasonably within the time frame of the program.**

☐ Yes → This means that the length of the program is sufficient to result in the participant changes listed in your logic model. You do not want to overreach in identifying outcomes that fall far outside the program’s time frame because they are hard to detect. For example, if your program is targeting adolescents, you do not want to list outcomes that will not occur until these adolescents become adults.

☐ No → If the outcomes fall outside the time frame of the program, you may need to reconsider whether you have identified the appropriate outcomes. Identifying outcomes that are close to the conclusion of the program makes the connection between program activities and outcomes stronger.

6. **Evaluation methods are reasonable, given the resources and time frame.**

☐ Yes → This means that you have the resources (e.g., time, money, staff) and expertise to both collect and analyze your evaluation data and that you have linked the timing of the evaluation with the timing of the program. For example, if your program is a two-month training, the evaluation data collection would need to begin before the program and end no sooner than the end of the program.
If your evaluation methods are not reasonable given your resources and expertise, consider hiring an external evaluator. For example, university faculty and graduate students can provide support for evaluation design and data analysis. If you do not have the resources to hire an external evaluator, consider whether you can simplify your evaluation design to fit with your program’s resources and staff expertise. Before implementing an evaluation, be sure you have the resources and expertise to do so.

7. **Evaluation measures provide appropriate assessments of the program outcomes.**

- Yes → This means that each program outcome has one or more evaluation measures associated with it.

- No → Chapter Four can help you identify evaluation measures to capture your program outcomes. If you do not have the resources to collect measures for all program outcomes, you can choose those that are most important to the program and funders or those that are most indicative of program success. However, this limits your ability to capture the full impact of your SPP.

8. **Outcomes correspond to the community needs listed.**

- Yes → This means that your program outcomes are helping to address an identified need. For example, if your program trains mental health professionals to identify patients who are at high risk for suicide, it should help address a shortage of training in that area. As a result, your program should not be competing with many other programs doing the same thing.

- No → If your program outcomes do not fill a need, your program may have trouble recruiting participants because it is competing with similar programs. Consider whether your investment of resources in the program could be redirected to focus on an area where there is an identified need.

If you did not check “Yes” to all of these items, you may need to calibrate the core components of your program. Revising your outcomes to more realistically reflect your program activities is a common way to improve alignment. For example, a training program that does not measure knowledge and change in targeted behaviors would have a poor evaluation design.
Summary

Congratulations, you have completed the first interactive chapter of the toolkit! This chapter provided guidance on how to develop a logic model (Worksheet 2.1) that will be used in subsequent chapters to plan the program evaluation. The core program components you identified in your logic model will be used to help you select evaluation measures and ensure that the time frame of your evaluation reflects program activities. Now that you have a complete logic model (Template 2.1) and have reviewed it using Checklists 2.1 and 2.2, you are ready to begin designing your SPP evaluation. Chapter Three will provide you with guidance about the type of evaluation design that is appropriate for your SPP and walk you through a series of interactive worksheets to ensure that you consider the factors that may influence your SPP evaluation design, such as resources, timing of evaluation data collection, and data sensitivity.
Chapter Three

Objective:
To provide guidance about the type of evaluation that is appropriate for your program based on the type of program and available resources and expertise.

Design an Evaluation for Your Program

In this chapter, we provide a brief summary of past evaluation studies of SPPs similar to your SPP. Evaluation studies are organized according to nine program types (e.g., gatekeeper trainings, crisis hotlines), so the chapter begins by helping you identify your program type. Next, we discuss several key issues to consider when selecting an evaluation design and drafting an evaluation plan. The chapter concludes with a review of the quality of information in the draft evaluation plan.

Identify Your Program Type

Understanding how similar SPPs have been evaluated could help identify evaluation approaches and measures that are useful for your program. Based on our literature review, we identified nine SPP types. Comprehensive approaches to suicide prevention may include one or more of these nine program types, so it is possible that your SPP will reflect more than one. It is still important to specify the program types that are relevant to your SPP because you will want to design an evaluation that captures your entire program. This will help you understand how each component contributes to your SPP's desired outcomes. Below, we briefly describe each of these program types, beginning with appropriate postvention response.

Appropriate Postvention Response

Because there is some evidence that suicides occur in clusters, and because suicides are traumatic events that may act as “triggers” for already vulnerable individuals,
there is guidance available on how to respond to suicides in a responsible way, both to help survivors grieve and, possibly, to prevent future suicides. The most recognized of these are guidelines for the media on how to appropriately cover suicides (Pirkis, 2009; Gould, 2001), but there are also emerging programs that are designed specifically to help communities respond to suicides (Campbell et al., 2004).

**Crisis Hotlines**

Crisis hotlines are telephone hotlines, live chat websites, and short message service (SMS) text lines. These crisis hotlines include the National Suicide Prevention Lifeline and the Veterans Crisis Line and are resources that an individual can contact when in distress. The intent of these hotlines is to provide immediate support and a point of access to care and to administer a short screening protocol to assess suicide risk and make referrals (Gould, Kalafat, et al., 2007; Kalafat, Gould, et al., 2007).

**Gatekeeper Training**

Gatekeeper training sessions educate friends, family members, clergy, and employees in work, school, and other social settings to identify when an individual they know is in distress, learn how and where to refer him or her for help (e.g., to a school guidance counselor or employee support program), and increase these gatekeepers' comfort in making personal referrals (Wyman, Brown, Inman, et al., 2008).

**Marketing Campaigns**

Marketing and outreach campaigns can be used to advertise crisis hotlines and create public awareness about the signs of suicide and symptoms of mental health problems.

**Means Restriction**

 Modifications to the physical environment have led to reductions in suicides. This can include policies that restrict or delay access to firearms or that limit the availability and modify the packaging of lethal medications, as well as the construction of fences or other safeguards on bridges or buildings and the installation of “breakaway” shower curtain or window rods (Florentine and Crane, 2010).

**Mental Health Interventions**

Therapeutic approaches are delivered by mental health providers to treat patients who are at risk of suicidal behavior. Examples include dialectical behavior therapy and cognitive therapy (G. Brown et al., 2005; Linehan, Comtois, et al., 2006).

**Provider Training**

Provider training approaches can fall in three domains: training of primary care physicians on mental health awareness (Rutz, von Knorring, and Walinder, 1989), training of providers who treat suicidal individuals with evidence-based therapies known to reduce suicidality, or more general training in suicide risk assessment and management (G. Brown et al., 2005; Linehan, Comtois, et al., 2006).
Screening Programs
These programs identify and refer individuals at risk of suicidal behavior through the use of standardized instruments administered in primary care and other non–mental health settings. For example, the Columbia Suicide Screen is an 11-item questionnaire embedded in a health survey administered to high school students during regular class time that screens students for mental health problems (Scott, Wilcox, Huo, et al., 2010; Shaffer, Scott, et al., 2004).

Coping Skills and Self-Referral Training
These programs teach individuals to self-identify signs of distress and seek needed support through self-referral. The skills and knowledge built through these trainings is similar to that of gatekeeper programs, but it is focused on encouraging individuals to engage in self-care rather than support the care of others. Examples include general health promotion programs aimed at increasing awareness of the signs of suicide and symptoms of mental health problems, as well as programs focused on enhancing individual protective factors and reducing risk factors associated with suicide. One such program, Signs of Suicide, is a school-based program that teaches youth participants to “acknowledge, care, and tell” about suicide, which includes recognizing the signs of suicide and depression in themselves and others and the specific steps they should take or encourage others to take in response to specific signs (Aseltine, James, et al., 2007).

You can use your work from Chapter Two to help you identify your program type(s). Table 3.1 cross-references each program type with responses to Question 2 on Worksheet 2.1. Use these examples to help you identify which program types best characterize your SPP.
Table 3.1
Program Types, by Response to Worksheet 2.1

<table>
<thead>
<tr>
<th>Program Type</th>
<th>Response Options from Question 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriate postvention</td>
<td>Respond appropriately in the aftermath of a suicide</td>
</tr>
<tr>
<td>response</td>
<td></td>
</tr>
<tr>
<td>Crisis hotlines</td>
<td>Operate a crisis hotline</td>
</tr>
<tr>
<td>Gatekeeper training</td>
<td>Train individuals (e.g., peers, clergy) to recognize individuals at risk for suicide</td>
</tr>
<tr>
<td>Marketing campaigns</td>
<td>Provide information on signs and symptoms associated with suicide or resources available to prevent suicide</td>
</tr>
<tr>
<td>Means restriction</td>
<td>Enforce policies or procedures that restrict access to lethal means (e.g., guns, lethal medications) or create a safer physical environment for individuals at risk for suicide</td>
</tr>
<tr>
<td>Mental health interventions</td>
<td>Provide mental health treatment</td>
</tr>
<tr>
<td>Provider training</td>
<td>Train mental health providers to assess suicide risk or treat patients at risk for suicide</td>
</tr>
<tr>
<td>Screening programs</td>
<td>Screen individuals for suicide risk or mental health problems</td>
</tr>
<tr>
<td>Coping skills and self-referral training</td>
<td>Train individuals to cope with stress or self-refer if they are at risk for suicide</td>
</tr>
</tbody>
</table>

Learn How Similar Programs Have Been Evaluated

You will be using your program type selection (or selections) to identify the relevant literature on SPPs similar to yours, so please consider your selection carefully. If your program includes activities across multiple types, feel free to select more than one type. Once you select your program type (Checklist 3.1), go to Appendix A to read a summary of prior evaluation studies. Doing this reading before you proceed is helpful in preparing you to select an evaluation design using the process outlined in the steps that follow. As mentioned earlier, understanding how similar SPPs have been evaluated could help you identify evaluation approaches and measures that are useful for your program. Brief descriptions of each of the programs types are provided at the beginning of this chapter.
Chapter Three: Evaluation Design

Checklist 3.1
Which Program Type Best Describes Your SPP?

☐ Appropriate postvention response  → See Section A.1 in Appendix A
☐ Crisis hotline  → See Section A.2 in Appendix A
☐ Gatekeeper training  → See Section A.3 in Appendix A
☐ Marketing campaign  → See Section A.4 in Appendix A
☐ Means restriction  → See Section A.5 in Appendix A
☐ Mental health intervention  → See Section A.6 in Appendix A
☐ Provider training  → See Section A.7 in Appendix A
☐ Screening program  → See Section A.8 in Appendix A
☐ Coping skills and self-referral training  → See Section A.9 in Appendix A

Learn the Types of Evaluation Designs

To help you select an evaluation design, we first provide some information on the types of evaluation designs and the expertise, cost, and ease of each. Below are the five most common types of evaluation designs, listed from most to least rigorous:

1. **Pre-/post-intervention evaluation with control group.** This type of evaluation requires you to randomly assign your pool of program participants to either participate in the program (intervention group) or not participate in the program (control group), then collect data at two time points: before the program starts and after the program ends.

2. **Pre-/post-intervention evaluation with comparison group.** This is the same as a pre-/post-intervention evaluation (number 4, below), except that you also collect data from a group that did not participate in the program but is similar in composition to the participating group.

3. **Interrupted time-series analysis.** This approach uses secondary data (i.e., census data, school data) to assess changes at multiple time points before and after the program.
4. **Pre-/post-intervention evaluation.** This type of evaluation involves collecting data from program participants at two time points: before the program starts and after the program ends.

5. **Retrospective pre-/post-intervention evaluation.** This type of evaluation involves collecting data from program participants at only one time point but asking them how their skills, knowledge, or behavior has changed since before the program.

When deciding on a type of evaluation design, you must consider the ease of execution, confidence in the evaluation result, cost of the design, and the expertise needed to gather and use evaluation data (see Table 3.2). As mentioned earlier, these designs are presented from the most rigorous (pre-/post-intervention evaluation with control group) to the least rigorous (retrospective pre-/post-intervention evaluation). Select the most rigorous design that your program has the money and expertise to implement. Ease of execution describes the relative ease of each evaluation design. However, remember that even a retrospective pre/post design can provide you with valuable data on the effectiveness of your program and ways that it can be improved.

### Identify Issues Pertinent to Your Evaluation Design

Several factors influence program evaluation design: (1) number of program participants, (2) program length and expected effects, (3) timing of the evaluation and intended audience, (4) data security and human-subject considerations, (5) evaluation expertise, and (6) resources available for the evaluation. In the following sections, we introduce a number of issues to consider for each of these factors. **Worksheet 3.1** provides a place for you to take notes about your SPP that may be relevant to each of these factors.

#### Number of Program Participants

The number of program participants depends on the outcomes you are measuring. If you are looking for a statistically significant effect—which is an effect that you are confident did not occur by chance—the number of participants in the evaluation or your sample size would be related to your ability to statistically detect an effect. You will need to take into account several things when conducting your evaluation.

- What do you expect to change as a result of the program?
- How many people could you enroll in your evaluation?
- Are there previous evaluation studies of similar programs that can be used to estimate program effects? (See Table 3.3 for evaluation studies that you can use to estimate program effects. The best estimates are from programs with similar activities, number of participants, and evaluation measures.)
- How are the intended outcomes of your program related to one another?

These factors will influence your ability to document a statistically significant effect, and consulting a statistician would help you to determine an adequate number of participants for your evaluation. There are statistical models that can help identify an adequate number based on the program’s expected effects.
### Table 3.2
Types of Evaluation Designs

<table>
<thead>
<tr>
<th>Method</th>
<th>Ease of Execution</th>
<th>Confidence in Result</th>
<th>Cost</th>
<th>Expertise Needed to Gather and Use Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-/post-intervention evaluation with control group</td>
<td>Hard to find group willing to be randomly assigned; ethical issues of withholding beneficial program from control participants</td>
<td>Provides excellent level of confidence that the program caused the change</td>
<td>High; doubles the cost of the evaluation</td>
<td>High</td>
</tr>
<tr>
<td>Pre-/post-intervention evaluation with comparison group</td>
<td>Can be hard to find group that is similar to program group</td>
<td>Provides good level of confidence that the program caused the change</td>
<td>High; doubles the cost of the evaluation</td>
<td>Moderate to high</td>
</tr>
<tr>
<td>Interrupted time series analysis</td>
<td>Requires several years of data collected in the same way, which can be hard to find</td>
<td>Tracks short- and long-term changes, but one cannot be sure that the program caused the change</td>
<td>Inexpensive (data usually collected by other sources)</td>
<td>Low</td>
</tr>
<tr>
<td>Pre-/post-intervention evaluation</td>
<td>Easy way to measure change</td>
<td>Only moderate confidence that the program caused the change</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Retrospective pre-/post-intervention evaluation</td>
<td>Easier than the standard pre/post evaluation</td>
<td>Only moderate confidence that the program caused the change and it may be hard for participants to recall how they were at the start</td>
<td>Inexpensive</td>
<td>Low</td>
</tr>
<tr>
<td>Program Type</td>
<td>Statistically Significant Effects Detected in Evaluation Studies</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate postvention response</td>
<td>Decreased negative emotions</td>
<td>Farberow, 1992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crisis hotline</td>
<td>Decreased number of callers with mental state of “at imminent risk” and increased number rated as “no suicide urgency”</td>
<td>King, Nurcombe, et al., 2003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marketing campaign</td>
<td>Decreased negative emotions and distress</td>
<td>Bryan et al., 2009, Daigle et al., 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means restriction</td>
<td>Decreased suicide rates</td>
<td>Yip et al., 2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental health intervention</td>
<td>Please refer to the evaluation findings for the specific therapeutic approach that you are employing. A summary of findings for multisystemic therapy, problem-solving therapy, LifeSPAN therapy, RUSH, and dialectical behavior therapy are included in Section A.6 in Appendix A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provider training</td>
<td>Improved attitudes and competence levels</td>
<td>Chan, Chien, and Tso, 2009a, 2009b, Rutz, 2001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening program</td>
<td>Increased referrals for mental health services</td>
<td>Husky, Kaplan, et al., 2011, Husky, McGuire et al, 2009, Husky, Miller, et al., 2011</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Coping skills and self-referral training | Decreased suicidal ideations, depression, hopelessness, stress, and anger  
Increased self-esteem and social network support | Eggert, Thompson, Herting, and Nicholas, 1995, King, Strunk, and Sorter, 2011                   |
It is important to note that program evaluations are often not about detecting statistical significance. For example, program staff may set benchmarks (e.g., 20 percent of participants) or be more focused on determining whether clients fall into a different clinical range from pre- to post-intervention.

### Timing of Evaluation and Intended Audience

The length of the program and the anticipated timeline for the evaluation both have implications for the timing of the evaluation. It is better to begin an evaluation before beginning to implement the program so you can collect some data from participants before they are involved with the program. You also want to collect data after participants are involved with the program (e.g., have received training or seen marketing campaign materials). However, the timing of that data collection depends on the length of the program. For example, coordinators of an eight-week gatekeeper training program may want to collect evaluation data right before the training begins (week 1) and right after the training ends (week 8). However, coordinators of a media campaign may want to wait until it has been operating for several months before collecting more data to allow the campaign to fully reach the target population.

Important grant and funding deadlines and reporting requirements may also drive the timing of the evaluation. To ensure that your program’s evaluation is timed appropriately, you must first identify the intended audience for the evaluation. How will the evaluation findings be used? For example, findings could be included in applications for new grant funding, progress reporting on existing grants, annual organizational reports, or briefings to a board of directors, community leadership group, program staff, or program participants. If your program plans to use findings in a variety of ways, identify the earliest that the evaluation would need to be completed and use that as an anchor for timing your evaluation.

When making decisions about the program evaluation design, consider the following questions from **Worksheet 3.1**:

- When does the evaluation need to be completed? This question will help you identify a deadline for completing your evaluation.
- Does the program have an end date? This question will help you identify the timing for post-program data collection.
- Is the program cyclical (e.g., runs for eight weeks twice a year)? This question will also help you identify the timing for data collection to coincide with program cycles.
  - If yes, when is the next time that the program will be offered?
  - If no, how many months has the program been operating?
- How will the evaluation data be used? This question will help you identify ways to report your evaluation findings (e.g., a presentation or briefing to the board, a report to your funder).

### Data Security and Human-Subjects Protection

Evaluation studies involving persons at high risk for suicide are limited, in part because of perceptions that potential risks for participants may exceed any potential benefits. Evaluators also perceive increased liability risks that require significant monitoring and treatment
options. Human subjects protection—or protecting program participants from harm during an evaluation—is of primary importance for any evaluation that collects data related to suicide.

The National Institute for Mental Health, a large funder of research, has released the report *Issues to Consider in Intervention Research with Persons at High Risk for Suicidality* (Pearson et al., 2001), which outlines design considerations and informed-consent issues, provides guidelines for implementing and monitoring risk-management protocols, and specifies the legal risks of research on persons at high risk for suicidal behavior. Implementation of these protocols and guidelines may take additional resources. Therefore, you may want to review the recommendations in the report prior to selecting an evaluation design. The report is available online at http://www.nimh.nih.gov/health/topics/suicide-prevention/issues-to-consider-in-intervention-research-with-persons-at-high-risk-for-suicidality.shtml.

**Evaluation Expertise**

*Evaluation expertise* refers to knowledge and experience in conducting program evaluation. Programs that cannot access evaluation expertise will be limited in the types of evaluation approaches they can feasibly implement. Evaluation expertise can come from academic partners, community-based private and nonprofit organizations, and in-house staff. Often, programs do not have staff or partners with any evaluation expertise. When designing your program’s evaluation, consider the type of evaluation expertise available:

- Do any program staff have evaluation expertise?
- Does the program have established relationships with any academic institutions or community-based organizations with evaluation expertise?

If your program does not have program staff with evaluation expertise or established relationships with organizations that can bring evaluation expertise, you might consider engaging an external evaluator to help support your evaluation efforts. The American Evaluation Association provides a searchable database of members available for evaluation consulting: http://www.eval.org/find_an_evaluator/evaluator_search.asp.

**Available Evaluation Resources**

All evaluations require resources. This can include supplies and equipment (e.g., access to computers, money for copies), staff time, financial resources (e.g., money to support an online survey subscription), and organizational resources (e.g., buy-in from leadership). In addition to the resources your program has set aside for the evaluation, you may be able to leverage additional resources from programs, institutions, and organizations in your community. Understanding the resources available in your community can help save your program time and money and prevent duplication of effort. To help design your evaluation consider the following:

- What kinds of resources does your program have to support the evaluation?
- What other resources are available in your community that could be used to support the evaluation?
Reflect on the questions below; they lay out some key issues that you will need to consider when designing your SPP evaluation.

**Detecting a Program Effect**

- What is expected to change as a result of the program?
- How many people could you enroll in your evaluation?
- Are there previous evaluation studies of similar programs that can be used to estimate the size of program effects?
- If you have more than one outcome, how are they related to one another?

Your notes:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

**Timing of Evaluation and Intended Audience**

- When does the evaluation need to be completed?
- Does the program have an end date?
- Is the program cyclical (e.g., runs for eight weeks twice a year)?
  - If yes, when is the next time that the program will be offered?
  - If no, how many months has the program been operating?
- How will the evaluation data be used?

Your notes:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Worksheet 3.1
Issues to Consider for My Program

Data Security and Human-Subjects Protection


Your notes:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Evaluation Expertise

- Do any program staff have evaluation expertise?
- Does the program have established relationships with any organizations with evaluation expertise?

Your notes:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Available Evaluation Resources

• What kinds of resources does your program have to support the evaluation?
• What other resources are available in your community that could be used to support the evaluation?

Your notes:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Worksheet 3.1
Issues to Consider for My Program

Select an Evaluation Design

Next, based on your answers to the questions in Worksheet 3.1, fill in a portion of your program’s evaluation plan (see Template 3.1). In the “Sample” column, specify the target population and estimated number of program participants (i.e., sample size) for the evaluation. In the “Data Collection” column, first specify the timing of your evaluation (i.e., the date or dates that data collection will occur). Again, these should be tied to the beginning and end dates of your evaluation. In the “Plan for Data Analysis” column, specify the intended audience for the evaluation and any specific plans you have for sharing the findings with this audience (e.g., “Conduct a briefing at the May 11 board meeting”). At this point, the field for frequency in the “Data Collection” column and the “Measures” column should remain blank. You will continue to add details to your evaluation plan in subsequent chapters. Information about available resources and evaluation expertise will be used to select evaluation measures and an analysis strategy.
### Template 3.1
**Evaluation Planner**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Data Collection</th>
<th>Measures</th>
<th>Plan for Data Analysis</th>
<th>Resources Needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target population</td>
<td>Timing</td>
<td></td>
<td>Intended audience</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Person responsible</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Assess the Quality of the Design

Use the following checklist (W. K. Kellogg Foundation, 2000) to review the information in Worksheet 3.1, your draft evaluation plan. Check off each item that is true for your draft evaluation. To help connect the Checklist 3.2 with Template 3.1, we have italicized the terms that appear in both.

☐ Whether your sample size will allow you to detect a clinically or statistically meaningful impact of your program (this could be clinical or statistical). Note that you may need the help of a statistician or evaluation expert to help determine whether your sample size is adequate.

☐ How the length of the program may influence your data collection timing

☐ Whether the evaluation will answer questions of interest to the intended audience(s) for the evaluation

☐ How grant and funding deadlines and reporting requirements may influence the timing of the data collection

☐ How the evaluation data will be used

☐ Whether any of the recommendations from the National Institute of Mental Health report Issues to Consider in Intervention Research with Persons at High Risk for Suicidality (Pearson et al., 2001) are relevant to the program.

Summary

This chapter provided guidance about the evaluation design that will be appropriate for your program based on the type of program and available resources and expertise (Worksheet 3.1). Now that you have completed this chapter, you should have a partially complete evaluation plan (Template 3.1) that is based on a careful consideration of a variety of key issues (Checklist 3.2). You can complete the evaluation plan by stepping through the remainder of the toolkit chapters. Chapter Four will help you select process and outcome evaluation measures.
Chapter Four: Evaluation Measures

Select Evaluation Measures for Your Program

This chapter first defines process and outcome evaluation measures and describes why it is important to collect both process and outcome evaluation measures. Then, it presents process and outcome measures used in prior evaluations of SPPs as possible options for use in your program’s evaluation. The chapter concludes with a review of the measures selected to ensure that they are appropriate for your program and the available level of resources and evaluation expertise.

Select Process Evaluation Measures

Process evaluation is a form of program evaluation designed to document and analyze the early development and actual implementation of a program, assessing whether and how well services are delivered as intended or planned. Process evaluation is also known as implementation assessment (Wholey, Hatry, and Newcomer, 2010; Rossi et al., 2004). Process data can include:

- Tracking participation or attendance
- Tracking participants (collecting demographic data on participants)
- Participant satisfaction surveys
- Measures of implementation activities (program fidelity measures, such as adherence to the program curriculum).
Reasons for Process Evaluation

Measuring the quality of program implementation can answer questions such as the following:

- How much of the program were participants exposed to?
- What are the characteristics of program participants?
- How satisfied are participants and program staff?
- Was the program implemented as intended?

A process evaluation may enhance your understanding of why program outcomes were or were not achieved. For example, if a gatekeeper training program intended to significantly increase participants’ comfort in referring their peers in crisis for help, and the outcome evaluation indicated that this program did not have the intended outcomes, program staff may conclude that the gatekeeper training program was ineffective.

Process evaluation data could help enrich the interpretation of these findings. If attendance data revealed that only 30 percent of program participants attended four or more of eight sessions, program staff may want to reanalyze the data to look only at individuals who attended all or almost all of the program sessions. Improving program retention may be a necessary first step before program effects can be detected.

Demographic data can also be useful in providing insight into whether the population served by the program reflects the intended target population. For example, if demographic data revealed that 25 percent of program participants were outside the target population’s expected age range, program staff may want to reanalyze the data to look only at participants within the expected age range. It may also be useful to compare your program’s evaluation participants with those in previous studies. If you do not find the same program impact as other studies, it may be due to demographic differences between your participants and those in other studies (e.g., you tested using middle school students and prior studies had been done on high school students). You may also find that the program works well with men but not women or vice versa. To look for these types of subgroup differences, you will need to collect demographic data.

Participant or program staff dissatisfaction can minimize program outcomes (e.g., by limiting the amount of knowledge that participants gain during the program). If program participants reported being extremely dissatisfied with their experience with the program, program staff may wish to reanalyze the data to determine the extent to which dissatisfaction increased as program outcomes decreased or to identify sources of dissatisfaction that could be addressed. Additionally, it may be difficult to sustain a program when participants are dissatisfied.

Finally, assessing program fidelity can show whether the implementation adheres to the program’s original design (Smith, Daunic, and Taylor, 2007). If fidelity data revealed that program staff were not fully implementing the required curriculum, improving the fidelity of implementation may be a necessary first step before program effects can be detected. For example, Knox et al. (2010) evaluated the U.S. Air Force Suicide Prevention program and found that suicide rates were significantly lower than before the program was launched, except in 2004, when the program was implemented less rigorously or with lower fidelity.
Possible Options Used in Prior Evaluation Studies

Understanding process evaluation measures that were used by other SPPs could help identify evaluation measures that will be useful for your program. Based on our literature review, we identified sample process measures used in evaluations of SPPs (see Table 4.1). These measures are organized according to the nine program types described in Chapter Three to make it easier for you to identify measures that are relevant for the type of SPP you are implementing. After you review the sample measures, specify which of the process evaluation measures your program plans to use in Template 3.1, your evaluation plan.

### Table 4.1
Sample Process Measures

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appropriate Postvention Response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of individuals trained to respond appropriately</td>
<td>Number of program participants who participated in all eight sessions of the Los Angeles Survivors After Suicide program</td>
<td>Farberow, 1992</td>
</tr>
<tr>
<td><strong>Crisis Hotline</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of call</td>
<td>Amount of time operator spent on a single call</td>
<td>Mishara, Houle, and Lavoie, 2005</td>
</tr>
<tr>
<td>Operator use of specific procedures</td>
<td>This includes tracking the extent to which hotline procedures were followed by reviewing recordings of phone calls. Tracking included determining whether</td>
<td>Mishara and Daigle, 1997</td>
</tr>
<tr>
<td></td>
<td>• A contract was made during the call and whether the caller upheld the contract</td>
<td>Mishara, Chagnon, and Daigle, 2007a,</td>
</tr>
<tr>
<td></td>
<td>• The operator asked the caller whether he or she had thoughts about killing him or herself and whether an attempt was in progress</td>
<td>2007b</td>
</tr>
<tr>
<td></td>
<td>• Rescue procedures were initiated when a caller initiated a suicide attempt and his or her life appeared to be in danger</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• The operator used an empathetic, supportive, and active listening style during the call</td>
<td></td>
</tr>
<tr>
<td><strong>Gatekeeper Training</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence to specific curriculum</td>
<td>Whether the program staff fully covered the curriculum outlined for each session of the program</td>
<td>Kalafat, Madden, et al., 2007 (unpublished)</td>
</tr>
</tbody>
</table>
### Table 4.1
Sample Process Measures

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gatekeeper Training (cont.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gatekeeper suggestions for improving the</td>
<td>An assessment of what gatekeepers liked most about the program, what they liked</td>
<td>Nelson, 1987</td>
</tr>
<tr>
<td>program</td>
<td>least, and suggestions for improving it</td>
<td>Orbach and Bar-Joseph, 1993</td>
</tr>
<tr>
<td>Gatekeeper satisfaction with training program</td>
<td>Gatekeeper's subjective evaluation of the effectiveness of the training program in providing them with the skills and knowledge necessary to identify and refer individuals at risk for suicide</td>
<td>Kato et al., 2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tompkins and Witt, 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cross et al., 2007</td>
</tr>
<tr>
<td>Gatekeeper perceptions of training program</td>
<td>Whether the program was too short/too long, boring/interesting, comforting/upsetting, informative/uninformative helpful/unhelpful for self/peers, etc.</td>
<td>Shaffer, Garland, et al., 1991</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tompkins, Witt, and Abraibesh, 2010</td>
</tr>
<tr>
<td><strong>Marketing Campaign</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating utility of marketing materials</td>
<td>Extent to which the <em>Friends for Life: Preventing Teen Suicide</em> video was clear and informative</td>
<td>Aseltine, 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daigle et al., 2006</td>
</tr>
<tr>
<td>Individuals reached by marketing materials</td>
<td>Counts of the total number of individuals who viewed the marketing materials; for example, the number of Air Force personnel exposed to the 30-minute Air Force Suicide Prevention Campaign briefing</td>
<td>Bryan et al., 2009</td>
</tr>
<tr>
<td><strong>Means Restriction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of individuals affected by means restriction policy</td>
<td>Counts of the total number of individuals in the geographic area or organization affected by the means restriction policy; for example, the number of Hong Kong residents affected by the removal of charcoal packs from retail stores</td>
<td>Yip et al., 2010</td>
</tr>
<tr>
<td><strong>Mental Health Intervention</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapeutic dosage</td>
<td>Extent to which patients are exposed to a specific therapeutic approach; can include the number of therapeutic sessions or amount of time patients spend in therapy</td>
<td>Brown et al., 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rotheram-Borus, Piacentini, et al., 2000</td>
</tr>
</tbody>
</table>
### Mental Health Intervention (cont.)

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient satisfaction</td>
<td>Extent to which patients are satisfied with specific therapeutic approaches</td>
<td>Huey et al., 2004</td>
</tr>
<tr>
<td>Patient attrition</td>
<td>Extent to which patients continued to attend recommended therapeutic treatment sessions</td>
<td>Linehan, Armstrong, et al., 1991</td>
</tr>
<tr>
<td>Provider-patient interactions</td>
<td>These can include observational measures of the extent to which mental health providers utilized specific therapeutic approaches (e.g., dialectical behavioral therapy, multisystemic therapy)</td>
<td>Huey et al., 2004</td>
</tr>
</tbody>
</table>

### Provider Training

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training dosage</td>
<td>Extent to which providers received the training curricula. For example, the number of training sessions providers attended or the number of hours of training</td>
<td>Kato et al., 2010</td>
</tr>
<tr>
<td>Patient satisfaction with provider</td>
<td>Whether patients were satisfied with the application of the skills and knowledge gained during the provider training</td>
<td>Gask, 2006</td>
</tr>
<tr>
<td>Provider perceptions of the training</td>
<td>Whether the provider felt better prepared to help a person at risk as a result of the training and whether the provider would recommend that other providers take the training</td>
<td>Mellonby et al., 2010; Ramberg and Wasserman, 2004a</td>
</tr>
</tbody>
</table>

### Screening Program

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number screened</td>
<td>Total number of individuals screened using a specific screening tool</td>
<td>Oyama et al., 2006; Scott et al., 2009</td>
</tr>
</tbody>
</table>

### Coping Skills and Self-Referral Training

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant interest in the training</td>
<td>Level of participant interest gauged by participant self-reporting and teacher perceptions of how interested participants were in the lessons</td>
<td>LaFromboise and Howard-Pitney, 1994</td>
</tr>
</tbody>
</table>

**NOTE:** The “Reference(s)” column includes references to the source articles describing how the measure was developed and references to evaluation studies and clinical trials in which the measure was used.
Select Outcome Evaluation Measures

Outcome evaluation is an assessment of how well a program’s activities or services have enacted expected changes in the target population or social condition (Rossi et al., 2004).

Reasons for Outcome Evaluation

Outcome evaluations help program staff determine whether their program is having the desired effect. Measuring program outcomes can answer questions about how participants’ knowledge, skills, attitudes, or behaviors have changed as a result of their participation in an SPP. SPPs face unique challenges in evaluating their programs.

Challenges Associated with Measuring Suicide

Suicide rates are often used in evaluations to track whether a program is effective. However, there are three common challenges to using suicide rates as a measure of your program’s effectiveness:

Challenge 1: As mentioned, base rates of the phenomenon being studied (deaths by suicide) are low, which makes it difficult to identify meaningful effects.

Challenge 2: Long time frames are necessary to observe prevention outcomes (i.e., a reduction in suicides).

Challenge 3: It is difficult to classify and track suicides, and there is known variability in how suicides are classified.

If you are interested in using deaths by suicide or suicide rates as an outcome measure, consider the following questions:

- Are the base rates of deaths by suicide high enough to detect a meaningful change? You would need a large geographic area and a high number of suicides to even consider this.
- Does your evaluation track suicide rates over several years?
- Are deaths by suicide all classified in the same way across the geographic area being tracked?

If you answered “no” to any of these questions, then deaths by suicide may not be an appropriate evaluation outcome. Remember, although you are not capturing deaths by suicide as an evaluation outcome, that does not mean that your program does not consider it important. However, you want to design an evaluation using outcomes that can show measurable changes within the time frame of your evaluation.
### Possible Options Used in Prior Evaluation Studies

Understanding the outcome measures used to evaluate other SPPs could help identify measures that are useful for your program. Based on our literature review, we identified measures that correspond to six intermediate and long-term outcomes of SPPs:

- Increased awareness of suicide signs and symptoms and self-care skills
- Improved identification of those at risk
- Increased access to care
- Improved provision of quality mental health care
- Reduced access to lethal means
- Reduced suicidal behaviors (ideation, attempts)
- Reduced suicides.

**Table 4.2** describes sample measures of each of these outcomes from prior evaluation studies. After you review the sample outcome measures, specify which of the outcome evaluation measures your program plans to use in **Template 3.1**, your evaluation plan. In the “Measures” column of **Template 3.1**, write a brief description of the measure(s) you have selected. The description can be taken from **Table 4.2**. Also list the reference(s) you will need to acquire for a copy of the measure. Some measures are not included in the articles referenced. To get a copy of these measures, you will need to contact the authors. Contact information for authors can typically be found as a footnote in the articles referenced.

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
</table>
| **Increased Awareness of Suicide Signs and Symptoms and Self-Care Skills** | **Problem-solving skills**  
Extent to which an individual applies problem-solving skills and has a proactive coping style (e.g., tackles problems head on)                                                                 | Thompson Eggert, and Herting, 2000*** |
| **Knowledge about suicide**             | **Extent of an individual’s knowledge about the signs and symptoms of suicide, as well as the mental health problems associated with suicide, such as depression and substance abuse (e.g., depression is an illness that a doctor can treat)** | Spirito et al., 1988***  
Shaffer, Garland, et al., 1991*** |
| **Attitudes about suicide**             | **Assessment of stigma associated with suicidal behaviors and the extent to which suicide can be prevented (e.g., “If somebody really wants to kill him/herself, there is not much I can do about it.”)**      | Spirito et al., 1988***  
Shaffer, Garland, et al., 1991*** |
### Table 4.2
Sample Outcome Measures

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Increased Awareness of Suicide Signs and Symptoms and Self-Care Skills (cont.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes toward mental health treatment</td>
<td>Extent to which individuals have a negative attitude toward mental health treatment or have concerns that it might affect their decision to seek treatment for a psychological problem from a mental health professional</td>
<td>Rotheram-Borus, Piacentini, Van Rossem, et al., 1996*** Britt et al, 2008***</td>
</tr>
<tr>
<td>Skills associated with help-seeking behaviors</td>
<td>Extent to which individuals have engaged in help-seeking behaviors in the past three months (e.g., In the past three months, have they received treatment from a psychologist?)</td>
<td>Aseltine and DeMartino, 2004***</td>
</tr>
<tr>
<td>Reasons for Living Inventory, Survival and Coping Scale</td>
<td>Assessment of positive expectancies about living as opposed to killing oneself and the importance of these beliefs in resisting suicide</td>
<td>Linehan, Goodstein, et al., 1983**</td>
</tr>
<tr>
<td><strong>Improved Identification of Those at Risk</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suicide intervention skills</td>
<td>Extent to which individuals were able to elicit a promise from a peer not to act on suicidal intentions until talking with someone first, expressed willingness to accompany the peer to a resource person, did not agree to keep a secret about the peer’s suicidal intentions, and displayed active crisis intervention skills</td>
<td>LaFromboise and Howard-Pitney, 1995***</td>
</tr>
<tr>
<td>Self-efficacy in identifying and referring individuals at risk</td>
<td>Extent to which individuals feel comfortable applying suicide prevention skills, active listening, problem-solving, anger management, and stress management skills to identify and refer individuals at risk for suicide to appropriate care; this measure also relates to access to care</td>
<td>LaFromboise and Howard-Pitney, 1995***</td>
</tr>
<tr>
<td>Screening for self-damaging, or impulsive behavior</td>
<td>Extent to which an individual engages in self-damaging behavior, including gambling, binge eating, substance misuse, and reckless driving</td>
<td>Arntz et al., 2003**</td>
</tr>
<tr>
<td>Screening for suicide risk</td>
<td>Use of systematic tool, such as the Symptom-Driven Diagnostic System for Primary Care, the Scale for Suicidal Ideation, or the Suicidal Ideation Screening Questionnaire, to screen individuals for suicide risk; screenings can take place in primary care, school, employment, and other non-mental health and mental health settings</td>
<td>Broadhead et al., 1995* Beck, Brown, and Steer, 1997* Cooper-Patrick, Crum, and Ford, 1994*</td>
</tr>
</tbody>
</table>
### Table 4.2
Sample Outcome Measures

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Improved Identification of Those at Risk (cont.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening for mental health and substance abuse problems</td>
<td>Use of systematic tool, such as the Symptom Checklist–90 or the Alcohol Use Disorders Identification Test, to screen individuals for mental health and substance abuse problems; screenings can take place in primary care, school, employment, and other non–mental health settings</td>
<td>Derogatis and Cleary, 1977*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Saunders et al., 1993**</td>
</tr>
<tr>
<td><strong>Increased Access to Care</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barriers to care</td>
<td>Extent to which an individual experiences barriers that might prevent him or her from seeking treatment for a psychological problem (e.g., depression) from a mental health professional (e.g., a psychologist or counselor)</td>
<td>Britt et al., 2008***</td>
</tr>
<tr>
<td>Mental health care shortage areas</td>
<td>Whether individuals live in a mental health catchment area designated as having a provider shortage</td>
<td>Gaynes et al., 2004***</td>
</tr>
<tr>
<td><strong>Improved Provision of Quality Mental Health Care</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment adherence</td>
<td>Extent to which patients follow treatment regime recommended by mental health providers</td>
<td>Rotheram-Borus, Piacentini, Cantwell, et al., 2000***</td>
</tr>
<tr>
<td>Presence and severity of mental health problems</td>
<td>Extent to which an individual is suffering from a mental health problem as defined by a structured clinical interview for Axis I <em>Diagnostic and Statistical Manual IV (DSM-IV)</em> disorders (e.g., depression, anxiety disorders)</td>
<td>First, Spitzer, and Williams, 2002*</td>
</tr>
<tr>
<td></td>
<td>Severity of depression symptoms as measured by the Hamilton Depression Scale, Beck Depression Inventory, or Center for Epidemiological Studies Depression Scale</td>
<td>Hamilton, 1960*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beck et al., 1961*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Radloff, 1977*</td>
</tr>
<tr>
<td>Service utilization</td>
<td>Extent to which an individual seeks out or makes use of treatment and resources (e.g., counseling, crisis lines)</td>
<td>Gould et al., 2012***</td>
</tr>
<tr>
<td>Medication use</td>
<td>Extent to which an individual is adhering to prescribed psychotropic medications</td>
<td>Linehan and Heard, 1987**</td>
</tr>
</tbody>
</table>
### Table 4.2
Sample Outcome Measures

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Improved Provision of Quality Mental Health Care (cont.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support</td>
<td>Presence or absence of supportive relationships (e.g., family, friends, confidants) and the extent to which an individual is satisfied with or supported by these relationships</td>
<td>Sarason et al., 1987**</td>
</tr>
<tr>
<td>Hopelessness</td>
<td>Extent to which an individual feels his or her life is bleak, is despairing, and believes he or she has no hope of being successful (e.g., “All I can see ahead of me are bad things”)</td>
<td>Kazdin, Rodgers, and Colbus, 1986* Beck et al., 1974*</td>
</tr>
<tr>
<td><strong>Reduced Access to Lethal Means</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery of means-restriction education</td>
<td>Whether means-restriction education or injury-prevention education was delivered to caregivers or family members advising them to dispose of or lock up suicide means</td>
<td>Kruesi et al., 1999***</td>
</tr>
<tr>
<td>Restriction of means</td>
<td>Presence or absence in the home of suicide means (e.g., firearms, over-the-counter medications, prescription medications, street drugs)</td>
<td>McManus et al., 1997***</td>
</tr>
<tr>
<td>Development of a plan for means restriction among individuals at risk for suicide</td>
<td>Extent to which mental health providers develop a plan to restrict access to firearms for individuals at risk for suicide</td>
<td>McManus et al., 1997***</td>
</tr>
<tr>
<td><strong>Suicidal Behaviors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suicidal ideations</td>
<td>Pierce Suicidal Intent Scale administered by physicians to patients includes 12 items (isolation, timing, precautions against discovery, seeking help, final acts, notes, predictable outcome, probability of death, patient’s expectation of lethality, lethality, premeditation, reaction to act)</td>
<td>Pierce, 1981**</td>
</tr>
<tr>
<td></td>
<td>Single-item assessment of suicidal ideations (e.g., “During the past 3 months, did you ever seriously consider attempting suicide?”)</td>
<td>Kann et al., 2000***</td>
</tr>
</tbody>
</table>
### Table 4.2
Sample Outcome Measures

<table>
<thead>
<tr>
<th>Sample Measure</th>
<th>Brief Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suicidal Behaviors (cont.)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suicide attempts</td>
<td>Suicide attempt self-injury interview</td>
<td>Linehan Comtois, et al., 2006**</td>
</tr>
<tr>
<td></td>
<td>Suicidal behavior questionnaire</td>
<td>Linehan, 1981**</td>
</tr>
<tr>
<td></td>
<td>Single-item assessment of suicide attempts</td>
<td>Kann et al., 2000***</td>
</tr>
<tr>
<td></td>
<td>(e.g., “During the past three months, did you actually attempt suicide?”)</td>
<td></td>
</tr>
<tr>
<td>Deaths by suicide</td>
<td>Number or rate of deaths by suicide (e.g., could be collected monthly or annually)</td>
<td>Kinn et al., 2011; DoD data are available from the Mortality Surveillance Division, Armed Forces Medical Examiner</td>
</tr>
</tbody>
</table>

**NOTE:** The “Reference(s)” column includes references to the source articles describing how the measure was developed and references to evaluation studies and clinical trials in which the measure was used.

* Measure must be purchased.

** Measure is available for free.

*** Unclear whether measure must be purchased or is available for free. Please contact the author for more information about this measure.

### Assess the Quality of Selected Measures

Use the following checklist (W. K. Kellogg Foundation, 2000) to review the information in **Template 3.1**, your draft evaluation plan. Check off each item that is true for your draft evaluation plan. The logic model (**Template 2.1**) should include the intended target population, as well as program activities and outcomes, and should be referred to when completing this checklist.
Checklist 4.1
To What Extent Do the Measures Selected Align with Your Program’s Target Population, Activities, and Outcomes?

☐ Fidelity data are linked directly to specific program activities. Refer to the program activities in your logic model.

☐ Demographic or attendance data collected are from the program participants. Refer to the target population in your logic model.

☐ Satisfaction data are collected from either the program participants or staff responsible for implementing the SPP.

☐ Outcome data are linked directly to a specified program outcome. Refer to the logic model for program outcomes.

Summary

Nice job selecting your evaluation measures! This chapter provided information about the process and outcome measures used in prior program evaluations (Tables 4.1 and 4.2). After using this chapter, you should have a completed measures portion of your evaluation plan (Template 3.1) and assessed quality of the measures (Checklist 4.1). Now you’re ready to start collecting data using those measures. Once you have evaluation the data collected, you can proceed to Chapters Five and Six, which will help you analyze your data and use them for program improvement, respectively.
Chapter Five

Analyze Your Program’s Evaluation Data

This chapter begins by summarizing how to establish a database that will allow you to link process and outcome measures over time. Next, we review three simple procedures to ensure the accuracy of the evaluation data. The chapter concludes with three primers that provide step-by-step instructions for analyzing evaluation data at an intermediate and more advanced level using Microsoft Excel® (version 2007 or later). The three primers are meant to be referenced after your program has collected its evaluation data and is ready to begin analysis. If you have not yet completed data collection from your evaluation, you can still review the beginning of this chapter to help prepare for data collection. A more basic introduction to inferential and descriptive data analysis is also available in the Defense Centers of Excellence Program Evaluation Guide (available at http://www.dcoe.health.mil/PsychologicalHealth/Program_Evaluation.aspx) and may serve as a useful foundation for users who are new to program evaluation. Users who are unfamiliar with statistical analysis may also wish to consult a statistician rather than trying to analyze their own program evaluation data.

How to Use This Chapter

The chapter provides guidance on how to

1. Create an evaluation database
2. Decide how to analyze your evaluation data
3. Analyze data using Microsoft Excel (version 2007 or later).
Your approach to the evaluation database and analysis will vary depending on the type of evaluation you are performing. If you are using any type of pre/post design (typical, retrospective, or with a comparison or control group), you will want to be able to link a program participant’s score on the pretest to his or her score on the post-test. Instructions on how to link these scores and enter them into a database for analysis are provided in the following section.

**Establish an Evaluation Database**

An evaluation database should contain rows and columns of data. Each row should represent a different program participant, and the columns represent different variables collected from each participant. If you are using a control or comparison group, you should also create a column that labels each participant as control or comparison (e.g., =0) or intervention (e.g., =1). Your evaluation database should contain unique identifiers for each participant, rather than participant names. Names should be separated from any data before storing or entering the data. These unique identifiers replace names and are needed to ensure that your program does not compromise participants’ confidentiality. If you are collecting process or outcome data that contain participant names or other potentially identifying information, the data will need to be stored securely (e.g., in a locked file cabinet) so that others cannot access this information.

**Unique Identifiers**

To ensure that you can link process and outcome evaluation data, you will need to first assign unique identifiers to all program participants. A unique identifier is usually a numerical code that identifies each participant uniquely without the use of names or other identifying information. An easy way to do this is to create a pre-filled template with numerical codes that can be assigned to specific participants by program staff.

The first example on the following page (Figure 5.1) shows a dosage log or attendance sheet that program staff can use to record the name of each student next to the numerical code that serves as the unique identifier. Once the program is over, the student names can be removed and kept in a single location to ensure data security. Data analysis can then proceed using only the code numbers.

The process evaluation data can then be linked to outcome data by program staff prior to collection. For example, if you are conducting a pre/post survey consider the second example on the following page (Figure 5.2). By completing the participant name and unique identifier and indicating whether the survey is the pre- or post-test on the first page of the survey (tear-off sheet)—and then repeating the code number on the second page of the survey—you can link specific codes with participant surveys. The code number in the example is the number next to the participant’s name on the attendance sheet (e.g., 6101, 7238).

**Accurate Data Entry**

Three procedures can greatly improve the accuracy of the data entry. First, establish ranges or data types within the program (see Figure 5.3). For example, in Excel, you can specify whether the information entered is a number, date, text string, etc., and limit the range of data entered (e.g., must fall between 1 and 4). Setting up these parameters to occur automatically can help staff minimize data entry errors.
Figure 5.1
Sample Attendance Sheet with Unique Identifiers

Prior to distributing this survey, complete the information above and write in the student code # and collection point (pretest or post-test) on the bottom of the next page.

Tear off this sheet when the questionnaire is handed to the student.

Figure 5.2
Survey Cover Sheet with Unique Identifier
Second, have two people assigned to data entry. The first person should be responsible for entering the data, and the second person should be responsible for reviewing the data entered. Third, have a supervisor spot-check 10–25 percent of all entries on an ongoing basis. If errors appear, share these errors with the data entry staff and develop a plan to prevent similar errors or inconsistencies in the future.

Analyze Evaluation Data

There are many ways to analyze your evaluation data. We summarize three options here: (1) using Excel for basic analyses, (2) using a statistical software package, and (3) hiring an external evaluator to do the analyses.

Microsoft Excel can be used to conduct some basic descriptive analyses (e.g., summarize participant characteristics); analyze pre- and post-test data to see whether program participants’ knowledge, skills, behaviors, etc., change; and determine whether program participants’ attendance, satisfaction, or other characteristics contributed to changes in their knowledge, skills, behaviors, etc.

There are statistical packages that help automate many analysis functions. The following are two common statistical packages:

• The Statistical Package for the Social Sciences (SPSS), owned and operated by IBM, ranges in price from $1,020–$5,790 for a basis statistics package that includes a 12-month subscription to a software support hotline. More information about SPSS is available at http://www-01.ibm.com/software/analytics/spss.

These statistical packages are efficient ways to analyze data, but they require users to be proficient with the techniques to use these types of programs and the statistical jargon needed to navigate them.

Another way to analyze evaluation data is to hire an external evaluator with expertise in data analysis and access to statistical software. While costs are involved for both of these options, your analysis can be done more efficiently, and an external evaluator can provide expertise to ensure the data analysis is executed appropriately and accurately reflects program outcomes. In addition, statistical packages afford other advantages, like being able to conduct interrupted time series analyses that are likely to be very useful for programs that collect data over time. The American Evaluation Association provides a searchable database of members available for evaluation consulting: http://www.eval.org/find_an_evaluator/evaluator_search.asp.

Using Excel to Analyze Data

To facilitate access to low-cost analysis options, we provide three analysis primers that describe how to use Excel 2010 to conduct descriptive analyses, run statistical models for detecting differences in your program’s target population, and link process and outcome data. These primers range from basic (i.e., descriptive analyses) to more advanced (i.e., statistical models for detecting differences). All primers were designed to be used with Microsoft Excel 2010, so the examples provide instructions specific to this program. Excel was selected because it is a common and basic processing program. Before selecting a primer, consider the types of questions each type of analyses will provide. You may not need to use all the primers, so use Table 5.1 to make a decision about your data analysis goals and which primer(s) would be most useful to you.

Before you use these primers to analyze data, you will need to enable the “Data Analysis” function in Excel. To do this, go to the data tab at the top of the Excel screen. If you do not see a button labeled “Data Analysis,” you will have to complete the following procedure:

1. Left-click on the green tab marked “File” at the top left of the screen. Then select “Options.”
2. In the left panel, click the “Add-Ins” button. At the bottom of the screen in the right panel next to “Manage,” make sure that “Excel Add-Ins” is selected. Then press “Go.”
3. A new window will open; ensure that both “Analysis ToolPak” options are checked and then click “OK.”
4. When you return to the data tab you should see a button called “Data Analysis.”
Table 5.1
Types of Analyses Addressed in Each Data Analysis Primer

<table>
<thead>
<tr>
<th>Primers in This Chapter</th>
<th>Uses for Each Primer</th>
</tr>
</thead>
</table>
| Primer 1: Calculating Descriptive Statistics for Your Program | Describe the key characteristics of program participants  
Summarize attendance  
Describe participant satisfaction  
Calculate averages and other descriptive data (e.g., percentages, frequencies, ranges, modes) for each outcome variable |
| Primer 2: Statistical Models for Detecting Differences in Your Program’s Target Population | Analyze pre- and post-test data to see whether program participants’ knowledge, skills, behaviors, etc., change (e.g., Does the program achieve its outcomes?) |
| Primer 3: Linking Process to Outcome Measures | Determine whether program participants’ attendance, satisfaction, or other characteristics contributed to changes in their knowledge, skills, behaviors, etc. |
Primer 1: Calculating Descriptive Statistics for Your Program

In this primer, we review how to calculate some basic descriptive statistics of your program. In your evaluation database, you are likely to have three types of variables: dichotomous, continuous, and categorical. The first step of any analysis is to determine what type of variables you have.

- **Dichotomous variables** are those for which there are only two possible outcomes—for example, male or female, true or false, yes or no.
- **Continuous variables** are those with multiple outcomes along a continuum—for example, age or height. Sometimes, there is a series of categorical variables (see below) that are added together to provide a “total score”; these are also often treated as continuous variables.
- **Categorical variables** are those for which there are only a few possible outcomes—for example, race/ethnicity, service branch, or rank. In some instances, categorical variables can be placed on a scale from high to low. An example would be a question that asks people how satisfied they were with a training program on a scale of five responses, from very satisfied to completely dissatisfied. These types of categorical variables are sometimes called ordinal variables.

**Analyzing Dichotomous Variables**

For dichotomous variables, it is almost always the case that you will be calculating a proportion (e.g., the percentage of the sample that is male or that had ever received a particular type of training). For these variables, you simply divide the numerator (e.g., the number of participants who are male) by the denominator (e.g., the total number of participants) and multiply by 100 to calculate the percent:

\[
\frac{\text{numerator}}{\text{denominator}} \times 100 = \%.
\]

**Example:**

\[
\frac{37}{96} = 0.385 \times 100 = 38.5\% \text{ of those offered the training completed it.}
\]

**Calculating Proportions in Microsoft Excel**

1. The key to calculating proportions in Microsoft Excel is that when entering the data for each participant, you insert values of 1 for the numerator (e.g., 1 = male) and 0 for everything else (e.g., 0 = female or unspecified).
Example: If you offered a training program to ten people (three women and seven men) and nine completed the training and one did not, your Excel chart would look like this:

<table>
<thead>
<tr>
<th>Person ID</th>
<th>Sex</th>
<th>Attended training</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note that, in this example, we have coded female as “1” and male as “0.” You can assign 1s and 0s to whichever response option you want, but it is critical that you document and are consistent in these assignments.

2. To calculate a proportion, you first need to calculate the total number of individuals in your sample, which will become your denominator. The easiest way to do this in Excel is to use the “Count” command or type, in the case of our example, “=COUNT(A2:A11)” in the cell. Everyone in the sample has an ID (nobody has a missing value), so you can ask Excel to count the number of unique IDs.

Example: The first row with a valid ID is A2, and the last row with a valid ID is A11. Thus, in cell A12, we type “=COUNT(A2:A11),” and Excel provides us with the total sample size.
3. Because we have converted our dichotomous variables to 0s and 1s, it is easy to count the total number of women or the total number of participants who completed the training by summing the values in the column. In Excel, the easiest way to do this is to highlight the values in the row and click the “AutoSum” button in the top right corner of the Excel toolbar, or you can type “=SUM(C2:C11)” in the cell. Excel will automatically sum the values in the row and place them in the next available cell in that column (in this case, C12).

4. Now, to calculate the proportion, you divide the numerator (B12) by the denominator (A12). In this example, we chose a cell to display the proportion by typing the following formula in cell D12: (=B12/A12).
5. If you want to calculate the percentage, simply highlight the cell in which you have just calculated the proportion and click the “%” button on the Excel toolbar.

Analyzing Continuous Variables
For continuous variables, you are most likely going to be calculating the mean, which is the average response across all unique IDs. For example, one might be interested in the mean duration, in minutes, of a call to a crisis hotline.

Calculating the Mean in Excel

1. In Excel, calculating a mean is straightforward. Simply highlight the cells for which you want the average and choose the “Average” button from the “AutoSum” dropdown menu in the toolbar.

Example: In our example, the average call duration was 10.8235 minutes.
2. When you use the mean, Excel typically includes more values after a decimal point than you might need. To reduce the number of values after the decimal point, choose the applicable cell and click the button indicated below.

---

Analyzing Categorical Variables

For categorical variables, you are most likely going to want to calculate frequencies. This is similar to calculating proportions for dichotomous variables, but for these variables, there are more than two response options. For example, you may want to calculate participants’ responses to a questionnaire that asks them to indicate on a five-point scale how likely they are to intervene with someone who is suicidal. For this analysis, you may want to present the proportion reporting to be very likely, somewhat likely, neither likely nor unlikely, somewhat unlikely, or very unlikely. It is important to note that this scale can also be considered an ordinal variable, which, as defined above, is a specific type of categorical variable. Thus, you can assign values to these response options (for example, very likely = 5, somewhat likely = 4, neither likely nor unlikely = 3, somewhat unlikely = 2, and very unlikely = 1) and calculate the mean response as if it were a continuous variable.

Calculating Frequencies in Excel

1. To calculate frequencies in Excel, it is easiest to use the “COUNTIF” command. First, make sure your responses are entered into the spreadsheet consistently. If there is any discrepancy (for example, misspelled words, extra spaces, letters that are or are not capitalized), the following strategy will not work. Excel requires the exact same spelling and capitalization to count responses as the same category; these counts are required to determine frequencies for each category.

---

Example: Your spreadsheet may look like this
2. Create new cells with all possible response options. In the column next to each, tell Excel to count the number of times within a range of cells the response you are interested in occurs.

**Example:** We placed our cursor in cell C14 and requested that Excel count the number of times the response “Very Likely” occurred between cells B3 and B12 using the command “=COUNTIF(B3:B12, “Very Likely”).” Here, we placed “Very Likely” in quotes because it is text. It tells us that “Very Likely” occurred three times. If we wanted to see how many times a number occurred, we would not need to use quotations.

3. We can then calculate the proportion as we did for the dichotomous variables above. As mentioned earlier, it might be useful to convert some categorical values into numbers and calculate a mean.

**Example:** First, we create a new column called “Numeric Response.” For each cell, we tell Excel that if the value in the column next to it is “very likely,” assign a 5; “somewhat likely,” assign a 4, and so on. This is a somewhat complicated formula that looks like this for each cell in our example:

```
=IF(B3=“Very Likely”,5,IF(B3=“Somewhat Likely”,4,IF(B3=“Neither Likely Nor Unlikely”,3,IF(B3=“Somewhat Unlikely”,2,IF(B3=“Very Unlikely”,1))))).
```
Once we do this for each cell, we can calculate the mean as if it were a continuous variable:

Summary
Users interested in calculating descriptive statistics may find it useful to abide by the following guidance. First, determine whether your variables are dichotomous, continuous, or categorical.

- If dichotomous, calculate a proportion.
- If continuous, calculate a mean.
- If categorical, calculate frequencies. If ordinal, consider calculating a mean.
Primer 2: Statistical Models for Detecting Differences in Your Program’s Target Population

In Primer 1, we identified different types of variables: dichotomous, continuous, or categorical. Here, we present guidance on how to detect differences across these variables. Differences may be among groups (for example, those who did or did not receive training) or among the same individuals (for example, responses before and after the training).

Dichotomous Variables: Differences in Proportions

When the goal is to detect differences in proportions, a common method is a chi-square test. Below, we summarize the steps to conduct a chi-square test. Steps 1–8 explain how to calculate chi-square for two groups (enlisted personnel and officers, in our example). If you have more than two groups (e.g., E1-4, E5-9, junior officers, senior officers) skip to Step 9. The ability to use this method depends on how many people you are studying.

Calculating Chi-Square in Excel

1. The first step is to create what is known as a contingency table. For dichotomous variables, this is a 2×2 table. It is important to note that the procedure we describe here does not work when one of the A, B, C, or D cells has five or fewer individuals. Although there are statistical procedures for handling the scenario of five or fewer individuals in a cell, we suggest consulting with a statistician.

Example: You may want to compare how many enlisted service members versus officers who were offered a course in suicide prevention actually took the training:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Enlisted</th>
<th>Officers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Took Training</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Did Not Take Training</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

2. You must have at least five observations in each cell above (A, B, C, and D) to continue. If you have fewer than five observations in one or more cells, your results will be biased. In this case, you need to perform a different test, which will require the assistance of a statistician. When you have at least five observations in each of the cells above (A, B, C, and D), you can conduct a chi-square test. Here, you will calculate what is called a chi-square test statistic. For each chi-square test statistic, there are two corresponding values: the degree of freedom and the p-value. For a 2×2 table, there will always be one degree of freedom. The p-value is used to indicate whether there is evidence that the two groups are different. It is generally accepted that p-values less than 0.05 indicate that the two groups differ.

3. To determine the chi-square statistic, we begin filling in the numbers in each of the cells (A, B, C, D) of the contingency table.
4. Then, we calculate subtotals across the rows and columns in the table using the “Auto-Sum” function. To do this, highlight the numbers you want to sum and then click the “AutoSum” button, or you can type “=SUM: (B2:C2)” into cell D2 to get the row 2 subtotal, for example.

5. Repeat this process until all the subtotals are filled in.

6. Next, we need to create another table that uses the numbers in the original table to create the number of “expected” outcomes in each cell. This new table becomes a hypothetical comparison group that will allow us to assess whether there was a difference between enlisted personnel and officer participation in the training. To do this, we enter
the expected value into each cell in the new 2×2 table (excluding the subtotal cell). The expected value for a given cell is defined as

\[
\text{expected value} = \frac{\text{row total} \times \text{column total}}{\text{table total}}.
\]

**Example:** To calculate the expected value for enlisted personnel who took the training (G2), in our new table, we type the formula “=(D2*B4)/D4” (which, in this case, is (30*50)/65). In this example, the row total is 30 (D2), the column total is 50 (B4), and the table total is 65 (D4). Repeat this process until all four cells are filled and look like the table on the right, below.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Observed</strong></td>
<td><strong>Enlisted</strong></td>
<td><strong>Officers</strong></td>
<td><strong>Sub total</strong></td>
<td><strong>Expected</strong></td>
<td><strong>Enlisted</strong></td>
<td><strong>Officers</strong></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Took Training</td>
<td>20</td>
<td>10</td>
<td>30</td>
<td>Took Training</td>
<td>23.08</td>
<td>6.92</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Did Not Take Training</td>
<td>30</td>
<td>5</td>
<td>35</td>
<td>Did Not Take Training</td>
<td>26.92</td>
<td>8.08</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Sub total</td>
<td>50</td>
<td>15</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. Next, we calculate the p-value based on the chi-square test. Click in a blank cell and type “=CHISQ.TEST(B2:C3,G2:H3).” Note that B2:C3 is the range of values in the original table, and G2:H3 is the range of values in the expected table. These numbers will differ if your tables are in different cells.

8. The p-value returned for this equation is 0.06921. Because this number is greater than 0.05, we can infer that we do not have evidence that these groups differ with respect to whether they took the training.

9. Do you have more than two groups to compare?

   *Performing a chi-square test across more than two groups is very similar to our earlier example. This time, however, we examine our data divided among a few more categories.*

10. To compare whether there is a difference between various rank groups’ tendency to attend a given training, we first create a table of our observed results (shown below). Once again, we will have to create the row and column subtotals, as well as the overall total.

11. We then create the expected outcomes to use as a comparison group. We do this for all cells.
**Example:** To calculate the expected value for the E1–E4 group that took the training, we choose a cell to display the formula and type “=(F2*B4)/F4).”

12. Finally, calculate the p-value for the observed data. To do this, click in a blank cell and type “=CHISQ.TEST(B2:E3,I2:L3).” Again, the cell numbers in this formula will differ if your tables use different cells.

**Example:** The p-value returned for this equation is 0.01265. Because this number is less than 0.05, we can infer that these groups do differ with respect to whether they took the training.

![](image)

13. Note that when using more than one group, this process does not tell us specifically which groups differ from which. For this, we will need to run a series of tests using 2x2 tables to indicate which group differs from the others. Unfortunately, we need to run a test for every possible pairing.

**Example:** In the example above, there are six possible pairings: E1s–E4s vs. E5s–E9s; E1s–E4s vs. junior officers; E1s–E4s vs. senior officers; E5s–E9s vs. junior officers; E5s–E9s vs. senior officers; and junior vs. senior officers. After conducting each of these six tests, we see evidence that there are differences in attendance among between junior officers and E1s–E4s (p-value = 0.03), junior officers and E5s–E9s (p-value = 0.01), and between junior officers and senior officers (p-value = 0.03). We can thus conclude that junior officers, of whom 75 percent took the training, were more likely to take the training than E1s–E4s (of whom 40 percent took the training), E5s–E9s (of whom 43 percent took the training), and senior officers (of whom 56 percent took the training). Steps for calculating percentages are described in Primer 1.
Continuous Variables: Differences in Means
When the goal is to detect how two groups differ for a continuous outcome, you will first need to ensure that the values that you want to compare for each group are normally distributed, which essentially means that they take the shape of a bell curve when their frequencies are plotted in a histogram (shown below).

Examining Whether a Variable Is Normally Distributed in Excel

1. To begin, we need to create our data.

Example: In this example, we will look at the distribution of the number of suicide prevention training courses that individuals in a certain group have attended. Below, we have a data set of 42 individuals (column A) and the number of suicide prevention training courses they completed, which range from one to ten courses (column B). For now, ignore column C.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Personal Identification</td>
<td>Number of Suicide Prevention Training Courses</td>
<td>Bins</td>
</tr>
<tr>
<td>2</td>
<td>001</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>002</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>003</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>004</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>005</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>006</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>007</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>008</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>009</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>010</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>011</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>012</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>013</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>014</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>015</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>016</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>017</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>018</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>019</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>020</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>021</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>022</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>023</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>024</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>025</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>026</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>027</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>028</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>029</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>030</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>031</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>032</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>033</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>034</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>035</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>036</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>037</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>038</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>039</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>040</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>041</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>042</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
Note that respondents 001–021 are enlisted personnel and respondents 022–042 are officers.

2. Next, we create column C, which adds the categories in which we want to display the data.

**Example:** In the example, we indicate that we want to categorize data as every single number of courses completed between one and ten. We could have easily grouped the data (e.g., one to three courses, four to six courses, seven to ten courses), especially if we were looking at data with more than ten discrete outcomes.

3. To plot a histogram of our data, first click the “Data Analysis” button.

4. Highlight the “Histogram” option and click “OK.” You will need to specify your input range and bin range (see example below).

**Example:** Using the data from the example table, in the input range box, either type “B2:B43,” or highlight the cells with the numeric values in the second column. In the bin range box, type “C2:C11,” or highlight the numeric values in the third column. This second column tells Excel how we want to group our data.

5. Make sure the “Chart Output” button option is selected and then click “OK.” A new Excel worksheet will be created that you can access using the worksheet tabs at the bottom left of the screen. This new tab will have two pieces of output. The first is a frequency table, which summarizes how many people responded with a specific answer. The second is a histogram, which shows up to the right of the table and provides a visual representation of the frequency data.

**Example:** In the example below, the frequency table (left) shows that seven people responded that they had participated in four suicide prevention courses. The frequencies are also displayed in the histogram (right).
6. To check that the data are normally distributed, we look to see whether there is a bell shape to the histogram. A bell-shaped histogram would tend to show a smaller frequency at the far right and left sides and a greater frequency in the middle. Note that reviewing a histogram is an imprecise way of determining the normality of your data. If possible, working with a statistician to conduct statistical tests of normality is the best way to determine whether your data are normally distributed.

Example: Our histogram has a bell-shape. Therefore, we can assume that our data are normally distributed.

7. Are the data normally distributed?

- If the variable is normally distributed in both groups: If in each of two groups the variable is normally distributed, you will conduct a student’s t-test. You will calculate a t-statistic, which (like the chi-square statistic presented earlier) has a corresponding p-value. Again, it is generally accepted that p-values less than 0.05 provide evidence that the two groups differ. Steps for conducting a t-test using Excel are provided on the following page.

- If the variable is not normally distributed in both groups: If the variable is not normally distributed in one or both of the two groups, you will need to conduct a Wilcoxon-Mann-Whitney test. The Wilcoxon-Mann-Whitney test is one of the family of nonparametric statistical tests. Nonparametric tests are based on ranks, rather than raw data, and ranking in Excel is error-prone and not straightforward. In addition, the formulas for these tests are unwieldy, and it is easy to make a mistake. We suggest contacting a statistician or data scientist if you find yourself in need of such a test.

8. Do you have more than two groups to compare?

- To compare means of normally distributed variables across more than two groups: If the variable of interest is normally distributed and you want to compare it across more than two groups, you will conduct a one-way analysis of variance (ANOVA). You will calculate an F-statistic, which has a corresponding p-value. Note that a p-value will only tell you whether the means are different; it will not tell you which means, or whether multiple means, differ from each other. For this process, you will need to calculate t-tests between each pair of groups being compared. Steps for conducting an ANOVA and t-tests using Excel are provided on the following page.

- To compare means of non-normally distributed variables across more than two groups: If the variable of interest is not normally distributed and you want to compare it across more than two groups, you will conduct a Kruskal-Wallis test. Again, using Excel for this test is not straightforward, and it is easy to make a mistake. We suggest contacting a statistician or data scientist if you find yourself in need of such a test.
Conducting a T-Test in Excel

1. To conduct a t-test, click the “Data Analysis” button.

2. Highlight the “T-Test: Two-Sample Assuming Unequal Variances” option and click “OK.”

3. You will need to specify your input range and variable 2 range (see example below). Then click “OK.”

Example: Using the previous example involving the number of suicide prevention training courses, respondents 001–020 are enlisted personnel and respondents 021–042 are officers. To compare these two groups (E1s–E4s and E5s–E9s vs. junior and senior officers), in the input range box, either type “A2:A22” or highlight the cells, and hit “Enter.” For the variable 2 range cell, either type “A23:A43” or highlight the cells, and hit “Enter.”

4. A new tab will be added to the current tabs at the bottom left of the screen. This new tab will have the t-test output.

5. We can then use the p-value to infer whether the findings are significant.

Conducting an ANOVA in Excel

1. To conduct an ANOVA, first click the “Data Analysis” button.

2. Highlight the “ANOVA: Single Factor” option and click “OK.”

3. You will need to specify your input range (see example). Then click “OK.”

Example: Using the previous example involving the number of suicide prevention training courses, respondents 001–021 are enlisted E1–E4s and E5–E9s, and respondents 022–042 are junior and senior officers. To compare these four groups (E1–E4s, E5–E9s, junior officers, and senior officers), you first need to calculate the means for each group. See Primer 1 for instructions on calculating means. In the input range box, either type “B2:E2” or highlight the cells.

<table>
<thead>
<tr>
<th></th>
<th>E1-E4</th>
<th>E5-E9</th>
<th>Junior Officers</th>
<th>Senior Officers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean number of suicide</td>
<td>5.2</td>
<td>4</td>
<td>4.7</td>
<td>5.833333333</td>
</tr>
<tr>
<td>prevention training</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>courses</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. A new tab will be added to the present tabs at the bottom left of the screen. This new tab will have the ANOVA output. Under “F” in the lower table is the F-statistic.

Example: The F-statistic in the previous example is 0.614. The p-value for this example is 0.641. Because the p-value is above 0.05, we cannot infer that the means differ more than would be expected by chance.
Categorical Variables: Differences in Frequencies

1. To test differences in frequencies (for example, the proportion of individuals who were very likely, somewhat likely, neither likely nor unlikely, somewhat unlikely, or very unlikely to ask for help before and after suicide prevention training), you can still use the chi-square tests described here.

2. In this case, you create an RxC contingency table, where R stands for “row” and C stands for “column.”

**Example:** You may want to test whether soldiers of different ranks reported different coping strategies:

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Very Likely</th>
<th>Somewhat Likely</th>
<th>Neither Likely nor Unlikely</th>
<th>Somewhat Unlikely</th>
<th>Very Unlikely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Course</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>After Course</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>J</td>
</tr>
<tr>
<td>Did Not Take Course</td>
<td>K</td>
<td>L</td>
<td>M</td>
<td>N</td>
<td>O</td>
</tr>
</tbody>
</table>

3. It is important to note that, in this case, the test will only tell you whether one proportion is different from another proportion; it will not tell you which proportions, or whether multiple proportions, differ from each other.
Primer 3: Linking Process to Outcome Measures

In this statistical primer, we describe some ways to measure associations between variables, which will enable us to test whether such variables as dosage/attendance, satisfaction, and fidelity are associated with program outcomes.

Both Variables Are Dichotomous

1. When the goal is to detect whether one dichotomous variable is associated with another variable, you will use the same rule as you did for testing differences in proportions in Primer 2. The first step is to create a 2×2 table.

Example: You may want to compare whether those who took a course in suicide prevention were more likely to intervene with a person in crisis than those who did not take the course:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intervened</th>
<th>Did Not Intervene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Took Training</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Did Not Take Training</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

2. In essence, you are still testing the difference in proportions: whether the proportion of those who intervened was larger among those who took the training than among those who did not take the training. That is,

\[ \frac{A}{A+B} > \frac{C}{C+D} \]

3. Recall from Primer 2 that when the number in each cell is larger than 5, you will conduct a chi-square test. See Primer 2 for instructions on how to conduct this test in Excel. You must have at least five observations in each cell above (A, B, C, and D) to continue. If you have fewer than five observations in one or more cells, your results will be biased. In this case, you need to perform a different test, which will require the assistance of a statistician.

Both Variables Are Continuous

1. In certain cases, you may want to examine how two continuous variables are related. In statistical terminology, we use estimates of correlation to quantify the relationship between two continuous variables.

Example: You might measure satisfaction with a class on a scale from 1 to 100, and you might measure an outcome as symptoms on a scale from 1 to 50.
2. In this case, as shown in Primer 2, you will first need to ensure that the values that you want to compare among each group are normally distributed, which essentially means that they take the shape of a bell curve when their frequencies are plotted in a histogram.

3. Are variables normally distributed?
   – When both variables are normally distributed, you will calculate the Pearson correlation coefficient (often indicated as \( r \)). When \( r \) is greater than 0, this indicates that the two variables are positively correlated (as one increases, the other increases). If \( r \) is less than 0, then the two variables are negatively correlated (as one increases, the other decreases). Finally, if \( r \) is equal to 0, the two variables are uncorrelated: There is no linear relationship between them. Like the test statistics we calculated in Primer 2, \( r \) also has a p-value. It is generally accepted that p-values less than 0.05 provide evidence that the two variables are significantly correlated (either negatively or positively).
   – When one or both variables are not normally distributed, you can still assess whether they are associated with each other. In this case, you will calculate the Spearman rank-correlation coefficient (often indicated as \( r_s \)). You can interpret \( r_s \) in the same way that you interpreted \( r \) above. When you have ten or more observations, \( r_s \) also has a p-value. It is generally accepted that p-values less than 0.05 provide evidence that two variables are significantly correlated (either negatively or positively).

**Calculating the Pearson Correlation Coefficient in Excel**

1. To calculate the Pearson correlation coefficient in Excel, we first need to obtain or enter the raw data.

   ![Excel Table Example]

   **Example:** We will create a list of observations for whether an individual intervened with a person in crisis and whether or not he or she took the training. The two variables, which we label “intervened” and “took training,” are called dichotomous variables, because they can only have two outcomes: yes (1) or no (0).
2. Now that the data are entered, in the data toolbar ribbon, click the “Data Analysis” button.

3. In the “Analysis Tools” list, highlight “Correlation” and click “OK.”

4. In the ensuing window, in the input range box, either drag and hold the cursor over all the numeric data or enter “A2:B16” in the box. Make sure that “Group by Columns” is selected and click “OK.”

5. A new Excel tab is created and, in this case, a 2×2 cell output table is also created. The output table shows the correlation coefficient between the two variables in our raw data.

6. Column 1 × column 1 and column 2 × column 2, by definition, should always equal 1. This means the two columns are perfectly correlated.

Example: We are interested in the column 1 × column 2 output. A strong correlation is close to 1, and a weak correlation is close to 0. The result is 0.207, which means that the two columns are weakly correlated.

Both Variables Are Categorical

When both variables are ordinal: In Primer 1, we explained that categorical variables are sometimes ordinal. If you have two ordinal variables (for example, satisfaction with a course on a scale from 1 to 5 and the number of training sessions a person attended), you can examine whether they are correlated using the Spearman rank-correlation coefficient, described earlier.

Neither variable is ordinal: To examine how two categorical variables are associated with each other when neither is ordinal, you can still use the chi-square to determine whether there are differences in proportions, described in Primer 2.

One Variable Is Dichotomous, and One Is Continuous

If you have one dichotomous variable (for example, did or did not intervene) and you are interested in whether it is associated with a continuous variable (for example, performance, measured on a scale from 1 to 100), you can use the methods for comparing means across two groups (student’s t-test, Wilcoxon-Mann-Whitney test) described in Primer 2.
One Variable Is Categorical, and One Is Dichotomous
If you have one dichotomous variable (for example, did or did not intervene) and you are interested in whether it is associated with a categorical variable (for example, satisfaction, measured on a scale from 1 to 5), you can use the methods for comparing proportions across more than two groups (chi-square test) described in Primer 2.

One Variable Is Categorical, and One Is Continuous
If you have one categorical variable (for example, satisfaction measured on a scale from 1 to 5) and you are interested in whether it is associated with a continuous variable (for example, performance, measured on a scale from 1 to 100), you can use the methods for comparing means across more than two groups (ANOVA or Kruskal-Wallis test) described in Primer 2.

Summary
Nice job completing your evaluation data analysis! This chapter provided information about the ways in which you can get support to analyze data (e.g., through a statistical software package or through an evaluation or statistics expert) and provided you with three primers to support some basic data analysis using Microsoft Excel. After using this chapter, you should have selected your evaluation data analysis strategy and begun implementing it—whether on your own or with help from the experts. Once your data analysis is complete, you can proceed to Chapter Six, which will help you use that analysis for program improvement.
Chapter Six

Use Evaluation Data to Improve Your Program

This chapter begins by assessing some basic information about your program’s evaluation to better interpret the results. This assessment is followed by a review of the evaluation findings to determine the extent to which the program achieved its intended outcomes. The chapter concludes with a series of small-scale assessments to identify appropriate changes needed to improve the quality of the program. If you have not yet completed the analysis of the data from your evaluation, you should do that before following the steps in this chapter.

Assess Participation in Your Evaluation

Before interpreting and applying the evaluation data, revisit the final evaluation design that was actually implemented to determine how well the evaluation participants reflect the intended program participants. You will need the information on who participated in the evaluation and how well that aligned with the program’s intent to help inform decisions about the types of program improvements needed. Answering the questions in Worksheet 6.1 will help you assess participation in your evaluation. Complete Worksheet 6.1 before interpreting your evaluation data.
Below is an example of how you might complete the worksheet:

If your program evaluation began in January 2012 and ended in August 2012, you would record this range as your period of reporting (Question A, January–August 2012). If your program intended to reach 50 leaders on your installation (Question B) and you were able to recruit 40 of those leaders (Question C) to attend your program at least once, the percentage of participants reached would be 40 divided by 50 × 100 = 80 percent. Of the 40 leaders who attended your program at least once, 36 participated in your evaluation, making the percentage participating in your evaluation 90 percent (36 / 40 × 100). These two percentages will help guide your response to Question H. Because the percentage of participants reached and percentage of that group who participated in your evaluation were both above 75 percent, your evaluation represented the intended population “very well.” This suggests that you should have strong confidence in your evaluation results. Your evaluation results may not accurately describe the effectiveness of your program if they do not or only somewhat represent the intended population. Improving program recruitment and retention strategies can help ensure that you get the needed participation to accurately assess your program’s effectiveness.
# Worksheet 6.1
Assessing Participation in Your Program’s Evaluation  
(adapted from unpublished research by Hunter et al.)

<table>
<thead>
<tr>
<th>A. What is the period of reporting?</th>
<th>B. How many participants did you plan to reach with your program?</th>
<th>C. How many attended your program even once?</th>
<th>D. How many people participated in the evaluation?</th>
</tr>
</thead>
<tbody>
<tr>
<td>_____________________________</td>
<td>_____________________________</td>
<td>_____________________________</td>
<td>_____________________________</td>
</tr>
</tbody>
</table>

E. % of participants reached: ________  
   (number of participants who attended your program even once/number of participants you planned to reach × 100)

F. % of participants in the evaluation: _____  
   (number of participants in the evaluation/number of participants who attended your program even once × 100)

G. Who took part in the evaluation?  
   - Program completers  
   - Regular attendees  
   - Everyone who ever attended  
   - Others

H. How well does your evaluation represent the population you intended to reach?  
   (Using the information above, check one.)  
   - **Not at all well**: This means that you did not reach the program participants you planned to reach (% of participants reached was less < 50%). It can also mean that you reached most or some of the participants you planned to reach (% of participants reached was > 50%), but few participated in the evaluation (% of participants in the evaluation was < 75%).  
   - **Somewhat well**: This means that you reached some of the program participants you planned to reach (% of participants reached was > 50%). Of those reached, most participated in the evaluation (% of participants in the evaluation was < 75%).  
   - **Very well**: This means that you reached most or all of the program participants you planned to reach (% of participants reached was > 75%), and most participated in the evaluation (% of participants in the evaluation was < 75%).
Interpret the Evaluation Data

Next, we will interpret evaluation findings by assessing whether they provide useful information about the extent to which the program achieved its intended outcomes. Worksheet 6.3 asks you to document, through a series of questions, what types of effects the program had on each outcome and whether the program met, missed (fell short of), or exceeded expectations. Continuous quality improvement (CQI) efforts should prioritize actions directed at program outcomes where the program missed expectations. The final column in the worksheet should be used to record any barriers that may have contributed to the “missed expectation” rating. Overall, you are trying to determine whether these results suggest that changes to the program are needed.

To use this worksheet, first transfer the outcomes and associated program activities from your logic model (Template 2.1) into the first and second columns, respectively. Then, record the results of your analysis (Chapter Five) in column 2. Specify any changes in participants’ behaviors, knowledge, skills, etc.

Column 2 also allows you to identify trends in the results. If you did not see any differences in participants as a result of your program, mark that things stayed the “same.” If you did see differences, specify whether these were improvements (i.e., “better”) or setbacks (i.e., “worse”) based on your specified outcome. For example, if you intended to improve attitudes toward help-seeking among service members through a media campaign, and the results of your evaluation of the campaign found that attitudes became more stigmatizing, this would be a setback and would get marked as “worse” on the worksheet.

Next, specify whether these results were what you expected in column 3. Refer back to your intended outcome in column 1 to help you determine whether you achieved the results you were expecting. Building off the media campaign example, if you expected that the campaign would improve attitudes and it did not, then you would check “missed” on the worksheet. Your program will need to take action to improve the outcomes for which evaluation results did not meet expectations (i.e., those marked “missed” in the third column), so mark these as “Yes” in the fourth column (“Action Needed?”). For outcomes for which action is needed, spend time reflecting on any barriers that could have kept your program from achieving that outcome and record them in the last column. Worksheet 6.2 provides an example of how to fill out this worksheet. Table 6.1, discussed next, is intended to help you reflect on possible barriers.
## Worksheet 6.2
Review Program Outcomes, with Example
(based on unpublished research by Hunter et al.)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Difference/Change in Any of the Outcomes?</th>
<th>Met Expectations?</th>
<th>Action Needed?</th>
<th>Potential Barriers (e.g., resources, expertise)?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example:</strong> After Springfield High School students participate in the program for two months, their recognition of suicide risk factors will increase by 20 percent.</td>
<td>15 percent increase in the Springfield High School students’ recognition of suicide risk factors</td>
<td><strong>What is the trend?</strong></td>
<td><strong>Did this meet your expectations for the program?</strong></td>
<td><strong>Yes</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Better</td>
<td>☑ Met</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>☑ Same</td>
<td>☑ Missed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>☑ Worse</td>
<td>☑ Exceeded</td>
<td></td>
</tr>
</tbody>
</table>

**Was this related to any program activities (as measured by process evaluation measures)?**
Yes, attendance data. These data showed that 65 percent of the students participated in the full two-month program.
Worksheet 6.3
Review Program Outcomes
(based on unpublished research by Hunter et al.)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Difference/Change in Any of the Outcomes?</th>
<th>Met Expectations?</th>
<th>Action Needed?</th>
<th>Potential Barriers (e.g., resources, expertise)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Was this related to any program activities (as measured by process evaluation measures)?</td>
<td>Did this meet your expectations for the program?</td>
<td>Action Needed?</td>
<td>Potential Barriers (e.g., resources, expertise)?</td>
</tr>
<tr>
<td></td>
<td>□ Better</td>
<td>□ Met</td>
<td>□ Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Same</td>
<td>□ Missed</td>
<td>□ No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Worse</td>
<td>□ Exceeded</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. What is the trend?
   - Better
   - Same
   - Worse
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Difference/Change in Any of the Outcomes?</th>
<th>Met Expectations?</th>
<th>Action Needed?</th>
<th>Potential Barriers (e.g., resources, expertise)?</th>
</tr>
</thead>
</table>

2. **Was this related to any program activities (as measured by process evaluation measures)?**

**What is the trend?**
- Better
- Same
- Worse

**Did this meet your expectations for the program?**
- Met
- Missed
- Exceeded

**Yes**
**No**

3. **Was this related to any program activities (as measured by process evaluation measures)?**

**What is the trend?**
- Better
- Same
- Worse

**Did this meet your expectations for the program?**
- Met
- Missed
- Exceeded

**Yes**
**No**
Make a Plan to Improve Your Program

Your review of the evaluation data could suggest a number of different actions. Table 6.1 presents a series of results-based scenarios and their associated program improvement strategies. If you indicated that the program did not meet expectations (i.e., “missed”) in Worksheet 6.3, the program activities did not result in a significant change on intended outcomes. If you indicated that the evaluation participants did not represent the target population well (i.e., “not at all well”), the program activities were not implemented with adequate dosage or fidelity. Your process evaluation data may also help you decide whether program activities were or were not implemented with adequate dosage (i.e., participants got most or all of the intervention) and fidelity (i.e., program content was delivered as intended). Identify which scenario best describes your evaluation results and then proceed to the small-scale CQI assessments, which will help you to select improvement strategies, if needed.

If changes to your program are needed, use the following CQI assessment to identify the specific changes needed and to plan key activities (see Checklist 6.1). The CQI assessment walks through a number of potential challenges to help you determine whether the problem is applicable to your program and then provides a description of potential changes to help address the problem. If you answer “no” to any of the questions about potential CQI challenges, review the actions to address challenges and select a couple of actions that your program can feasibly pursue.
### Table 6.1
Results-Based Scenarios and Associated Strategies for Program Improvement

<table>
<thead>
<tr>
<th>Program activities resulted in a significant change on intended outcomes.</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>
| Yes | The program seems to be working as designed; continue implementation and evaluation.  
- Keep monitoring process and outcomes. | Changes may be needed because your program did not achieve the intended results. Lack of significant results could be related to the following barriers:  
- A small sample size  
- A misalignment between program activities and intended outcomes  
- A mismatch between the program and its participants (e.g., participants started high on outcome measures, leaving limited room to improve)  
- A mismatch between evaluation design or measures and program outcomes.  
Address these potential barriers before implementing the program. Addressing these barriers may require you to change your program activities and your evaluation design/measures. | |
| No | Changes may be needed to address potential barriers to fidelity and dosage; you might want to reevaluate the program to determine whether results improve.  
- Focus on strengthening areas that were not implemented with fidelity.  
- Improve recruiting or retention strategies to ensure that participants get an adequate dosage. | Changes may be needed because the program does not seem to be working as designed. Significant changes to program activities and evaluation design may be required. Assess whether  
- There is a mismatch between the program and the intended population  
- There are adequate resources to deliver the program (e.g., Do facilitators have enough training? Do participants have incentive to attend program?).  
Proceed to Checklist 6.1. | |
A. Did participants represent your target population?

☐ Yes

☐ No → Review referral sources to determine whether you have the right relationships in place to get appropriate referrals. Review eligibility criteria to ensure that they are clear enough to recruit appropriate program participants.

B. Was the program delivered as intended?

☐ Yes

☐ No → Improve staff training on how to implement the program and self-assess fidelity.

C. Was attendance adequate?

☐ Yes

☐ No → Revisit recruitment and retention practices to identify where improvements can be made. Assess whether there are any logistical barriers that might make it difficult for participants to attend (e.g., transportation). Consider whether changing the time or place of the program would improve participation. Consider whether the program is appropriate for the population served.

D. Did you have the resources needed to implement the program completely and as intended?

☐ Yes

☐ No → Review your program’s resources for implementation and evaluation to determine whether you have the right staff, resources, and partnerships to deliver the program. Try to leverage additional resources from untapped sources in your community. A community resources assessment may help inform this effort.
Chapter Six: Using Evaluation Data

Checklist 6.1
What CQI Actions Are Needed to Improve the Program?

E. Were the outcomes you expected reasonable/appropriate for the program?

☐ Yes

☐ No  Revisit the goals and logic model that you developed and revise them to be more reasonable/appropriate for your program.

F. Was your process and outcome evaluation appropriate?

☐ Yes

☐ No  Update the process and/or outcome evaluation plan to be more appropriate for your program.

G. Were you using an evidence-based program?1

☐ Yes

☐ No  Consider changing your program content to use an evidence-based approach. Evidence-based programs are included in the National Registry of Evidence-Based Practices and Programs.

Transfer over the relevant improvement actions from any items on Checklist 6.1 for which you answered “No” to Worksheet 6.4. Then, record who will participate in the action, who will be responsible for the action, the resources needed, location details, and the target date for improvement. Making a plan for program improvement using Worksheet 6.4 will help you identify the activities necessary to achieve those objectives and specify a target date for completion of program improvement activities. If possible, complete the program improvement activities prior to implementing the program again.

1 An evidence-based program is a program that has been determined to be effective through rigorous scientific evaluations and randomized controlled trials, has a significant and sustained effect on intended outcomes, and has been tested using large longitudinal studies or multiple replications (Evidence-Based Prevention and Intervention Support Center, 2011).
## Worksheet 6.4
Program Improvement Plan

<table>
<thead>
<tr>
<th>Improvement Action</th>
<th>Who Will Participate</th>
<th>Who Is Responsible</th>
<th>Resources Needed/Source</th>
<th>Location/Details</th>
<th>Date of Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

Congratulations on completing the evaluation toolkit! We hope you found it helpful for planning and implementing an evaluation and using evaluation data for program improvement. If you skipped any steps, consider going back and completing those portions of the toolkit. Each chapter builds on information from the previous chapter:

- **Chapter Two** helps you identify the core components of your SPP and organize these components into a logic model to clearly visualize the relationships and dependencies between components. This chapter also contains tools to help you review your logic model, assessing whether it is complete and reasonable.

- **Chapter Three** helps you select an evaluation design appropriate for each of your program’s core components by sharing information from prior evaluations.

- **Chapter Four** helps you select process and outcome evaluation measures that are appropriate for your evaluation design by sharing information from prior evaluations.

- **Chapter Five** describes how to enter and analyze the evaluation data collected using evaluation measures.

- **Chapter Six** helps you interpret the process and outcome evaluation results to guide improvements.

Fully completing each chapter will help you ensure that you have taken advantage of the comprehensive guidance provided. The content of the toolkit is intended to be used for continuous quality improvement. So, once you complete **Chapter Six** and determine whether improvements are needed, you will need to begin again at **Chapter Two** by updating your program’s core components with any improvements. The content and worksheets are intended to be reused, even after you have started an evaluation or enhanced an existing evaluation. Consider reviewing the toolkit annually to continue improving your program and refining your evaluation.
Appendix A

Summary of Program Evaluation Studies, by Program Type

This appendix provides summaries of prior evaluation studies, categorized according to the program types outlined in Chapter Three. The evaluation studies are organized as follows:

- **Section A.1:** Evaluation Studies of Appropriate Postvention Response Programs (pp. 116–117)
- **Section A.2:** Evaluation Studies of Crisis Hotlines (pp. 118–123)
- **Section A.3:** Evaluation Studies of Gatekeeper Training (pp. 124–133)
- **Section A.4:** Evaluation Studies of Marketing Campaigns (pp. 134–136)
- **Section A.5:** Evaluation Study of Means Restriction (p. 137)
- **Section A.6:** Evaluation Studies of Mental Health Interventions (pp. 138–145)
- **Section A.7:** Evaluation Studies of Provider Training (pp. 146–150)
- **Section A.8:** Evaluation Studies of Screening Programs (pp. 151–155)
- **Section A.9:** Evaluation Studies of Coping Skills and Self-Referral Training (pp. 156–160).
A.1. Evaluation Studies of Appropriate Postvention Response Programs

Description of Program:

The Los Angeles Survivors After Suicide program provides eight sessions, once a week, conducted by two leaders (a mental health professional and a survivor of suicide who has gone through the program and received additional training). After eight weeks, participants are invited to attend monthly meetings for as long as they wish.

Population:

Friends and families of people who die by suicide (survivors)

Description of Evaluation:

Using an intervention and control group, the study examined the results of the program through pre- and postintervention questionnaires.

Summary of Evaluation Findings:

By the end of the program, the intervention group had significantly decreased ratings on all emotions except feeling suicidal themselves (which was low to start with). The control group had only decreased anxiety. The intervention group’s ratings also decreased more than those of the control group.

Sample Size:

n = 82 participants (60 intervention, 22 control)

Design:

Nonequivalent control group

Reference:

Farberow, 1992
Appendix A: Evaluation Studies

A.1. Evaluation Studies of Appropriate Postvention Response Programs

Description of Program:
Local Outreach to Survivors of Suicide (LOSS) is an active postvention intervention. Crisis center staff, as well as volunteer survivors who have substantial training, respond at the scene of suicide. LOSS team members provide comfort to survivors, explain the protocols used to investigate the scene, and help answer questions related to the many other responders at the scene.

Population:
All adult suicide survivors seen at Baton Rouge Crisis Intervention Center from 1999 to 2005

Description of Evaluation:
The study compared suicide survivors who received the LOSS active postvention with those who did not to determine whether there were differences in the elapsed time between death and seeking services for bereavement, clinical problems since death, and engagement in group treatment.

Summary of Evaluation Findings:
Those who received LOSS support

- Presented for an intake sooner (48.5 + 74.9 days, range 2–532 vs. 97.5 + 147.0 days, range 3–883)
- Were more likely to attend support group meetings (69.3% vs. 59%, p < 0.05) and attended more group meetings, t(164) = 2.22, p < 0.05.

The study found no differences in clinical problems since death.

Sample Size:
n = 356 participants (150 received LOSS; 206 did not receive LOSS)

Design:
Retrospective comparison group

Reference:
Cerel and Campbell, 2008
A.2. Evaluation Studies of Crisis Hotlines

Description of Program:

A Kansas City crisis hotline was available to citizens in the Kansas City area. It offered an initial telephone interview and made referrals to local mental health clinics, family physicians, and others.

Population:

Residents of the Kansas City area experiencing a personal crisis

Description of Evaluation:

An evaluation of the demographic data records from a three-year period (September 1, 1966, to August 31, 1969) of crisis hotline calls were compared with data on those who had died by suicide. Names were compared to see whether the names of those who had died by suicide were found in the call logs of the hotline.

Summary of Evaluation Findings:

The findings support the hypothesis that suicide attempters and those who die by suicide constitute two epidemiological populations, albeit overlapping, and that the crisis intervention method of suicide prevention programs can reach the first group but not the second. In other words, the demographics of the callers more closely resembled the attempters group than the suicide completion group.

Sample Size:

n = 543 telephone clients for the 3-year period; n = 389 recorded suicides in the same period in the counties studied

Design:

Quasi-experimental

Reference:

Bidwell, Bidwell, and Tsai, 1971
A.2. Evaluation Studies of Crisis Hotlines

Description of Program:
National Suicide Prevention Lifeline centers attempt to reduce callers’ current crisis or suicidal states and provide referrals to mental health care. Many federal and community public-awareness campaigns reference the Lifeline.

Population:
U.S. residents experiencing a suicidal crisis

Description of Evaluation:
Lifeline callers who had received a mental or behavioral health care referral were interviewed two weeks after their call to assess depression, referral follow-through, and barriers to utilization. The sample included both suicidal callers and nonsuicidal callers.

Summary of Evaluation Findings:
Only 51.6% of subjects actually sought mental health services after their referrals. The most common barrier cited was the caller’s perception about mental health problems, followed by the caller’s financial problems. The study observed no significant difference in the rates of mental health service utilization between suicidal and nonsuicidal crisis callers.

Sample Size:
\( n = 654 \) callers (376 suicidal callers, 278 crisis callers)

Design:
Nonexperimental (no control group)

Reference:
Gould, Munfakh, et al., 2012
A.2. Evaluation Studies of Crisis Hotlines

Description of Program:
Kids Help Line provides telephone counseling assistance for youth under age 18 and operates 24 hours a day, seven days a week.

Population:
Youth under 18 years old experiencing a personal crisis

Description of Evaluation:
Independent raters quantified changes in suicidality over the course of a call/counseling session by reviewing the first five minutes when suicidality became evident and the last five minutes of the call.

Summary of Evaluation Findings:
The study showed decreases in callers' mental state and suicide ideation from the beginning to the end of the call, a decrease in the number of callers rated to be at "imminent risk," and an increase in those rated as "no suicide urgency."

Sample Size:
n = 101 callers randomly taped over 12 months

Design:
Nonexperimental (no control group)

Reference:
R. King et al., 2003
A.2. Evaluation Studies of Crisis Hotlines

Description of Program:
Suicide-Action Montreal and Carrefour Intervention Suicide is a crisis hotline to prevent suicides and is run by volunteers who receive 32 hours of training.

Population:
Residents in the Montreal and Carrefour area experiencing a personal crisis

Description of Evaluation:
Trained observers listened to and coded calls in real time.

Summary of Evaluation Findings:
The study found an overall decrease in depressed mood from the beginning to the end of calls, but depression decreased in only 14% of calls and remained the same in 85% of calls. There was also a significant decrease in suicidal urgency from the beginning to the end of calls (urgency decreased in 27% of calls), especially for nonchronic callers. Contracts were made in 68% of calls, more frequently with chronic callers.

Sample Size:
n = 110 volunteers were observed helping 263 callers

Design:
Nonexperimental (no control group)

Reference:
Mishara and Daigle, 1997
A.2. Evaluation Studies of Crisis Hotlines

Description of Program:
1-800-SUICIDE is a national suicide prevention crisis hotline in the United States.

Population:
U.S. residents experiencing a personal crisis

Description of Evaluation:
Trained observers listened to and coded calls in real time.

Summary of Evaluation Findings:
Empathy, respect, supportive approach, good contact, and collaborative problem-solving were significantly related to positive outcomes at the end of the call, including the caller feeling better or happier, less confused, more resourceful, less helpless, and more confident. Active listening was not related to outcomes.

Sample Size:
n = 782 crisis intervention workers at 14 call centers

Design:
Nonexperimental (no control group)

Reference:
Mishara, Chagnon, and Daigle, 2007a
### A.2. Evaluation Studies of Crisis Hotlines

**Description of Program:**
A prevention program at the Los Angeles Suicide Prevention Center provided a 24-hour call hotline.

**Population:**
Residents of the Los Angeles County area in personal crisis

**Description of Evaluation:**
The study involved a comparison of suicide rates in Los Angeles County before and after the introduction of the suicide prevention service. Comparisons were also made with the suicide rates in other California counties. (One of the other three counties had a prevention program; two did not.)

**Summary of Evaluation Findings:**
Researchers did not find a decrease in the suicide rate in Los Angeles Country after implementation of the program but, rather, an increase. The suicide rate seemed to increase slightly with the rise in the number of calls.

**Sample Size:**
Sample sizes were not given. The total sample included all citizens of the four counties studied.

**Design:**
Two counties that did not have a suicide prevention program served as controls.

**Reference:**
Weiner, 1969
A.3. Evaluation Studies of Gatekeeper Training

Description of Program:
The school-based suicide prevention program Signs of Suicide (SOS) teaches high school students to respond to the signs of suicidal thoughts and behaviors as a mental health emergency.

Population:
High school students

Description of Evaluation:
The study evaluated the safety, efficacy, and feasibility of implementing the SOS program using data collected from 92 schools during the 2000–2001 school year.

Summary of Evaluation Findings:
The evaluation showed that there was a nearly 60% increase in help-seeking behavior among students following the training (help-seeking behavior being defined as seeking counseling for depression or suicidal ideation). The training program had excellent overall ratings on a questionnaire.

Sample Size:
n = 92 high schools

Design:
Quasi-experimental

Reference:
Aseltine, 2003
A.3. Evaluation Studies of Gatekeeper Training

Description of Program:
The Suicide Awareness Program (SAP) is a 90-minute instructional training course for community volunteers.

Population:
High school students

Description of Evaluation:
Data were collected two weeks prior to the training and then again at the completion of the training using the same questionnaire.

Summary of Evaluation Findings:
After the SAP training, participants had a mean score of 4.53 for their awareness of suicide warning signs, an increase from 3.97 before the training.

Sample Size:
n = 76

Design:
Quasi-experimental

Reference:
Tsai, Lin, Chang, Chang, et al., 2010
A.3. Evaluation Studies of Gatekeeper Training

Description of Program:

Teens Who Choose Life is an educational program that consists of a 15-minute filmstrip and a 40-minute follow-up discussion, including skill-building and intervention tools and a self-esteem handout. The program focuses on the relationship between mental illness and suicide and explores intervention and coping skills.

Population:

High school students

Description of Evaluation:

All participants were surveyed pretest and 30 days later (between these times, the test group participated in the program). The study compared two groups on the basis of changes in attitudinal responses.

Summary of Evaluation Findings:

In the baseline period, a sizable proportion of surveyed students did not hold “sensible or accurate” views of suicide. Participants in the program experienced a significant shift from undesirable to desirable responses in six of the eight tested areas.

Sample Size:

n = 324 students (203 program participants, 121 control group participants)

Design:

Quasi-experimental

Reference:

Ciffone, 1993
A.3. Evaluation Studies of Gatekeeper Training

Description of Program:

Question, Persuade, Refer is a one-hour community gatekeeper training program geared toward community members who may serve in gatekeeper roles for citizens in the community. The training is focused on suicide education and awareness, active listening, and role-play.

Population:

Community residents who may serve in gatekeeper roles (e.g., nurses, college resident advisors, school personnel)

Description of Evaluation:

The goals of the pilot study were to (1) study training outcomes, including skill development, from a brief gatekeeper training; (2) assess the feasibility of incorporating active learning principles (i.e., role-play practice) into standardized gatekeeper training; and (3) examine employee satisfaction with, and diffusion of information from gatekeeper training conducted in the workplace (e.g., sharing the training information with family/friends).

Summary of Evaluation Findings:

Pre- and postintervention analyses showed positive changes in participants’ knowledge about suicide and attitudes (self-efficacy) about intervening with suicidal individuals. A subset of participants engaged in role-play practice of gatekeeper skills after the training and rated the experience positively.

Sample Size:

n = 76 (n = 26 role-play participants; n = 30 agreed to a six-week follow-up)

Design:

Quasi-experimental

Reference:

Cross, Matthieu, et al., 2007
A.3. Evaluation Studies of Gatekeeper Training

<table>
<thead>
<tr>
<th>Description of Program:</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Youth Suicide Prevention school program is a three-year pilot program that trains high school students (four-hour training course), school personnel, and parents (90-minute seminar) to recognize suicidal behavior among young people and to respond in a manner that is most likely to prevent a fatal outcome.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Population:</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school students</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description of Evaluation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>The evaluation involved a questionnaire administered to a sample of 189 students in eight schools prior to exposure to the four-hour training course. A post-test using the same instrument was administered following exposure to the course to a comparable sample of 181 additional students from the same schools. The researchers obtained feedback evaluations from an additional sample of students, school staff, and parents who completed the course.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Summary of Evaluation Findings:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significant gains in the first two years of the program in understanding of youth suicide prevention techniques among high school students who took the course. School staff and parents taking the seminars indicated that they most appreciated receiving practical advice and resources that they could use in responding to a potentially suicidal young person.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Size:</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 189 students (181 at post-test); 390 program evaluators (218 students, 132 staff, 40 parents)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonexperimental (no control group)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson, 1987</td>
</tr>
</tbody>
</table>
A.3. Evaluation Studies of Gatekeeper Training

Description of Program:
The Curriculum-Based Suicide Prevention Program for Teenagers was a 1.5-hour program provided in a regular classroom by teachers who had previously received approximately six hours of training. It targeted a predominantly white suburban or rural student body and emphasized support networks in alleviating stress, confronting one’s peers, and taking advantage of community resources.

Population:
High school students

Description of Evaluation:
Students were assessed through a pretest and then again at a four- to six-week follow-up. The authors of this evaluation felt that four to six weeks was not sufficient to appropriately measure whether survey respondents were able to apply the knowledge learned.

Summary of Evaluation Findings:
The study failed to find convincing evidence of any program effect. There was also no evidence that the program had an effect on suicide attempt rates among exposed teens. Of the exposed group, 2.5% reported having made a (first) suicide attempt during the 18 months of follow-up, compared with 2.7% of the control group.

Sample Size:
Total n = 381; n = 174 received training and n = 207 in the control group who were available/willing after 18 months

Design:
Experimental

Reference:
Vieland et al., 1991
A.3. Evaluation Studies of Gatekeeper Training

**Description of Program:**

The Aboriginal Community Gatekeeper Training provided eight free, one-day gatekeeper training workshops to increase the ability of the local community to identify individuals at risk of suicide, mobilize local informal helping networks, and, where necessary, facilitate help-seeking behavior.

**Population:**

Members of the aboriginal community, i.e., health workers, youth workers, and education assistants

**Description of Evaluation:**

Paired sample t-tests were performed to assess whether there were changes in knowledge, intention to help, intention to refer to the mental health service, and confidence in identifying a suicidal person.

**Summary of Evaluation Findings:**

Evaluation of the workshops demonstrated an increase in participants’ knowledge about suicide, greater confidence in identifying people who are suicidal, and high levels of intention to provide help.

**Sample Size:**

n = 44 workshop attendees agreed to participate in the voluntary evaluation.

**Design:**

Quasi-experimental

**Reference:**

Capp et al., 2001
A.3. Evaluation Studies of Gatekeeper Training

Description of Program:
The Samaritans of New York public education suicide awareness and prevention program is designed to train lay and professional staff in effective suicide prevention practices and how to “befriend” a person in crisis. Two three-hour training sessions were conducted in 2005.

Population:
Community- and school-based staff

Description of Evaluation:
The study aimed to examine baseline and post-training differences among participants who attended the program regarding their knowledge about suicide and suicide prevention and their ability to intervene with individuals at risk for suicide. The study also explored the influence of previous exposure to suicidal individuals, as well as other predictors of gains subsequent to training.

Summary of Evaluation Findings:
Results indicated increased self-efficacy after suicide prevention training. Trainees with higher levels of education and previous contact with suicidal individuals were significantly more likely to indicate gains in self-efficacy after training.

Sample Size:
\( n = 365 \)

Design:
Quasi-experimental

Reference:
Clark et al., 2010
A.3. Evaluation Studies of Gatekeeper Training

Description of Program:
Raising Awareness of Personal Power (RAPP) is a curriculum-based suicide education program designed for high school and middle/junior high school students. RAPP educates students about depression and bipolar disorder, suicidal warning signs, a three-step process (i.e., Listen, Ask, and Take Action) for responding, and local resources. The program, generally presented by trained peer counselors and adult RAPP volunteers, includes a lecture and activities, such as interactive games, role-playing, and the analysis of several stories. The duration was one class period, which ranged from 50 to 75 minutes.

Population:
High school and middle/junior high students

Description of Evaluation:
The study used three methodological approaches: a rolling group design (RGD), an internal referencing strategy (IRS), and a minimum-competency approach (MC).

Summary of Evaluation Findings:
Findings for each approach were as follows:

1. The RGD approach found improvements in the knowledge, attitude, and efficacy variables that were significant for both the treatment and the control groups.
2. The IRS approach showed improvement on training materials, but there was little improvement on the untrained-yet-relevant materials.
3. The MC approach results showed that the RAPP program exceeded the minimum standards.

Sample Size:
\[ n = 33 \]

Design:
Quasi-experimental

Reference:
Cigularov et al., 2008
A.3. Evaluation Studies of Gatekeeper Training

Description of Program:
Question, Persuade, Refer (QPR) is a universal gatekeeper training program intended to improve school staff’s knowledge, appraisals (including willingness to assume a gatekeeper role), and self-reported suicide identification behaviors by students.

Population:
Staff in secondary schools

Description of Evaluation:
Staff from 20 middle schools and 12 high schools were randomized to either receive QPR or be assigned to the control group.

Summary of Evaluation Findings:
Intent-to-treat analyses showed that training increased school staff’s self-reported knowledge (effect size [ES] 0.41), appraisals of efficacy (ES 1.22), and service access (ES 1.07).

Sample Size:
102,000 students in the Cobb County (Marietta, Georgia) school district (10 middle and 6 high schools randomized to receive QPR and 10 middle and 6 high schools randomized to the control group)

Design:
Experimental (randomized control trial)

Reference:
Wyman, Brown, Inman, et al., 2008
A.4. Evaluation Studies of Marketing Campaigns

Description of Program:
The Air Force Suicide Prevention Campaign is a 30-minute briefing regularly required of all U.S. Air Force personnel that consists of 16 slides and four videos. It focuses on identifying warning signs, responding to those at risk, the impact on suicide survivors, and barriers to help-seeking behavior.

Population:
U.S. Air Force personnel

Description of Evaluation:
The study compared participants’ pre- and post-briefing surveys to assess the emotional impact of the briefing. Changes in the emotional status of suicidal and nonsuicidal subjects, as well as survivors and nonsurvivors, were also compared.

Summary of Evaluation Findings:
Suicidal female subjects showed a decrease in negative emotions. In general, there was a decrease in distress and an increase in calmness between pre- and post-briefing responses. There was no observed difference in the emotional changes between survivors and nonsurvivors.

Sample Size:
\[ n = 286 \text{ subjects (173 nonsurvivors, 113 survivors; 249 nonsuicidal, 32 suicidal)} \]

Design:
Quasi-experimental

Reference:
Bryan et al., 2009
A.4. Evaluation Studies of Marketing Campaigns

Description of Program:
Suicide Prevention Week is a yearly one-week media campaign that uses newspapers, radio, and television to change the behavior and attitudes of suicidal individuals and influence the public.

Population:
Men aged 20–40 in Quebec

Description of Evaluation:
Researchers surveyed a random sample of male Quebec residents regarding attitudes, knowledge, intentions, behaviors, and exposure to the campaign. Those who were not exposed to the campaign formed the control group. Data were also collected daily on the number of suicides and suicide attempts and the use of various resources.

Summary of Evaluation Findings:
Respondents who had exposure to Suicide Prevention Week demonstrated significantly more knowledge about suicide than those who had not, but no differences were observed in attitudes or help-seeking intentions. There was also no significant difference in observed behaviors, with the exception of suicide-related websites visited. No unintended negative effects of Suicide Prevention Week were observed.

Sample Size:
n = 1,020 respondents (190 test group, 830 control group)

Design:
Quasi-experimental

Reference:
Daigle et al., 2006
A.4. Evaluation Studies of Marketing Campaigns

Description of Program:

Youth Suicide: Recognizing the Signs is a video campaign intended to teach the signs of suicide and how to respond.

Population:

Parents of youth aged 10 or older

Description of Evaluation:

The evaluation included a pre- and post-test of knowledge of suicide, response to suicidality, perceptions of suicide, and intention to respond to suicidal individuals.

Summary of Evaluation Findings:

After the video participants’ knowledge, response, and intention to help improved.

Sample Size:

n = 112 parents

Design:

Nonexperimental (no control group)

Reference:

Maine, Shute, and Martin, 2001
A.5. Evaluation Study of Means Restriction

Description of Program:
Means restriction intervention removed all charcoal packs from open shelves of major retail outlets in the intervention region for 12 months.

Population:
Hong Kong residents

Description of Evaluation:
The evaluation compared charcoal-burning deaths between the intervention region and a nearby region where the means restriction intervention was not instituted. The evaluation was a double-blind trial (the removal of charcoal packs was not publicly announced, and public and frontline staff at supermarkets were not aware of the intervention). The control group was a region where charcoal packs were displayed as usual.

Summary of Evaluation Findings:
Suicide rates from charcoal burning decreased in the intervention region (p < 0.05) but not in the control region.

Sample Size:
Tuen Mun (intervention region) had 502,000 inhabitants. Yuen Long (control region) had 534,000 inhabitants.

Design:
Experimental, double-blind trial

Reference:
Yip et al., 2010
A.6. Evaluation Studies of Mental Health Interventions

Description of Program:

A 10-session cognitive therapy intervention program designed to prevent repeat suicide attempts among adults who recently attempted suicide.

Population:

Individuals who attempted suicide and who received a medical or psychiatric evaluation within 48 hours of the attempt

Description of Evaluation:

The evaluation used a randomized control trial design to assess the impact of the cognitive therapy as compared with enhanced usual care with tracking and referral services.

Summary of Evaluation Findings:

Cognitive therapy was effective in preventing suicide attempts among adults who recently attempted suicide. Participants in the therapy group had a significantly lower reattempt rate ($W_1^2 = 3.9; \ p = 0.049$) and were 50% less likely to reattempt suicide than participants in the usual care group (hazard ratio, 0.51; 95% confidence interval, 0.26–0.997). The severity of self-reported depression was significantly lower for the therapy group than for the usual care group at six months ($p = 0.02$), 12 months ($p = 0.009$), and 18 months ($p = 0.046$). The therapy group reported significantly less hopelessness than the usual care group at 6 months ($p = 0.045$).

Sample Size:

n = 120 individuals (60 to receive usual care; 60 to receive cognitive therapy and usual care)

Design:

Experimental (randomized control trial)

Reference:

G. Brown et al., 2005
A.6. Evaluation Studies of Mental Health Interventions

Description of Program:

Multisystemic therapy (MST) is a home-based model of service provision/treatment delivery grounded in social-ecological and systems theories and used to treat youth with mental health problems.

Population:

Youth referred for emergency psychiatric hospitalization (suicidal, homicidal, psychosis, or threat to self or others due to mental illness)

Description of Evaluation:

Multiple evaluation studies have been conducted on MST comparing it to hospitalization, by randomizing youth to receive one or the other.

Summary of Evaluation Findings:

The evaluations found that those who received MST exhibited

- Reductions in self-reported suicide attempts over the course of 16 months following recruitment (Huey et al., 2004)
- Higher levels of satisfaction with treatment (Huey et al., 2004)
- Reductions in externalizing symptoms and improvements in family functioning and school attendance (Henggeler et al., 1999; Rowland et al., 2005)

Also, 25% of youth randomized to MST were subsequently hospitalized during the two weeks following referral to MST (Schoenwald et al., 2000).

Sample Size:

Henggeler et al., 1999; Huey et al., 2004; and Schoenwald et al., 2000: n = 113 patients (57 MST, 56 hospitalization)
Rowland et al., 2005: n = 31 patients (15 MST, 16 usual services)

Design:

Henggeler et al., 1999; Huey et al., 2004; and Schoenwald et al., 2000: Control group received hospitalization.
Rowland et al., 2005: Control group received usual services.

Reference:

Henggeler et al., 1999; Huey et al., 2004; Rowland et al., 2005; Schoenwald et al., 2000
A.6. Evaluation Studies of Mental Health Interventions

Description of Program:
RUSH is a ten-week treatment program based on dialectical behavior therapy that facilitates and encourages self-help and life enhancement skills through twice-weekly group skills training.

Population:
Offenders exhibiting borderline personality disorder characteristics

Description of Evaluation:
A pre-/post-test design evaluated participants’ level of stress, anxiety, and depression. Observational evaluations were made throughout the program.

Summary of Evaluation Findings:
The majority of participants experienced a decline in symptomology between the beginning and end of treatment—especially on the stress subscale. High levels of group cohesion, motivation, and commitment were observed.

Sample Size:
\( n = 29 \) patients

Design:
Nonexperimental (no control group)

Reference:
Eccleston and Sorbello, 2002
A.6. Evaluation Studies of Mental Health Interventions

Description of Program:
Problem-solving therapy is a short-term, secondary prevention approach designed to decrease the intensity of suicidal ideation and associated psychopathology by bolstering problem-solving abilities and coping skills.

Population:
Participants were selected from the Psychology 100 Research Experience Program at Ohio State University and local campus and community service providers.

Description of Evaluation:
The evaluation assessed whether the intervention increased the use of positive coping mechanisms and decreased suicidal ideation, depression, and hopelessness.

Summary of Evaluation Findings:
Using a results classification system (maintained, improved, or stayed the same), the pattern of results indicated little substantial shifting at post-treatment (83% of control vs. 81% of treatment maintained; 13% vs. 13%, improved; and 4% vs. 6% worsened).

Sample Size:
\[ n = 110 \]

Design:
Those in the treatment group watched a 35-minute video focused on problem-solving and coping styles.

Reference:
Fitzpatrick, Witte, and Schmidt, 2005
A.6. Evaluation Studies of Mental Health Interventions

Description of Program:

The Suicide Prevention Program is a “mandated assessment” response to suicide threats and attempts that requires students at the University of Illinois to receive four sessions of professional assessment immediately following a suicide incident. The first of these weekly sessions is held within a week of the incident.

Population:
College students

Description of Evaluation:

The study compared the suicide rate among students at the University of Illinois during the program years to the rate before the program’s initiation. It also compared this trend to contemporary data from peer institutions that did not use the program.

Summary of Evaluation Findings:

The university observed a 45.3% reduction in the rate of suicide between the pre-intervention and intervention years. Furthermore, rates at the university were declining in the same years when the rates at peer institutions were steady or increasing. However, the program was not effective in reducing the rate of suicide among graduate students.

Sample Size:

One-group comparison: 1,041,951 students (274,942 control; 767,009 intervention);
Two-group comparison: 2,057,780 students (249,812 intervention; 1,807,968 control)

Design:
Quasi-experimental

Reference:
Joffe, 2008
A.6. Evaluation Studies of Mental Health Interventions

Description of Program:
LifeSPAN therapy is a cognitive therapy approach provided through the Early Psychosis Prevention and Intervention Centre and includes eight to ten individual sessions.

Population:
Young people (aged 15–29)

Description of Evaluation:
The study evaluated the effectiveness of the LifeSPAN therapy through a randomized controlled trial.

Summary of Evaluation Findings:
The evaluation found that standard care reduces suicidality, and adding LifeSPAN therapy leads to an improvement in cognitive factors linked to suicide risk (e.g., hopelessness).

Sample Size:
n = 56 were randomized to either the

treatment group (n = 31) or a control group (n = 25). Thirteen of the 56 dropped out during the next ten weeks, and one dropped out before the six-month follow-up. This left 21 patients in each group (treatment and control) at the end of ten weeks.

Design:
Treatment group (n = 21) (LifeSPAN therapy plus standard clinical care) and control group (n = 21) (standard clinical care)

Reference:
Power et al., 2003
Description of Program:
Dialectical behavior therapy (DBT) is a one-year behaviorally oriented outpatient psychotherapy approach for bipolar disorder patients with the goals of reducing self-injurious and life-threatening behaviors, reducing therapy-interfering behaviors, and reducing quality-of-life-interfering behaviors. It consists of weekly individual and group therapy sessions that include skill training and contingency management.

Population:
Suicide attempters meeting the criteria for borderline personality disorder

Description of Evaluation:
Multiple evaluation studies have been conducted comparing DBT to “treatment as usual” for parasuicidal patients with bipolar disorder and personality disorder over the course of a year. Measures from these studies include the number of parasuicidal acts, maintenance in therapy, amount of psychiatric inpatient treatment, depression, hopelessness, reasons for living, anger, social adjustment, work performance, and suicide ideation. Turner (2000) examined the comparative effectiveness of DBT and client-centered therapy (CCT).

Summary of Evaluation Findings:
The evaluations found that those who received DBT exhibited

- Reductions in the frequency and medical risk of parasuicidal behavior (Linehan, Armstrong, et al. 1991; Linehan, Heard, and Armstrong, 1993; Bohus et al., 2004)
- Lower attrition rates compared with the control group (Linehan, Armstrong, et al., 1991)
- Fewer days of inpatient hospitalization (Linehan, Armstrong, et al., 1991; Linehan, Heard, and Armstrong, 1993; Linehan, Comtois, et al., 2006; Bohus et al., 2004) compared with the control group
- Less anger than the control group (Linehan, Heard, and Armstrong, 1993; Linehan, Tutek, et al., 1994); however, this was not replicated in Bohus et al., 2004
- Better self-reported social adjustment (Linehan, Heard, et al., 1993; Linehan, Tutek, et al., 1994)
- Higher scores on the Global Assessment Scale (Linehan, Heard, et al., 1993; Linehan, Tutek, et al., 1994)
- Better work performance than the control group (Linehan, Heard, et al., 1993)
### Appendix A: Evaluation Studies

#### A.6. Evaluation Studies of Mental Health Interventions

**Summary of Evaluation Findings (cont.):**

- Fewer self-mutilating behaviors and self-damaging, impulsive behavior than the control group (Bohus et al., 2004); however, Verheul et al. (2003) found that DBT is more effective for patients with high severity of self-mutilating acts than for low-severity patients.

- Higher retention in therapy (Verheul et al., 2003).

There was mixed evidence as to whether DBT reduced depression, hopelessness, suicide ideation, and inability to find reasons for living. A 1991 study by Linehan, Armstrong, et al., found that DBT did not have any significant effects on these outcomes. However, a later study by Linehan, Comtois, et al. (2006) found that DBT significantly reduced suicide ideation and depression and improved reasons for living among patients. Turner (2000) found that DBT subjects showed more improvement than client-centered therapy subjects at both six and 12 months on their parasuicide rating, Brief Symptom Inventory, number of suicide attempts, and hospitalization days. At 12 months, DBT subjects showed more improvement in impulsiveness, anger, and depression. There was no difference in anxiety between treatments.

**Sample Size:**

- Linehan, Armstrong, et al., 1991: n = 44 subjects (22 DBT, 22 control)
- Linehan, Heard, et al., 1993: n = 39 patients (19 DBT, 20 control) for PHI; 20 of these (9 DBT, 11 control) for all other assessments
- Linehan, Tutek, et al., 1994: n = 39 patients (19 DBT, 20 control) for the Parasuicide History Interview; 20 of these (nine DBT, 11 control) for all other assessments
- Linehan, Comtois, et al., 2006: n = 101 patients (52 DBT and 49 community treatment by experts)
- Turner, 2000: n = 24 patients (12 DBT, 12 CCT)
- Verheul et al., 2003: n = 58 patients (27 DBT, 31 treatment as usual)
- Bohus et al., 2004: n = 50 patients (31 DBT, 19 waiting list/usual treatment)

**Design:**

- Bohus et al., 2004: Quasi-experimental
- All other studies: Experimental

**Reference:**

- Linehan, Armstrong, et al., 1991; Linehan, Heard, et al., 1993; Linehan, Tutek, et al., 1994; Linehan, Comtois, et al., 2006; Turner, 2000; Verheul et al., 2003; Bohus et al., 2004
A.7. Evaluation Studies of Provider Training

Description of Program:
A nurse education program for nurses working in general hospitals aimed to increase nurses’ knowledge of suicide prevention, promote the adoption of a positive attitude by nurses in caring for patients at risk of suicide and their families, and enhance nurses’ competence in suicide prevention.

Population:
Registered general nurses

Description of Evaluation:
The evaluation used a mixed-methods design relying on a single group (convenience sample) for pretest and post-test analysis and focus group interviews.

Summary of Evaluation Findings:
Statistically significant positive changes were found between the pre- and post-test measures of participants’ attitudes and competence levels. Qualitative data showed that participants had applied the new knowledge in clinical practice.

Sample Size:
n = 54 nurses (18 in focus groups)

Design:
Nonexperimental (no control group)

Reference:
Chan, Chien, and Tso, 2009a, 2009b
A.7. Evaluation Studies of Provider Training

Description of Program:
STORM is a one- to two-day training program conducted by mental health nurses that includes lectures, group discussions, role-playing and video feedback in the areas of assessment, crisis management, problem-solving, and crisis intervention.

Population:
Mental health care providers

Description of Evaluation:
A pre-/post-test design evaluated participants acting as their own controls using questionnaires and interviews to measure changes in attitude, confidence, skills, competence, and satisfaction.

Summary of Evaluation Findings:
The training resulted in significant positive changes in attitudes and confidence, but the study found no statistically significant changes in skill acquisition.

Sample Size:
n = 458 (baseline), 458 (immediately post-training), 143 (four-month post-training)

Design:
Nonexperimental (one group pre-/post-test)

Reference:
Gask et al., 2006
A.7. Evaluation Studies of Provider Training

Description of Program:
CALM is a two-hour workshop for mental health care providers to learn how reducing access to lethal means can prevent suicide. The training includes presentations, a video screening, instruction, discussion, and role-playing in addition to written materials.

Population:
Mental health care providers

Description of Evaluation:
Participants attended CALM training and immediately answered a questionnaire comparing their pre-CALM and post-CALM attitudes and beliefs. This post-test was compared to the follow-up responses of the same group six to eight weeks later.

Summary of Evaluation Findings:
Participating in CALM led to a significant increase in commitment and skills for counseling using means reduction. These beliefs and attitudes were sustained in the weeks following the program. However, there was a significant decrease in perceptions of the effectiveness of means reduction between post-test and follow-up.

Sample Size:
168 (post-test); 111 (follow-up)

Design:
Nonexperimental (one group pre-/post-test)

Reference:
Johnson et al., 2011
A.7. Evaluation Studies of Provider Training

Description of Program:
The Gotland Educational Program was directed at all general practitioners on the island of Gotland, Sweden, and consisted of two two-day sessions of oral and written information, group work, case reports, and discussions focused on suicide prevention.

Population:
General practitioners

Description of Evaluation:
The study compared baseline measures, prior to the program, with trends in suicide rates, prescription rates, sick leave, inpatient care, and general practitioner attitudes over the next five years on Gotland.

Summary of Evaluation Findings:
General practitioners reported an increase in competency, knowledge, and support of comprehensive treatment. Additionally, the suicide rate, the number of psychiatric referrals, and depression-related sick leave, and inpatient care decreased on Gotland following the program. Lithium and antidepressant prescriptions also increased.

Sample Size:
Not reported

Design:
Nonexperimental (one group pre-/post-test)

Reference:
Rutz, 2001
A.7. Evaluation Studies of Provider Training

Description of Program:
The suicide prevention education program in Slovenia is a four-hour educational program for primary care physicians focused on recognition and treatment of depression and suicide prevention.

Population:
Primary care doctors who had written at least five prescriptions of antidepressants at the baseline period

Description of Evaluation:
The study was conducted in three neighboring regions in Slovenia (Celje, Ravne na Koroškem, and Podravska). The study compared prescription rates in the regions for antidepressants and anxiolytics before and after the intervention. The study also compared suicide rates three years before and after the intervention.

Summary of Evaluation Findings:
The training program was beneficial for primary care physicians’ ability to recognize and manage depression. However, there was no significant decrease in local suicide rates.

Sample Size:
Physicians from Celje and Ravne na Koroškem (n = 82; educational group) attended the program, and the other half (n = 108; control group 1) and physicians from the Podravska region (n = 164; control group 2) did not attend the program.

Design:
The intervention group received the training; the control group did not.

Reference:
Roskar et al., 2010
A.8. Evaluation Studies of Screening Programs

Description of Program:
A participatory screening program included (1) public health education from 1991 to 2000 and (2) screening for depression with follow-up from 1991 to 1997, using public health and primary care resources in the town with the assistance of the psychiatric care resources in the neighboring county.

Population:
Elderly residents (65 years and older)

Description of Evaluation:
Quasi-experimental design with intervention and reference municipalities were used to estimate changes in the risk of dying by suicide before and after the ten-year implementation.

Summary of Evaluation Findings:
The risk for women in the intervention area was reduced by 64%, whereas there was no significant change in the risk for men in the intervention area or for either men or women in the reference area. The risk reduction for women in the intervention area was greater than the secular trend.

Sample Size:
Oyama et al., 2006: approx. 1,400 (426 residents in the intervention town, 1,000 in the reference town)
Scott et al., 2009: n = 1,729 students completed a questionnaire; 356 who screened positive and 285 who screened negative completed the Diagnostic Interview Schedule for Children

Design:
Oyama et al., 2006: Quasi-experimental
Scott et al., 2009: A sample of those who screened negative also completed the Diagnostic Interview Schedule for Children

Reference:
Oyama et al., 2006; Scott et al., 2009
A.8. Evaluation Studies of Screening Programs

Description of Program:
The Columbia Health Screen (CHS) and Post-Screening Structured Interview is a 14-item questionnaire that includes questions on reporting needing help with an emotional problem. Youth who screened positive on the CHS were referred for further mental health screening.

Population:
High school students

Description of Evaluation:
The study randomized classrooms to receive the screening or not and then assessed differences in referrals to mental health services among those identified via screening versus those identified via usual process (e.g., parents, teachers).

Summary of Evaluation Findings:
Screened students were 21 times more likely to receive a referral for mental health services.

Sample Size:
Husky, McGuire, et al., 2009: n = 483 students offered screening
Husky, Miller, et al., 2011: n = 890 (467 received screening, 423 in control group)

Design:
Husky, McGuire, et al., 2009: Nonexperimental (no control group)
Husky, Miller, et al., 2011: Classrooms were randomized to screening condition or control

Reference:
Husky, McGuire, et al., 2009; Husky, Miller, et al., 2011
A.8. Evaluation Studies of Screening Programs

Description of Program:
The Columbia Suicide Screen (CSS) is an 11-item questionnaire embedded in a health survey administered during regular class time.

Population:
High school students

Description of Evaluation:
Several evaluation studies have assessed the CSS. In all studies, youth who screened positive and a control group were reassessed with the Diagnostic Interview Schedule for Children, and school staff were also asked to assess all youths who completed the instrument for emotional problems. Scott et al. (2010) and Shaffer, Scott, et al. (2004) also tested optimal positive-screen algorithm using sensitivity, specificity, positive-predictive value.

Summary of Evaluation Findings:
34% of those with mental health problems were identified by screening only, 13% were identified by school officials, and 35% were identified by both. Of those, 18% with mental health problems were not identified by either (Scott et al., 2009). Three screening algorithms were also tested and identified 96%, 92%, and 89% of youth with mental health problems (corresponding positive screens are 35%, 24%, 17%; Scott et al., 2010). The best algorithm relied on suicide ideation or previous attempt and a score greater than or equal to 3 for the screening questions about whether they were unhappy, withdrawing, irritable, and anxious (Shaffer, Scott, et al., 2004).

Sample Size:
Scott et al., 2009, 2010: n = 1,729 students completed the questionnaire; 356 who screened positive and 285 who screened negative completed the DISC
Shaffer, Scott, et al., 2004: n = 1,729 students completed questionnaire; 489 who screened positive and 356 who screened negative completed the DISC

Design:
Scott et al., 2009, 2010: A sample of those who screened negative also completed the Diagnostic Interview Schedule for Children
Shaffer, Scott, et al., 2004: A matched sample of those who screened negative also completed the DISC

Reference:
Scott et al., 2009, 2010; Shaffer, Scott, et al., 2004
A.8. Evaluation Studies of Screening Programs

Description of Program:
Diagnostic Predictive Scales–8 (DPS-8) is an 84-item computerized, voluntary, self-report assessment offered to youth at no cost. The DPS-8 screens for mental health problems, including suicide.

Population:
Adolescents seen in a pediatric health center

Description of Evaluation:
Clinicians reviewed the results of the screening at a visit, and the study assessed whether youth who screened positive received pediatric mental health care or were referred to specialty mental health care.

Summary of Evaluation Findings:
Fourteen percent of youth who completed the scale screened positive for one or more mental disorders. Screening was associated with an increased likelihood of receiving pediatric mental health care or referral for specialty mental health care.

Sample Size:
n = 483 patients offered screening

Design:
Nonexperimental (no control group)

Reference:
Husky, Kaplan, et al., 2011
A.8. Evaluation Studies of Screening Programs

Description of Program:
The Universal Youth Suicide Screening program screens high school students for mental health problems and suicide risk.

Population:
High school students

Description of Evaluation:
Classes at six high schools were randomized to either the experimental or control group. The experimental group was surveyed on mood, depression, drug use, suicide ideation, suicide history, while the control group was surveyed on just mood, depression, and drug use (no suicide questions). Two days later, both groups were reassessed on mood, depression, suicide history, and ideation.

Summary of Evaluation Findings:
The two groups did not differ significantly in distress levels or suicidal ideation. Furthermore, there was no evidence of an iatrogenic effect of asking about suicide, even among high-risk subjects.

Sample Size:
n = 2,342 students (1,172 experimental, 1,170 control)

Design:
Experimental

Reference:
Gould, Marrocco, et al., 2005

Description of Program:
The Personal Growth Class is a semester-long class involving a small-group work component that focuses on social support, weekly monitoring of activities targeting changes in mood management, school performance and attendance, drug involvement, and life skills training in self-esteem enhancement, decisionmaking, personal control, and interpersonal communication.

Population:
High school students

Description of Evaluation:
The evaluation compared youth in the class to a random sample of youth not at risk for school failure over time.

Summary of Evaluation Findings:
The intervention resulted in improvements in outcomes (suicide risk behaviors, depression, hopelessness, stress, anger, self-esteem, and social network support) among the intervention group as compared with the control group. However, no differences in personal control were detected.

Sample Size:
Not reported

Design:
There was both an assessment-only condition and a random sample of youth not at risk for school failure to serve as a normative comparison group.

Reference:
Eggert et al., 1995

Description of Program:
Surviving the Teens Suicide Prevention and Depression Awareness Program is a four-session classroom intervention designed to increase knowledge, decrease the stigma of depression and other mental health disorders, improve coping, increase help-seeking, increase the likelihood of depressed teens seeking help, increase family and school connectedness, and decrease suicidal and other risk-taking behaviors.

Population:
High school students

Description of Evaluation:
A pre- and post-tests were used to examine the effect of the intervention on suicide ideation, suicidal behavior, self-efficacy, and behavioral intent regarding help-seeking.

Summary of Evaluation Findings:
The evaluation found that youth participants exhibited reductions in considering suicide, making a suicide plan, and attempting suicide, as well as increases in self-efficacy and behavioral intentions toward help-seeking.

Sample Size:
n = 1,030 students

Design:
Nonexperimental (no control group)

Reference:
K. King, Strunk, and Sorter, 2011

Description of Program:
A problem-solving intervention was designed to improve coping and problem-solving to decrease suicidal ideation and behavior. The program included an experiential-affective group, psycho-educational classes, and an extended problem-solving group. Treatment was group-based and outpatient but administered in a partial or day hospital venue.

Population:
Suicidal adolescents and young adults

Description of Evaluation:
Suicidal patients were randomly assigned to the experimental group vs. treatment as usual. Treatment as usual involved a combination of inpatient and outpatient care.

Summary of Evaluation Findings:
The evaluation found no group differences at one, six, or 12 months. Both groups improved over time in terms of symptoms, problem-solving, suicidality, and other outcomes. However, the experimental group had less attrition than control group.

Sample Size:
\( n = 264 \) (143 experimental, 121 control)

Design:
Experimental with random assignment

Reference:
Rudd, Rajab, and Orman, 1996

Description of Program:
Reaching Young Europe was based on the Reaching Young People program and consisted of 25 weekly one-hour sessions focused on increasing coping skills and resilience.

Population:
Young children (kindergarten through second grade)

Description of Evaluation:
The evaluation assessed the program’s implementation through weekly reports from program coordinators and program facilitators’ reports of participation and appreciation after each session. The evaluation of program effects assessed changes in social skills, including assertion and empathy, as well as changes in coping skills.

Summary of Evaluation Findings:
According to the authors, the results showed that the program could be successfully implemented, and some activities were appreciated by certain age groups more than others. They also reported that the evaluation showed significant improvement in overall social skills (teacher-reported) among first graders, assertion among second graders, and empathy among kindergarteners and first graders. The program had no significant effects on coping skills.

Sample Size:
Evaluation of program implementation: two kindergarten groups, one after-school club, and 12 grade school classes. Evaluation of program effects: 214 participating children and 109 control children.

Design:
Implementation evaluation: no control group.
Evaluation of effects: control group, not randomized

Reference:
Mishara and Ystgaard, 2000

Description of Program:
The program involved sending letters to patients who refused continued interaction with the health care system. Letters were sent monthly for four months, then every two months for eight months, and finally every three months for four years—a total of five years and 24 contacts. Content of the letters was personalized but always contained the sentiment that “It has been some time since you were here at the hospital. . . . If you wish to drop us a note we would be glad to hear from you.”

Population:
Patients who had refused ongoing care after being hospitalized because of a depressive or suicidal state

Description of Evaluation:
The study assessed differences in suicide rates among 3,005 patients five and 15 years after hospital discharge. The study assessed differences by whether patients accepted or declined ongoing treatment and, among those who refused treatment, whether they were periodically contacted by letter.

Summary of Evaluation Findings:
Five years after initial discharge from the hospital, patients in the group that refused treatment and then were contacted by letter had a lower suicide rate (3.9%) than patients who chose continued treatment (6.2%), and patients who refused treatment and were not contacted by letter (4.6%). Fifteen years after initial discharge from the hospital, patients who refused treatment and were not contacted by letter had the lowest suicide rates (5.7%), followed by patients who refused treatment and were contacted by letter (6.4%), and patients who chose to continue treatment (8.2%).

Sample Size:
3,005 individuals were divided into three groups: those who continued to seek treatment (n = 1,939); those who refused treatment and were randomized to receive letters (n = 389); and those who refused treatment and were randomized into the control group (i.e., did not receive letters; n = 454). An additional 223 individuals were not included in the study because it was unclear whether they were seeking treatment.

Design:
Experimental (randomized control trial)

Reference:
Motto and Bostrom, 2001
Appendix B
Glossary of Terms

Evidence-based program
A program that has been determined to be effective through rigorous scientific evaluations and randomized controlled trials, has a significant and sustained effect on intended outcomes, and has been tested using large longitudinal studies or multiple replications (Evidence-Based Prevention and Intervention Support Center, 2011).

Fidelity
Adherence of implementation to a program’s original design (Smith, Daunic, and Taylor, 2007).

Logic model
A graphical depiction of the rationale and expectations of a program (Leviton et al., 2010). A logic model clarifies the causal relationships among program resources, activities, and outcomes (McLaughlin and Jordan, 1999; Wholey, Hatry, and Newcomer, 2010).

Outcomes
Changes or benefits resulting from activities and outputs. Typically, programs have short-, intermediate-, and long-term outcomes (Leviton, et al., 2010; Wholey, Hatry, and Newcomer, 2010).

Outcome evaluation
An assessment of how well the program’s activities or services have brought about expected changes in the target population or social condition (Rossi, Lipsey, and Freeman, 2004).

Outputs
The products, goods, and services provided to the program’s participants (Wholey, Hatry, and Newcomer, 2010).

Process evaluation
A form of program evaluation designed to document and analyze the early development and actual implementation of a program, assessing whether and how well services are delivered as intended or planned. Also known as implementation assessment (Wholey, Hatry, and Newcomer, 2010; Rossi, Lipsey, and Freeman, 2004).
Program
A set of activities, tied together through shared resources (e.g., staff, funding, space, materials), meant to influence a targeted population’s knowledge, attitudes, or behavior to accomplish a specific goal or goals.

Theory of change
A model that describes the mechanisms through which the initiative’s inputs and activities are thought to lead to desired outcomes (Leviton et al., 2010).


———, “Intervening with Practitioners to Improve the Quality of Prevention: One Year Findings from a Randomized Trial of Assets–Getting To Outcomes,” *Journal of Primary Prevention*, forthcoming.


Evidence-Based Prevention and Intervention Support Center, Evidence-Based Intervention Reports, University Park, Pa., July 2011. As of March 19, 2013: http://www.episcenter.psu.edu/EBIReports


Linehan, Marsha M., unpublished work on multisystemic therapy, 1981.


O’Carroll, Patrick W., Lloyd B. Potter, and James A. Mercy, “Programs for the Prevention of Suicide Among Adolescents and Young Adults,” Morbidity and Mortality Weekly Report, Vol. 43, No. RR-6, April 22, 1994, pp. 1–7.


World Health Organization, Western Pacific Region, Towards Evidenced-Based Suicide Prevention Programmes, 2010. As of March 19, 2013: http://www.wpro.who.int/mnh/TowardsEvidencebasedSPP.pdf


Evaluating suicide prevention programs can be challenging because suicide is a rare event, data on suicides often lag by several years, and programs tend to have multiple components, making it difficult to discern which characteristics contributed to a given outcome. The RAND Suicide Prevention Program Evaluation Toolkit was designed to help program staff overcome these common challenges to evaluating and planning improvements to their programs. It begins by walking users through the process of developing a program logic model that ties program activities to intermediate outcomes, helping staff better understand the drivers of any changes in long-term outcomes, such as suicide rates. It then offers information about the latest evaluation research, helps users design an evaluation that is appropriate for their program type and available resources and expertise, supports the selection of measures for new evaluations and to augment or enhance ongoing evaluations, and offers basic guidance on how to analyze and use evaluation data for program improvement. Through checklists, worksheets, and templates, the toolkit takes users step by step through the process of identifying whether their programs produce beneficial effects, ultimately informing the responsible allocation of scarce resources. The toolkit’s design and content are the result of a rigorous, systematic review of the program evaluation literature to identify evaluation approaches, measures, and tools used elsewhere and will be particularly useful to coordinators and directors of suicide prevention programs in the U.S. Department of Defense, Veterans Health Administration, community-based settings, and state and local health departments. A companion report, Development and Pilot Test of the RAND Suicide Prevention Program Evaluation Toolkit, offers additional background on the toolkit’s design and refinement.