

FOURTH EDITION

Health Program Planning and Evaluation

**A Practical, Systematic Approach
for Community Health**

L. Michele Issel

Rebecca Wells

FOURTH EDITION

Health Program Planning and Evaluation

**A Practical, Systematic Approach
for Community Health**

L. Michele Issel, PhD, RN

Professor of PhD Program
University of North Carolina at Charlotte
College of Health and Human Services
Charlotte, North Carolina

Rebecca Wells, PhD, MHSA

Professor
The University of Texas
School of Public Health
Houston, Texas



JONES & BARTLETT
LEARNING



World Headquarters
Jones & Bartlett Learning
5 Wall Street
Burlington, MA 01803
978-443-5000
info@jblearning.com
www.jblearning.com

Jones & Bartlett Learning books and products are available through most bookstores and online booksellers. To contact Jones & Bartlett Learning directly, call 800-832-0034, fax 978-443-8000, or visit our website, www.jblearning.com.

Substantial discounts on bulk quantities of Jones & Bartlett Learning publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones & Bartlett Learning via the above contact information or send an email to specialsales@jblearning.com.

Copyright © 2018 by Jones & Bartlett Learning, LLC, an Ascend Learning Company

All rights reserved. No part of the material protected by this copyright may be reproduced or utilized in any form, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the copyright owner.

The content, statements, views, and opinions herein are the sole expression of the respective authors and not that of Jones & Bartlett Learning, LLC. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not constitute or imply its endorsement or recommendation by Jones & Bartlett Learning, LLC and such reference shall not be used for advertising or product endorsement purposes. All trademarks displayed are the trademarks of the parties noted herein. *Health Program Planning and Evaluation: A Practical, Systematic Approach for Community Health, Fourth Edition* is an independent publication and has not been authorized, sponsored, or otherwise approved by the owners of the trademarks or service marks referenced in this product.

There may be images in this book that feature models; these models do not necessarily endorse, represent, or participate in the activities represented in the images. Any screenshots in this product are for educational and instructive purposes only. Any individuals and scenarios featured in the case studies throughout this product may be real or fictitious, but are used for instructional purposes only.

This publication is designed to provide accurate and authoritative information in regard to the Subject Matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the service of a competent professional person should be sought.

15842-7

Production Credits

VP, Executive Publisher: David D. Cella
Publisher: Michael Brown
Associate Editor: Danielle Bessette
Vendor Manager: Nora Menzi
Senior Marketing Manager: Sophie Fleck Teague
Manufacturing and Inventory Control Supervisor: Amy Bacus
Composition and Project Management: S4Carlisle Publishing Services

Cover Design: Scott Moden
Director of Rights & Media: Joanna Gallant
Rights & Media Specialist: Merideth Tumasz
Media Development Editor: Shannon Sheehan
Cover Image: © Lynne Nicholson/Shutterstock
Printing and Binding: Edwards Brothers Malloy
Cover Printing: Edwards Brothers Malloy

Library of Congress Cataloging-in-Publication Data

Names: Issel, L. Michele, author. | Wells, Rebecca, 1966- author.
Title: Health program planning and evaluation: a practical, systematic approach for community health/L. Michele Issel and Rebecca Wells.
Description: Fourth edition. | Burlington, MA: Jones & Bartlett Learning, [2018] | Includes bibliographical references and index.
Identifiers: LCCN 2017010386 | ISBN 9781284112115 (pbk.)
Subjects: | MESH: Community Health Services—organization & administration | Program Development—methods | Health Planning—methods | Program Evaluation—methods | United States
Classification: LCC RA394.9 | NLM WA 546 AA1 | DDC 362.12068—dc23 LC record available at <https://lccn.loc.gov/2017010386>

6048

Contents

List of Figures	xi
List of Tables	xiii
List of Exhibits	xvii
Preface to the Fourth Edition	xix
Acknowledgments	xxv
List of Acronyms	xxvii

SECTION I **The Context of Health Program Development** **1**

Chapter 1 Context of Health Program Development and Evaluation	3
History and Context	4
Concept of Health	4
Health Programs, Projects, and Services	4
History of Health Program Planning and Evaluation	5
Evaluation as a Profession	8
Who Does Planning and Evaluations?	10
Roles of Evaluators	10
Planning and Evaluation Cycle	11
Interdependent and Cyclic Nature of Planning and Evaluation	11
Using Evaluation Results as the Cyclical Link	13
Program Life Cycle	13
The Fuzzy Aspects of Planning	14
Paradoxes	14
Assumptions	16
Uncertainty, Ambiguity, Risk, and Control	17

Introduction to the Types of Evaluation	19
Mandated and Voluntary Evaluations	20
When Not to Evaluate	21
The Public Health Pyramid	21
Use of the Public Health Pyramid in Program Planning and Evaluation	23
The Public Health Pyramid as an Ecological Model	23
The Town of Layetteville in Bowe County	25
Across the Pyramid	25
Discussion Questions	27
Internet Resources	27
References	27

Chapter 2 Relevance of Diversity and Disparities to Health Programs	29
Health Disparities	30
Diversity and Health Disparities	32
Diversity and Health Programs	33
Measurement	33
Interventions	38
Influences of Sociocultural Diversity on Interventions	38
Influences of Biological Diversity on Interventions	39
Approaches to Developing Programs	39
Profession and Provider Diversity	40
The Three Health Provider Sectors	43
Diversity Within Healthcare Organizations and Programs	43
Organizational Culture	44
Cultural Competency Continuum	44
Enhancing Cultural Competency	48

Stakeholders and Coalitions50
 Across the Pyramid51
 Discussion Questions53
 Internet Resources.....53
 References54

SECTION II Defining the Health Problem 57

Chapter 3 Community Health Assessment for Program Planning 59

Defining Community59
 Community as Context and Intended Recipient60
 Defining Terms: Based, Focused, and Driven.....61
 Types of Needs62
 Types of Strengths.....63
 Approaches to Planning64
 Incremental Approach.....64
 Apolitical Approach66
 Advocacy Approach66
 Communication Action Approach67
 Comprehensive Rational Approach67
 Strategic Planning Approach68
 Summary of Approaches69
 Models for Planning Public Health Programs69
 Mobilizing for Action through Planning and Partnership (MAPP)70
 Community Health Improvement Process (CHIP).....70
 Protocol for Assessing Community Excellence in Environmental Health (PACE-EH).....70
 In Summary70
 Perspectives on Assessment71
 Epidemiological Perspective72
 Public Health Perspective.....74
 Social Perspective74
 Asset Perspective.....74
 Rapid Perspective75

Types of Assessments.....75
 Organizational Assessment75
 Marketing Assessment.....76
 Needs Assessment76
 Community Health Assessment.....77
 Workforce Assessment.....77
 Steps in Planning and Conducting the Assessment77
 Form and Develop the Team.....78
 Create a Vision.....79
 Involve Community Members79
 Define the Population80
 Define the Problem to Be Assessed81
 Investigate81
 Prioritize.....82
 Make a Decision.....82
 Implement and Continue.....83
 Anticipate Data-Related and Methodological Issues83
 Across the Pyramid85
 Discussion Questions.....85
 Internet Resources.....86
 References87

Chapter 4 Characterizing and Defining the Health Problem 91

Collecting Data From Multiple Sources.....91
 Public Data91
 Primary Data92
 Observational Data92
 Archival Data93
 Proprietary Data93
 Published Literature93
 Data Beyond Street Lamp.....93
 Collecting Descriptive Data94
 Magnitude of the Problem.....94
 Dynamics Leading to the Problem.....94
 Population Characteristics96
 Attitudes and Behaviors96
 Years of Life and Quality of Life.....96

Statistics for Describing Health Problems99

 Descriptive Statistics..... 100

 Geographic Information Systems: Mapping..... 101

 Small Numbers and Small Areas 101

 Epidemiology Rates 102

Stating the Health Problem 102

 Diagramming the Health Problem 102

 Writing a Causal Theory of the Health Problem 108

Prioritizing Health Problems 110

 Nominal Group Technique..... 111

 Basic Priority Rating System..... 111

 Propriety, Economics, Acceptability, Resources, and Legality (PEARL) Component..... 113

 Prioritizing Based on Importance and Changeability 114

Across the Pyramid 115

Discussion Questions and Activities..... 117

Internet Resources..... 118

References 118

SECTION III Health Program Development and Planning 121

Chapter 5 Program Theory and Interventions Revealed.....123

Program Theory 124

 Process Theory 125

 Effect Theory 125

Interventions 126

 Finding and Identifying Interventions..... 126

 Types of Interventions 127

 Specifying Intervention Administration and Dosage..... 128

 Interventions and Program Components 130

 Characteristics of Good Interventions..... 131

Path to Program Outcomes and Impacts 134

 Components of the Effect Theory..... 135

 Matching Levels: Audience, Cause, Intervention, and Effects 137

Generating the Effect Theory 138

 Involve Key Stakeholders 138

 Draw Upon the Scientific Literature 138

 Diagram the Causal Chain of Events..... 140

 Check Against Assumptions 141

Functions of Program Theory 141

 Provide Guidance..... 141

 Enable Explanations 142

 Form a Basis for Communication..... 142

 Make a Scientific Contribution 143

Across the Pyramid 143

Discussion Questions and Activities..... 144

Internet Resources..... 144

References 145

Chapter 6 Program Objectives and Setting Targets147

Program Goals and Objectives 147

 Goals..... 147

 Foci of Objectives..... 148

 Objectives and Indicators..... 151

 Good Goals and Objectives..... 154

Using Data to Set Target Values 156

 Decisional Framework for Setting Target Values..... 156

 Stratification and Object Target Values 159

 Use of Logic Statements to Develop Targets..... 160

 Options for Calculating Target Values 160

Caveats to the Goal-Oriented Approach 170

Across the Pyramid 171

Discussion Questions and Activities..... 171

Internet Resources..... 172

References 172

SECTION IV Implementing and Monitoring the Health Program 173

Chapter 7 Process Theory for Program Implementation 175

Organizational Plan Inputs 175

- Human Resources 177
- Physical Resources..... 179
- Transportation..... 180
- Informational Resources 180
- Time 180
- Managerial Resources 180
- Fiscal Resources 182

Organizational Plan Outputs 182

- Time Line..... 182
- Operations Manual 182
- Organizational Chart..... 184
- Information System..... 185

Inputs to Service Utilization Plan 185

- Social Marketing..... 185
- Eligibility Screening..... 185
- Queuing..... 189
- Intervention Delivery 189

Services Utilization Plan Outputs 191

- Summary: Elements of Organizational and Services Utilization Plans..... 192

Alternative Plan Formats 192

- Logic Models 193
- Business Plans 195

Across the Pyramid 195

Discussion Questions and Activities..... 197

Internet Resources..... 197

References 198

Chapter 8 Monitoring Implementation Through Budgets and Information Systems 201

Budgets and Budgeting 201

- Budgeting Terminology..... 202

Budgeting as Part of Planning..... 204

- Monetize and Compute Program Costs..... 204
- Budget for Start-Up and Evaluation Costs... 205
- Break-Even Analysis..... 205
- Budget Justification 207

Budget as a Monitoring Tool 209

- Budget Variance..... 209
- Types of Cost Analyses..... 209

Information Systems..... 213

- Health Informatics Terminology..... 214
- Information Systems Considerations 214

Across the Pyramid 216

Discussion Questions and Activities..... 217

Internet Resources..... 217

References 218

Chapter 9 Implementation Evaluation: Measuring Inputs and Outputs 219

Assessing the Implementation..... 219

- Implementation Documentation 220
- Implementation Assessment 221
- Implementation Evaluation..... 221

Efficacy, Effectiveness, and Efficiency..... 222

Data Collection Methods 223

Quantifying Inputs to the Organizational Plan..... 223

- Human Resources 228
- Physical Resources..... 229

Quantifying Outputs of the Organizational Plan..... 230

- Information Systems..... 230
- Monetary Resources..... 230

Quantifying Inputs to the Services Utilization Plan..... 230

- Participants and Recipients 230
- Intervention Delivery and Fidelity..... 231

Quantifying Outputs of the Services Utilization Plan..... 234

- Coverage as Program Reach 234
- Participant-Related Issues..... 238
- Program Logistics..... 240

Across the Pyramid 241
 Discussion Questions and Activities 242
 Internet Resources 243
 References 243

Chapter 10 Program Quality and Fidelity: Managerial and Contextual Considerations. 245

The Accountability Context 246
 Program Accountability 246
 Professional Accountability 246
 Performance and Quality: Navigating the Interface 247
 Quality Improvement Approaches 248
 Quality Improvement Tools 248
 Relevance to Health Programs 251
 Performance Measurement 252
 Informatics and Information Technology 253
 Creating Change for Quality and Fidelity 255
 Interpreting Implementation Data 255
 Maintaining Program Process Quality and Fidelity 257
 Managing Group Processes for Quality and Fidelity 258
 When and What Not to Change 259
 Formative Evaluations 259
 Across the Pyramid 259
 Discussion Questions 260
 Internet Resources 260
 References 261

SECTION V Outcome and Impact Evaluation of Health Programs 263

Chapter 11 Planning the Intervention Effect Evaluations. 265

Developing the Evaluation Questions 266
 Characteristics of the Right Question 267
 Outcome Documentation, Outcome Assessment, and Outcome Evaluation 268

Evaluation and Research 268
 Rigor in Evaluation 270
 Variables from the Program Effect Theory 271
 Outcome and Impact Dependent Variables 271
 Causal Factors as Independent Variables 273
 Antecedent, Moderating, and Mediating Factors as Variables 273
 Measurement Considerations 275
 Units of Observation 275
 Types of Variables (Levels of Measurement) 275
 Timing 278
 Sensitivity of Measures 278
 Threats to Data Quality 279
 Missing Data 279
 Reliability Concerns 280
 Validity of Measures 281
 Contextual Considerations in Planning the Evaluation 281
 Evaluation Standards 281
 Ethics 282
 Stakeholders 282
 Across the Pyramid 284
 Discussion Questions and Activities 284
 Internet Resources 285
 References 285

Chapter 12 Choosing Designs for Effect Evaluations. 287

Evaluation Design Caveats 288
 Considerations in Choosing a Design 289
 Using Designs Derived from Multiple Paradigms: An Example 294
 Choosing the Evaluation Design 294
 Identifying Design Options 294
 Overview of the Decision Tree 295
 Designs for Outcome Documentation 298
 Designs for Outcome Assessment: Establishing Association 301

Designs for Outcome Evaluation: Establishing Causation	307
Practical Issues with Experimental Designs	307
Designs and Failures	309
Across the Pyramid	311
Discussion Questions	312
Internet Resources	312
References	312

Chapter 13 Sampling Designs and Data Sources for Effect Evaluations315

Sampling Realities	315
Sample Construction	317
Hard-to-Reach Populations	318
Sample Size	318
Calculating Response Rates	319
Sampling for Effect Evaluations	322
Sampling for Outcome Assessment.....	322
Sampling for Outcome Evaluation	324
Data Collection Methods	324
Surveys and Questionnaires	325
Secondary Data	328
Big Data	329
Physical Data	330
Across the Pyramid	330
Discussion Questions and Activities	330
Internet Resources	331
References	331

Chapter 14 Quantitative Data Analysis and Interpretation335

Data Entry and Management	335
Outliers	337
Linked Data	337
Sample Description	338
Thinking About Change	339
Change as a Difference Score	339

Issues with Quantifying Change from the Program	339
Relationship of Change to Intervention Effort	342
Clinical and Statistical Significance	343
Across Levels of Analysis	343
Statistical Answers to the Questions	345
Description	346
Comparison	348
Association	349
Prediction	350
Interpretation	353
Four Fallacies of Interpretation	353
Ecological Fallacy	354
Across the Pyramid	356
Discussion Questions and Activities	356
Internet Resources	357
References	357

Chapter 15 Qualitative Methods for Planning and Evaluation359

Qualitative Methods Throughout the Planning and Evaluation Cycle	359
Qualitative Methods	360
Individual In-Depth Interview	361
Written Open-Ended Questions	362
Focus Group	363
Observation	364
Data Collection	364
Case Study	365
Innovative Methods	366
Scientific Rigor	368
Sampling for Qualitative Methods	369
Analysis of Qualitative Data	372
Overview of Analytic Process	372
Software	374
Issues to Consider	374
Presentation of Findings	375
Across the Pyramid	376

Discussion Questions and Activities 377
 Internet Resources 377
 References 377

**SECTION VI Additional Considerations
 for Evaluators 381**

**Chapter 16 Program Evaluators’
 Responsibilities 383**

Ethical Responsibilities 383
 Ethics and Planning 383
 Institutional Review Board Approval
 and Informed Consent 385
 Ethics and Evaluation 387
 HIPAA and Evaluations 388
 Responsible Spin of Data and Information 389
 Persuasion and Information 389
 Information and Sensemaking 391

Reporting Responsibly 392
 Report Writing 392
 Making Recommendations 394
 Misuse of Evaluations 397
 Responsible Contracts 398
 Organization–Evaluator Relationship 398
 Health Policy 399
 Responsible for Evaluation Quality 400
 Responsible for Dissemination 401
 Responsible for Current Practice 402
 Across the Pyramid 404
 Discussion Questions and Activities 405
 Internet Resources 405
 References 405

Index 409

List of Figures

- Figure 1-1** The Planning and Evaluation Cycle
- Figure 1-2** The Public Health Pyramid
- Figure 1-3** The Pyramid as an Ecological Model
- Figure 2-1** Effects of Diversity Throughout the Planning and Evaluation Cycle Stage in the Planning and Evaluation Cycle
- Figure 3-1** Connections Among Program, Agency, and Community
- Figure 3-2** Venn Diagram of Community-Based, Community-Focused, and Community-Driven
- Figure 3-3** The Planning and Evaluation Cycle
- Figure 4-1** Generic Model of a Theory of Causes
- Figure 4-2** Diagram of Theory of Causes/ Determinants of Receiving Immunizations, as Contributing to Adult Immunization Rates, Using the Layetteville Example
- Figure 4-3** Diagram of Theory of Causes/ Determinants for Deaths from Gunshot Wounds, as Contributing to Adolescent Death Rates, Using the Layetteville Example
- Figure 4-4** Diagram of Theory of Causes/ Determinants for Neural Tube Defects, as Contributing to Rates of Congenital Anomalies, Using the Bowe County Example
- Figure 4-5** Theory of Causes/Determinants with Elements of the BPRS Score: Size, Seriousness, and Interventions
- Figure 5-1** Model of Program Theory
- Figure 5-2** The Effect Theory Showing the Causal Theory Using Community Diagnosis Elements
- Figure 5-3** Effect Theory Example: Effect Theory for Reducing the Rate of Congenital Anomalies
- Figure 5-4** Two Roots of Program Failure
- Figure 6-1** Using Elements of Program Theory as the Basis for Writing Program Objectives
- Figure 6-2** Diagram Showing Relationship of Effect Theory Elements to Process and Outcome Objectives
- Figure 6-3** Calculations of Options 1 Through 4 Using a Spreadsheet
- Figure 6-4** Calculations of Options 5 Through 8 Using a Spreadsheet
- Figure 6-5** Calculations of Options 9 and 10 Using a Spreadsheet
- Figure 7-1** Amount of Effort Across the Life of a Health Program
- Figure 7-2** Diagram of the Process Theory Elements Showing the Components of the Organizational Plan and Services Utilization Plan
- Figure 7-3** Process Theory for Neural Tube Defects and Congenital Anomalies Health Problem
- Figure 7-4** Effect and Process Theory for Neural Tube Defect Prevention Program
- Figure 8-1** Relevance of Process Theory to Economic Evaluations
- Figure 8-2** Information System Processes Throughout the Program Planning Cycle
- Figure 9-1** Elements of the Process Theory Included in a Process Evaluation
- Figure 9-2** Roots of Program Failure

- Figure 9-3** Examples of Organizational Plan Inputs and Outputs That Can Be Measured
- Figure 9-4** Examples of Services Utilization Inputs and Outputs That Can Be Measured
- Figure 10-1** List of Quality Improvement Tools with Graphic Examples
- Figure 11-1** Planning and Evaluation Cycle, with Effect Evaluation Highlights
- Figure 11-2** Diagram of Net Effects to Which Measures Need to Be Sensitive
- Figure 11-3** Using the Effect Theory to Identify Effect Evaluation Variables
- Figure 11-4** Effect Theory of Reducing Congenital Anomalies Showing Variables
- Figure 12-1** Relationship Between the Ability to Show Causality and the Costs and Complexity of the Design
- Figure 12-2** Decision Tree for Choosing an Evaluation Design, Based on the Design's Typical Use
- Figure 12-3** Three Sources of Program Failure
- Figure 13-1** Probability and Nonprobability Samples and Their Usage
- Figure 14-1** Contributing Factors to the Total Amount of Change
- Figure 14-2** Summary of the Three Decisions for Choosing an Analytic Approach
- Figure 14-3** Five Ways That the Rate of Change Can Be Altered
- Figure 16-1** Making Recommendations Related to the Organizational and Services Utilization Plans
- Figure 16-2** Making Recommendations Related to the Program Theory
- Figure 16-3** The Planning and Evaluation Cycle with Potential Points for Recommendations

List of Tables

- Table 1-1** Comparison of Outcome-Focused, Utilization-Focused, and Participatory Focused Evaluations
- Table 1-2** Evaluation Standards Established by the Joint Commission on Standards for Educational Evaluation
- Table 1-3** Fuzzy Aspects Throughout the Planning and Evaluation Cycle
- Table 1-4** A Summary of the *Healthy People 2020* Priority Areas
- Table 2-1** Examples of Cultural Tailoring Throughout the Program Planning and Evaluation Cycle
- Table 2-2** Indicators Used to Measure Race in Different Surveys
- Table 2-3** Professional Diversity Among Health Professions
- Table 2-4** Cultural Continuum with Examples of the Distinguishing Features of Each Stage
- Table 3-1** Three Elements of Community, with Their Characteristics
- Table 3-2** Summary of the Six Approaches to Planning, with Public Health Examples
- Table 3-3** Comparison of Models Developed for Public Health Planning
- Table 3-4** A Comparison of the Five Perspectives on Community Health and Needs Assessment
- Table 4-1** Haddon's Typology for Analyzing an Event, Modified for Use in Developing Health Promotion and Prevention Programs
- Table 4-2** Quality-of-Life Acronyms and Definitions
- Table 4-3** Global Leading Causes of Disability-Adjusted Life-Years (DALYs) and Years of Life Lost (YLL)
- Table 4-4** Numerators and Denominators for Selected Epidemiological Rates Commonly Used in Community Health Assessments
- Table 4-5** Existing Factors, Moderating Factors, Key Causal Factors, Mediating Factors, and Health Outcome and Impact for Five Health Problems in Layetteville and Bowe County
- Table 4-6** Relationship of Problem Definition to Program Design and Evaluation
- Table 4-7** Criteria for Rating Problems According to the BPRS
- Table 4-8** Program Prioritization Based on the Importance and Changeability of the Health Problem
- Table 4-9** Examples of Sources of Data for Prioritizing Health Problems at Each Level of the Public Health Pyramid
- Table 4-10** Examples of Required Existing, Causal, and Moderating Factors Across the Pyramid
- Table 5-1** Examples of Interventions by Type and Level of the Public Health Pyramid
- Table 5-2** Comparison of Effect Theory, Espoused Theory, and Theory-in-Use
- Table 5-3** Examples of Types of Theories Relevant to Developing Theory of Causative/Determinant Factors or Theory of Intervention Mechanisms by Four Health Domains

Table 5-4	Examples of Types of Theories Relevant to Developing the Organizational Plan and Services Utilization Plan Components of the Process Theory	Table 7-3	Examples of Partial- and Full-Coverage Programs by Level of the Public Health Pyramid
Table 6-1	Aspects of Process Objectives as Related to Components of the Process Theory, Showing the TAAPS Elements	Table 7-4	Template for Tracking Services Utilization Outputs Using Example Interventions and Hypothetical Activities
Table 6-2	Domains of Individual or Family Health Outcomes with Examples of Corresponding Indicators and Standardized Measures	Table 7-5	Hypothetical Logic Model of a Program for Reducing Congenital Anomalies
Table 6-3	Bowe County Health Problems with Indicators, Health Outcomes, and Health Goals	Table 7-6	Generic Elements of a Business Plan, with Their Purpose and Corresponding Element of the Process Theory and Logic Model
Table 6-4	Effect Objectives Related to the Theory of Causal/Determinant Factors, Theory of the Intervention Mechanisms, and Theory of Outcome to Impact, Using Congenital Anomalies as an Example, Showing the TREW Elements	Table 8-1	Formulas Applied for Options A and B
Table 6-5	Effect Objectives Related to the Theory of Causal/Determinant Factors, Theory of the Intervention Mechanisms, and Theory of Outcome to Impact, Using Adolescent Pregnancy as an Example, Showing the TREW Elements	Table 9-1	Methods of Collecting Process Evaluation Data
Table 6-6	Matrix of Decision Options Based on Current Indicator Value, Population Trend of the Health Indicator, and Value of Long-Term Objective or Standard	Table 9-2	Example of Measures of Inputs and Outputs of the Organizational Plan
Table 6-7	Framework for Target Setting: Interaction of Data Source Availability and Consistency of Information	Table 9-3	Examples of Measures of Inputs and Outputs of the Services Utilization Plan
Table 6-8	Summary of When to Use Each Option	Table 9-4	Matrix of Undercoverage, Ideal Coverage, and Overcoverage
Table 6-9	Range of Target Values Derived from Options 1 Through 10, Based on the Data from Figures 6-3 Through 6-5	Table 9-5	Examples of Process Evaluation Measures Across the Public Health Pyramid
Table 7-1	List of Health Professionals with a Summary of Typical Legal and Regulatory Considerations	Table 10-1	Types of Program Accountability, with Definitions and Examples of Process Evaluation Indicators
Table 7-2	Relationship of Test Sensitivity and Specificity to Overinclusion and Underinclusion	Table 10-2	Comparison of Improvement Methodologies and Program Process Evaluation
		Table 10-3	Definitions of Terms Used in Performance Measurement
		Table 10-4	Partial List of Existing Performance Measurement Systems Used by Healthcare Organizations, with Their Websites
		Table 11-1	Three Levels of Intervention Effect Evaluations
		Table 11-2	Differences Between Evaluation and Research

- Table 11-3** Advantages and Disadvantages of Using Each Type of Variable
- Table 11-4** Examples of Nominal, Ordinal, and Continuous Variables for Different Health Domains
- Table 11-5** Example Time Line Showing the Sequence of Intervention and Evaluation Activities
- Table 11-6** Summary of Evaluation Elements
- Table 12-1** Contribution of Various Disciplines to Health Program Evaluation
- Table 12-2** Summary of Main Designs and Their Use for Individual or Population-Level Evaluations
- Table 12-3** Approaches to Minimizing Each of the Three Types of Program Failure
- Table 13-1** Probability and Nonprobability Samples and Their Usage
- Table 13-2** Comparison of Main Types of Samples with Regard to Implementation Ease, Degree of Representativeness, and Complexity of Sampling Frame
- Table 13-3** Example of Data Sources for Each Health and Well-Being Domain
- Table 13-4** Interaction of Response Bias and Variable Error
- Table 14-1** Calculation of Effectiveness and Adequacy Indices: An Example
- Table 14-2** Intervention Efficiency as a Relation of Effect Size and Causal Size
- Table 14-3** Factors That Affect the Choice of a Statistical Test: Questions to Be Answered
- Table 14-4** Analysis Procedures by Level of Intervention and Level of Analysis
- Table 14-5** Commonly Used Parametric and Nonparametric Statistical Tests for Comparison, Association, and Prediction
- Table 14-6** Main Types of Comparison Analyses Used by Level of Analysis and Assuming That the Variables Are at the Same Level of Measurement
- Table 14-7** Main Types of Association Analyses Used by Level of Analysis, Assuming That Variables Are the Same Level of Measurement
- Table 14-8** Example of Statistical Tests for Strength of Association by Level of Measurement, Using Layetteville Adolescent Antiviolence Program
- Table 14-9** Examples of Statistical Tests by Evaluation Design and Level of Measurement, with Examples of Variables
- Table 14-10** Main Types of Prediction Analyses Used by Level of Analysis, Assuming That Variables Are at the Same Level of Measurement
- Table 15-1** Comparison of Qualitative Perspectives with Regard to the Basic Question Addressed and the Relevance to Health Program Planning and Evaluation
- Table 15-2** Comparison of Major Qualitative Perspectives with Regard to the Method Used
- Table 15-3** Summary of Key Benefits and Challenges to Using Qualitative Methods in Planning and Evaluation
- Table 15-4** Sampling Considerations for Each of the Qualitative Methods Discussed
- Table 15-5** Summary of Types of Sampling Strategies Used with Qualitative Designs
- Table 15-6** Example of Interview Text with Final Coding
- Table 15-7** Suggested Qualitative Methods by Pyramid Level and Planning Cycle
- Table 16-1** Ethical Frameworks and Principles for Planning Health Programs
- Table 16-2** Comparison of Types of IRB Reviews
- Table 16-3** Eight Elements of Informed Consent, as Required in 45 CFR 46
- Table 16-4** Effect of Rigor and Importance of Claims on Decision Making
- Table 16-5** List of Ways to Make Graphs More Interpretable
- Table 16-6** Examples of Dissemination Modes, Audiences, and Purposes

List of Exhibits

- Exhibit 2-1** Checklist to Facilitate Development of Cultural and Linguistic Competence Within Healthcare Organizations
- Exhibit 2-2** Checklist to Facilitate Cultural Competence in Community Engagement
- Exhibit 7-1** Example of an Abbreviated Time Line for a Short-Term Health Program
- Exhibit 7-2** Chapter Text Paragraph Rewritten at an Eighth-Grade Reading Level
- Exhibit 8-1** Example of a Scenario Needing a Break-Even Analysis
- Exhibit 8-2** Example of a Budget Used for a Break-Even Analysis for Bright Light on an Excel Spreadsheet
- Exhibit 8-3** Break-Even Table Shows Number of Paying Students Needed to Break Even
- Exhibit 8-4** Example of a Budget Showing Year-to-Date Variance
- Exhibit 8-5** Types of Cost Analyses
- Exhibit 9-1** Formulas for Measures of Coverage
- Exhibit 9-2** Example of Narrative Background about Coverage and Dosage Measures
- Exhibit 9-3** Examples of Coverage Measures Using an Excel Spreadsheet
- Exhibit 9-4** Examples of Calculating Dosage for the Congenital Anomalies Prevention Program Using Excel

Preface to the Fourth Edition

The fourth edition of *Health Program Planning and Evaluation* has stayed true to the purpose and intent of the previous editions. This advanced-level text is written to address the needs of professionals from diverse health disciplines who find themselves responsible for developing, implementing, or evaluating health programs. The aim of the text is to assist health professionals to become not only competent health program planners and evaluators but also savvy consumers of evaluation reports and prudent users of evaluation consultants. To that end, the text includes a variety of practical tools and concepts necessary to develop and evaluate health programs, presenting them in language understandable to both the practicing and novice health program planner and evaluator.

Health programs are conceptualized as encompassing a broad range of programmatic interventions that span the social-ecological range, from individual-level to population-level programs. Examples of programs cited throughout the text are specific yet broadly related to improving health and reflect the breadth of public health programs. The examples have been updated once again to reflect current best practices. Maintaining a public health focus provides an opportunity to demonstrate how health programs can target different levels of a population, different determinants of a health problem, and different strategies and interventions to address a health problem. In addition, examples of health programs and references are selected to pique the interests of the diverse students and practicing professionals who constitute multidisciplinary program teams. Thus, the content and examples presented throughout

the text are relevant to health administrators, medical social workers, nurses, nutritionists, pharmacists, public health professionals, physical and occupational therapists, and physicians.

This textbook grew from teaching experiences with both nurses and public health students and their need for direct application of the program planning and evaluation course content to their work and to their clients and communities. Today programs need to be provided through community-based healthcare settings to address broad public health issues and expand the individual to population focus. The distinction between individual patient health and population health is a prerequisite for the thinking and planning—in terms of aggregates and full populations—by students from clinical backgrounds.

In most graduate health professions programs, students take a research methods course and a statistics course. Therefore, this evaluation text avoids duplicating content related to research methods and statistics while addressing and extending that content into health program development, implementation, and evaluation. In addition, because total quality management and related methodologies are widely used in healthcare organizations, areas of overlap between quality improvement methodologies and traditional program evaluation approaches are discussed. This includes ways that quality improvement methodologies complement program evaluations. Sometimes evaluations are appropriate; sometimes they are not. Enthusiasm for providing health programs and performing evaluation is tempered with thoughtful notes of caution in the hope that students will avoid potentially serious and costly program and evaluation mistakes.

► Unique Features

The *Fourth Edition* has retained the three unique features that distinguish this text from other program planning and evaluation textbooks: use of the public health pyramid, consistent use of a model of the program theory throughout the text, and role modeling of evidence-based practice.

The public health pyramid explains how health programs can be developed for individuals, aggregates, populations, and service delivery systems. Use of the pyramid is also intended as a practical application of the social-ecological perspective that acknowledges a multilevel approach to addressing health problems. The public health pyramid contains four levels: direct services to individuals; enabling services to aggregates; services provided to entire populations; and, at the base, infrastructure. In this textbook, the pyramid is used as an organizing structure to summarize the content of each chapter in the “Across the Pyramid” sections. In these sections, specific attention is paid to how key concepts in a given chapter might vary across each pyramid level. Summarizing the chapter content in this manner reinforces the perspective that enhancing health and well-being requires integrated efforts across the levels of the public health pyramid. Health program development and evaluation is relevant for programs intended for individuals, aggregates, populations, and service delivery systems, and this fact reinforces the need to tailor program plans and evaluation designs to the level at which the program is conceptualized. Using the pyramid also helps health professionals begin to value their own and others’ contributions within and across the levels and to transcend disciplinary boundaries.

The second unique feature of this text is that one conceptual model of program planning and evaluation is used throughout the text: the program theory. The program theory is like a curricular strand, connecting content across the chapters, and activities throughout the planning and evaluation cycle. The program theory, as a conceptual model, is composed of elements.

Articulating each of the component elements of the program theory sharpens the student’s awareness of what must be addressed to create an effective health program. One element of the program theory is the effect theory, which focuses on how the intervention results in the program effects. The effect theory had its genesis in the concepts of action and intervention hypotheses described by Rossi and Freeman; those concepts were dropped from later editions of their text. We believe these authors were onto something with their effort to elucidate the various pathways leading from a problem to an effect of the program. Rossi and colleagues’ ideas have been updated with the language of moderating and mediating factors and an emphasis on the intervention mechanisms.

Throughout the current edition of this textbook, emphasis is given to the effect theory portion of the program theory. The effect theory describes relationships among health antecedents, causes of health problems, program interventions, and health effects. The hypotheses that comprise the effect theory need to be understood and explicated to plan a successful health program and to evaluate the “right” elements of the program. The usefulness of the effect theory throughout the planning and evaluation cycle is highlighted throughout this text; for example, the model is used as a means of linking program theory to evaluation designs and data collection. The model becomes an educational tool by serving as an example of how the program theory is manifested throughout the stages of planning and evaluation, and by reinforcing the value of carefully articulating the causes of health problems and consequences of programmatic interventions. Students and novice program planners may have an intuitive sense of the connection between their actions and outcomes, but they may not know how to articulate those connections in ways that program stakeholders can readily grasp. The effect theory and the process theory—the other main element of the program theory—provide a basis from which to identify and describe these connections.

The third unique feature of this text is the intentional role modeling of evidence-based practice. Use of published, empirical evidence as the basis for practice—whether clinical practice or program planning practice—is the professional standard. Each chapter of this book contains substantive examples drawn from the published scientific health and health-related literature. Relying on the literature for examples of programs, evaluations, and issues is consistent with the espoused preference of using scientific evidence as the basis for making programmatic decisions. Each chapter offers multiple examples from the health sciences literature that substantiate the information presented in the chapter.

► Organization of the Book

The book is organized into six sections, each covering a major phase in the planning and evaluation cycle. Chapter 1 introduces the fictitious city of Layetteville and the equally fictitious Bowe County. In subsequent chapters, chapter content is applied to the health problems of Layetteville and Bowe County so that students can learn how to use the material on an ongoing basis. In several chapters, the case study is used in the “Discussion Questions and Activities” section to provide students with an opportunity to practice applying the chapter content. In recognition of the availability of parts of the text in digital format, each use of the Layetteville case stands on its own in reference to the chapter’s content.

Section I explores the context in which health programs and evaluations occur. Chapter 1 begins with an overview of definitions of health, followed by a historical context. The public health pyramid is introduced and presented as an ecological framework for thinking of health programs. An overview of community is provided and discussed as both the target and the context of health programs. The role of community members in health programs and

evaluations is introduced, and emphasis is given to community as a context and to strategies for community participation throughout the program development and evaluation process. Chapter 2 focuses on the role of diversity in the planning and evaluation cycle and its effects on the delivery and evaluation of health programs. Although a discussion of diversity-related issues could have been added to each chapter, the sensitive nature of this topic and its importance in ensuring a successful health program warranted it being covered early in the text and as a separate chapter. Cultural competence is discussed, particularly with regard to the organization providing the health program and with regard to the program staff members.

Section II contains two chapters that focus on the task of defining the health problem. Chapter 3 covers planning perspectives and the history of health program planning. Effective health program developers understand that approaches to planning are based on assumptions. These assumptions are exemplified in six perspectives that provide points of reference for understanding diverse preferences for prioritizing health needs and expenditures and therefore for tailoring planning actions to fit the situation best. Chapter 3 also reviews perspectives on conducting a community needs assessment as foundational to decision making about the future health program. Essential steps involved in conducting a community health and needs assessment are outlined as well.

Chapter 4 expands on key elements of a community needs assessment, beginning with a review of the data collection methods appropriate for a community needs assessment. This discussion is followed by a brief overview of key epidemiological statistics. Using those statistics and the data, the reader is guided through the process of developing a causal statement of the health problem. This causal statement, which includes the notion of moderating and mediating factors in the pathway from causes to problem, serves as the basis for the effect theory of the program. Once the causal statement has been

developed, prioritization of the problem is needed; four systems for prioritizing in a rational manner are reviewed in Chapter 4.

Following prioritization comes planning, beginning with the decision of how to address the health problem. In many ways, the two chapters in Section III form the heart of planning a successful health program. Unfortunately, students generally undervalue the importance of theory for selecting an effective intervention and of establishing target values for objectives. Chapter 5 explains what theory is and how it provides a cornerstone for programs and evaluations. More important, the concept of intervention is discussed in detail, with attention given to characteristics that make an intervention ideal, including attention to intervention dosage. Program theory is introduced in Chapter 5 as the basis for organizing ideas related to the selection and delivery of the interventions in conjunction. The effect theory element of the program theory is introduced and the components of the effect theory are explained. Because the effect theory is so central to having an effective program intervention and the subsequent program evaluation, it is discussed in conjunction with several examples from the Layetteville and Bowe County case.

Chapter 6 goes into detail about developing goals and objectives for the program, with particular attention devoted to articulating the interventions provided by the program. A step-by-step procedure is presented for deriving numerical targets for the objectives from existing data, which makes the numerical targets more defensible and programmatically realistic. We focus on distinguishing between process objectives and outcome objectives through the introduction of two mnemonics: TAAPS (Time frame, Amount of what Activities done by which Participants/program Staff) and TREW (Timeframe, what portion of Recipients experience what Extent of Which type of change).

Section IV deals with the task of implementing a health program. Chapter 7 provides an in-depth review of key elements that constitute the process theory element of the program theory—specifically, the organizational plan and services utilization plan. The distinction between

inputs and outputs of the process theory is highlighted through examples and a comprehensive review of possible inputs and outputs. Budgeting for program operations is covered in this chapter as well. Chapter 8 is devoted entirely to fiscal data systems, including key aspects of budgeting, and informatics. Chapter 9 details how to evaluate the outputs of the organizational plan and the services utilization plan. The practical application of measures of coverage is described, along with the need to connect the results of the process evaluation to programmatic changes. Program management for assuring a high-quality program that delivers the planned intervention is the focus of Chapter 10.

Section V contains chapters that are specific to conducting the effect evaluations. These chapters present both basic and advanced research methods from the perspective of a program effect evaluation. Here, students' prior knowledge about research methods and statistics is brought together in the context of health program and services evaluation. Chapter 11 highlights the importance of refining the evaluation question and provides information on how to clarify the question with stakeholders. Earlier discussions about program theory are brought to bear on the development of the evaluation question. Key issues, such as data integrity and survey construction, are addressed with regard to the practicality of program evaluation. Chapter 12 takes a fresh approach to evaluation design by organizing the traditional experimental and quasi-experimental designs and epidemiological designs into three levels of program evaluation design based on the design complexity and purpose of the evaluation. The discussion of sampling in Chapter 13 retains the emphasis on practicality for program evaluation rather than taking a pure research approach. However, sample size and power are discussed because these factors have profound relevance to program evaluation. Chapter 14 reviews statistical analysis of data, with special attention to variables from the effect theory and their level of measurement. The data analysis is linked to interpretation, and students are warned about potential flaws in how numbers are understood. Chapter 15 provides a review of qualitative designs

and methods, especially their use in health program development and evaluation.

The final section, Section VI, includes just one chapter. Chapter 16 discusses the use of evaluation results when making decisions about existing and future health programs. Practical and conceptual issues related to the ethics issues that program evaluators face are addressed. This chapter also reviews ways to assess the quality of evaluations and the professional responsibilities of evaluators.

Each chapter in the book concludes with a “Discussion Questions and Activities” section. The questions posed are intended to be provocative and to generate critical thinking. At the graduate level, students need to be encouraged to engage in independent thinking and to foster their ability to provide rationales for decisions. The discussion questions are developed from this point of view. In the “Internet Resources” section, links are provided to websites that support the content of the chapter. These websites have been carefully chosen as stable and reliable sources.

► Additions to and Revisions in the Fourth Edition

The fourth edition of *Health Program Planning and Evaluation* represents continuous improvement, with corrections and updated references. Classical references and references that remain state of the art have been retained.

The *Fourth Edition* has retained the original intent—namely, to provide students with the ability to describe a working theory of how the intervention acts upon the causes of the health problem and leads to the desired health results. Some content has been condensed in order to allow enough room to describe current evaluation approaches adequately for both new and experienced practitioners. For instance, Chapter 1 now includes participatory evaluations in addition to outcome- and utilization-focused evaluations. In addition to disciplines traditionally recognized

in western medical care, Chapter 2 now includes acupuncture and massage therapy as examples of health professional diversity. Discussion of the nuances of cultural competency has been refined, in light of the continuing importance and challenges of this area. Community strengths have been given more attention in Chapter 3 in recognition of the powerful potential of shifting from a “deficit-based” to an “asset-based” perspective on health planning. Chapter 4 now devotes greater attention to the health evaluation potential of data from social media such as Facebook and Twitter, as well as geospatial data, including attendant concerns about privacy, and also notes implications of the increasingly prevalent public rankings of community health status. Examples of infrastructure-level interventions within the public health pyramid have been added in Chapter 5. Discussion of financial modeling options in Chapter 8 now includes simulation modeling, an exciting if also resource-intensive option to conducting real-world experiments, which are, of course, inevitably expensive themselves. Chapters 9 and 15 include emerging data collection techniques such as participant self-reports, video, photos, and audio recordings that may make public health evaluation more inclusive of the people such interventions seek to serve. Chapter 13 includes updates on surveying, reflecting the decreased numbers of people with land-line phones, long a mainstay of health evaluations. Options for online surveying have been updated in Chapter 14; given the rapid evolution of big data such as those available from social media, billing, and medical records, discussion of this topic has been updated in Chapter 13 as well. Finally, Chapter 16 now includes bioethics—the application of ethical and philosophical principles to medical decision making—as an increasingly salient component of responsible health evaluation.

In sum, we have worked hard to sustain this book’s conceptual and empirical rigor and currency in the *Fourth Edition* while maintaining accessibility for a range of health evaluators. Above all, we hope this book is useful to our readers’ vitally important efforts to improve health.

Acknowledgments

We are indebted to the many people who supported and aided us in preparing this fourth edition of *Health Program Planning and Evaluation: A Practical, Systematic Approach for Community Health*. We remain grateful to the numerous students over the years who asked questions that revealed the typical sticking points in their acquiring and understanding of the concepts and content, as well as where new explanations were needed. Through their eyes we have learned there is no one way to explain a complex notion or process. Their interest and enthusiasm for planning and evaluating health programs was a great motivator for writing this book.

Several additional colleagues helped fine-tune this text. We are especially indebted to Arden

Handler at the School of Public Health, University of Illinois at Chicago, for taking time to contribute to this textbook. Her devotion to quality and clarity has added much to the richness of otherwise dry material. We remain deeply indebted to Deborah Rosenberg, also at the School of Public Health University of Illinois at Chicago, for sharing her innovative and quintessentially useful work on developing targets for program objectives. Special thanks as well to Joseph Chen, at the University of Texas School of Public Health, for his many contributions to updating the literature cited across many chapters and for his contribution on big data. Last, but not least, thanks to Mike Brown, publisher at Jones & Bartlett Learning, for his encouragement and patience over the years.

List of Acronyms

ABCD	Asset-based community development	DHHS	U.S. Department of Health and Human Services
ACA	Affordable Care Act	DSM-5	<i>Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition</i>
AEA	American Evaluation Association	EBM	Evidence-based medicine
AHRQ	Agency for Healthcare Research and Quality	EBP	Evidence-based practice
ANOVA	Analysis of variance	EHR	Electronic health record
APHA	American Public Health Association	EMR	Electronic medical record
BPRS	Basic priority rating system	FTE	Full-time equivalent
BRFSS	Behavioral Risk Factor Surveillance System	GAO	U.S. Government Accountability Office
BSC	Balanced Score Card	GNP	Gross Product
CAHPS	Consumer Assessment of Health Plans	GPRA	Government Performance and Results Act
CARF	Commission on Accreditation of Rehabilitation Facilities	HEDIS	Healthcare Effectiveness Data and Information Set
CAST-5	Capacity Assessment of Title-V	HIPAA	Health Insurance Portability and Accountability Act
CBA	Cost-benefit analysis	HIT	Health information technology
CBPR	Community-based participatory research	HMOs	Health maintenance organizations
CDC	Centers for Disease Control and Prevention	HRQOL	Health-related quality of life
CEA	Cost-effectiveness analysis	HRSA	Health Resources and Services Administration (part of DHHS)
CER	Cost-effectiveness ratio	i-APP	Innovation-Adolescent Preventing Pregnancy (Program)
CFIR	Consolidated Framework for Implementation Research	ICC	Intraclass correlation
CFR	Code of Federal Regulations	IRB	Institutional review board
CHIP	Community Health Improvement Process	JCAHO	Joint Commission on the Accreditation of Healthcare Organizations
CI	Confidence interval	MAPP	Mobilizing for Action through Planning and Partnership
CPT	Current Procedural Terminology	MBO	Management by objectives
CQI	Continuous quality improvement		
CUA	Cost-utility analysis		
DALY	Disability-adjusted life-year		

xxviii List of Acronyms

MCHB	Maternal and Child Health Bureau (part of HRSA)	PSA	Public service announcement
NACCHO	National Association of City and County Health Officers	QALY	Quality-adjusted life-year
NAMI	National Alliance on Mental Illness	RAR	Rapid assessment and response
NCHS	National Center for Health Statistics	RARE	Rapid assessment and response and evaluation
NCQA	National Commission on Quality Assurance	RE-AIM	Reach, Effectiveness, Adoption, Implementation, and Maintenance model
NFPS	National Family Planning Survey	RR	Relative risk
NHANES	National Health and Nutrition Examination Survey	SAMHSA	Substance Abuse and Mental Health Services Administration
NHIS	National Health Interview Survey	SCHIP	State Child Health Insurance Program
NIH	National Institutes of Health	SES	Socioeconomic status
NPHPS	National Public Health Performance Standards	SMART	Specific, measurable, achievable, realistic, and time (objective)
OHRP	Office for Human Research Protections	TAAPS	T ime frame, A mount of what A ctivities done by which P articipants/program S taff
OMB	Office of Management and Budgeting	TQM	Total quality management
OR	Odds ratio	TREW	Time frame, what portion of Recipients experience what Extent of Which type of change
PACE-EH	Protocol for Assessing Excellence in Environmental Health	UOS	Units of service
PAHO	Pan American Health Organization	WHO	World Health Organization
PDCA	Plan-Do-Check-Act	WIC	Special Supplemental Nutrition Program for Women, Infants, and Children
PEARL	Property, economic, acceptability, resource, legality system	YHL	Years of healthy life
PERT	Program Evaluation and Review Technique	YLL	Years of life lost
PPIP	Putting Prevention into Prevention	YPLL	Years of potential life lost
PRECEDE	Predisposing, Reinforcing, and Enabling Factors in Community Education Development and Evaluation (model)		



SECTION I

The Context of Health Program Development

CHAPTER 1

Context of Health Program Development and Evaluation

Health is not a state of being that can easily be achieved through isolated, uninformed, individualistic actions. *Health* of individuals, of families, and of populations is a state in which physical, mental, and social well-being are integrated to enable optimal functioning. From this perspective, achieving and maintaining health across a life span is a complex, complicated, intricate affair. For some, health is present irrespective of any special efforts or intention. For most of us, health requires, at a minimum, some level of attention and specific information. It is through health programs that attention is given focus and information is provided or made available, but that does not guarantee that the attention and information are translated into actions or behaviors needed to achieve health. Thus, those providing health programs, however large or small, need to understand both the processes whereby those in need of attention and health information can receive what is needed, and also the processes by which to learn from the experience of providing the health program.

The processes and effects of health program planning and evaluation are the subjects of this text. The discussion begins here with a brief overview of the historical context. This background sets the stage for appreciating the considerable number of publications on the topic of health program planning and evaluation, and for acknowledging the professionalization of evaluators. The use of the term *processes* to describe the actions involved in health program planning and evaluation is intended to denote action, cycles, and open-endedness. This chapter introduces the planning and evaluation cycle, and the interactions and iterative nature of this cycle are stressed throughout the text. Because health is an individual, aggregate, and population phenomenon, health programs need to be conceptualized across those levels. The public health pyramid, introduced in this chapter, is used throughout the text as a tool for conceptualizing and actualizing health programs for individuals, aggregates, and populations.

► History and Context

An appropriate starting point for this text is reflecting on and understanding what “health” is, along with having a basic appreciation for the genesis of the fields of health program planning and evaluation. A foundation in these elements is key to becoming an evaluation professional.

Concept of Health

To begin the health program planning and evaluation cycle requires first reflecting on the meaning of health. Both explicit and implicit meanings of health can dramatically influence what is considered the health problem and the subsequent direction of a program. The most widely accepted definition of *health* is that put forth by the World Health Organization (WHO), which for the first time defined health as more than the absence of illness and as the presence of well-being (WHO, 1947).

Since the publication of the WHO definition, health has come to be viewed across the health professions as a holistic concept that encompasses the presence of physical, mental, developmental, social, and financial capabilities, assets, and balance. This idea does not preclude each health profession from having a particular aspect of health to which it primarily contributes. For example, a dentist contributes primarily to a patient’s oral health, knowing that the state of the patient’s teeth and gums has a direct relationship to his or her physical and social health. Thus the dentist might say that the health problem is caries. The term *health problem* is used, rather than *illness*, *diagnosis*, or *pathology*, in keeping with the holistic view that there can be problems, deficits, and pathologies in one component of health while the other components remain “healthy.” Using the term *health problem* also makes it easier to think about and plan health programs for aggregates of individuals. A community, a family, and a school can each have a health problem that is the focus of a health program intervention. The extent to which the health program planners have

a shared definition of health and have defined the scope of that definition influences the nature of the health program.

Health is a matter of concern for more than just health professionals. For many Americans, the concept of health is perceived as a right, along with civil rights and liberties. The right to health is often translated by the public and politicians into the perceived right to have or to access health care. This political aspect of health is the genesis of health policy at the local, federal, and international levels. The extent to which the political nature of health underlies the health problem of concern being programmatically addressed also influences the final nature of the health program.

Health Programs, Projects, and Services

What distinguishes a program from a project or from a service can be difficult to explain, given the fluidity of language and terms. The term *program* is fairly generic but generally connotes a structured effort to provide a specific set of services or interventions. In contrast, a *project* often refers to a time-limited or experimental effort to provide a specific set of services or interventions through an organizational structure. In the abstract, a *service* can be difficult to define but generally includes interaction between provider and client, an intangibility aspect to what is provided, and a nonpermanence or transitory nature to what is provided. Using this definition of service, it is easy to see that what is provided in a health program qualifies as a service, although it may not be a health service.

A *health program* is a totality of an organized structure designed for the provision of a fairly discrete health-focused intervention, where that intervention is designed for a specific target audience. By comparison, *health services* are the organizational structures through which providers interact with clients or patients to meet the needs or address the health problems of the clients or patients. Health programs, particularly

in public health, tend to provide educational services, have a prevention focus, and deliver services that are aggregate or population-focused. In contrast, health services exist exclusively as direct services. Recognizing the distinction between health programs and health services is important for understanding the corresponding unique planning and evaluation needs of each.

History of Health Program Planning and Evaluation

The history of planning health programs has a different lineage than that of program evaluation. Only relatively recently, in historical terms, have these lineages begun to overlap, with resulting synergies. Planning for health programs has the older history, if public health is considered. Rosen (1993) argued that public health planning began approximately 4,000 years ago with planned cities in the Indus Valley that had covered sewers. Particularly since the Industrial Revolution, planning for the health of populations has progressed, and it is now considered a key characteristic of the discipline of public health.

Blum (1981) related *planning* to efforts undertaken on behalf of the public well-being to achieve deliberate or intended social change as well as providing a sense of direction and alternative modes of proceeding to influence social attitudes and actions. Others (Dever, 1980; Rohrer, 1996; Turnock, 2004) have similarly defined planning as an intentional effort to create something that has not occurred previously for the betterment of others and for the purpose of meeting desired goals. The purpose of planning is to ensure that a program has the best possible likelihood of being successful, defined in terms of being effective with the least possible resources. Planning encompasses a variety of activities undertaken to meet this purpose.

The quintessential example of planning is the development and use of the *Healthy People* goals. In 1979, *Healthy People* (U.S. Department of Health, Education, and Welfare [DHEW], 1979) was published as an outgrowth of the

need to establish an illness prevention agenda for the United States. The companion publication, *Promoting Health/Preventing Disease* (U.S. Department of Health and Human Services [DHHS], 1980), marked the first time that goals and objectives regarding specific areas of the nation's health were made explicit, with the expectation that these goals would be met by the year 1990. *Healthy People* became the framework for the development of state and local health promotion and disease prevention agendas. Since its initial publication, the U.S. goals for national health have been revised and published as *Healthy People 2000* (DHHS, 1991), *Healthy Communities 2000* (American Public Health Association [APHA], 1991), *Healthy People 2010* (DHHS, 2000), and *Healthy People 2020* (DHHS, 2011), with development of *Healthy People 2030* underway. Other nations also set health status goals and international organizations, such as the World Health Organization (WHO) and Pan American Health Organization (PAHO), develop health goals applicable across nations.

The evolution of *Healthy People* goals also reflects the accelerating rate of emphasis on nationwide coordination of health promotion and disease prevention efforts and a reliance on systematic planning to achieve this coordination. The development of the *Healthy People* publications also reflects the underlying assumption that planning is a rational activity that can lead to results. However, at the end of each 10-year cycle, many of the U.S. health objectives were not achieved, reflecting the potential for planning to fail. Given this failure potential, this text emphasizes techniques to help future planners of health programs to be more realistic in setting goals and less dependent upon a linear, rational approach to planning.

The *Healthy People 1990* objectives were developed by academics and clinician experts in illness prevention and health promotion. In contrast, development of the goals and health problems listed in *Healthy People 2010* and *Healthy People 2020* incorporated ideas generated at public forums and through Internet commentary; these ideas later were revised and refined by expert panels before final publication of the

objectives. Greater participation of the public during the planning stage of health programs has become the norm. In keeping with the emphasis on participation, the role and involvement of stakeholders are stressed at each stage of the planning and evaluation cycle.

The history of evaluation, from which the evaluation of health programs grew, is far shorter than the history of planning, beginning roughly in the early 1900s, but it is equally rich in important lessons for future health program evaluators. The first evaluations were done in the field of education, particularly as student assessment and evaluation of teaching strategies gained interest (Patton, 2008). Assessment of student scholastic achievement is a comparatively circumscribed outcome of an educational intervention. For this reason, early program evaluators came from the discipline of education, and it was from the fields of education and educational psychology that many methodological advances were made and statistics developed.

Guba and Lincoln (1987) summarized the history of evaluations by proposing generational milestones or characteristics that typify distinct generations. Later, Swenson (1991) built on their concept of generations by acknowledging that subsequent generations of evaluations will occur. Each generation incorporates the knowledge of early evaluations and extends that knowledge based on current broad cultural and political trends.

Guba and Lincoln (1987) called the first generation of evaluations in the early 1900s “the technical generation.” During this time, nascent scientific management, statistics, and research methodologies were used to test interventions. Currently, evaluations continue to incorporate the rationality of this generation by using activities that are systematic, science based, logical, and sequential. Rational approaches to evaluations focus on identifying the best-known intervention or strategy given the current knowledge, measuring quantifiable outcomes experienced by program participants, and deducing the degree of effect from the program.

The second generation, which lasted until the 1960s, focused on using goals and objectives

as the basis for evaluation. Second-generation evaluations were predominantly descriptive. With the introduction in the 1960s of broad innovation and initiation of federal social service programs, including Medicare, Medicaid, and Head Start, the focus of evaluations shifted to establishing the merit and value of the programs. Because of the political issues surrounding these and similar federal programs, determining whether the social policies were having any effect on people became a priority. Programs needed to be judged on their merits and effectiveness. The U.S. General Accounting Office (GAO; now called the Government Accountability Office) had been established in 1921 for the purpose of studying the utilization of public finances, assisting Congress in decision making with regard to policy and funding, and evaluating government programs. The second-generation evaluation emphasis on quantifying effects was spurred, in part, by reports from the GAO that were based on the evaluations of federal programs.

Typically, the results of evaluations were not used in the “early” days of evaluating education and social programs. That is, federal health policy was not driven by whether evaluations showed the programs to be successful. Although the scientific rigor of evaluations improved, their usefulness remained minimal. Beginning in the 1980s, however, the third generation of evaluations—termed “the negotiation generation” or “the responsiveness generation”—began. During this generation, evaluators began to acknowledge that they were not autonomous and that their work needed to respond to the needs of those being evaluated. As a result of this awareness, several lineages have emerged. These lineages within the responsiveness generation account for the current diversity in types, emphases, and philosophies related to program evaluation.

One lineage is utilization-focused evaluation (Patton, 2012), in which the evaluator’s primary concern is with developing an evaluation that will be used by the stakeholders. Utilization-focused evaluations are built on the following premises (Patton, 1987): Concern for use of the evaluation pervades the evaluation from beginning to end;

evaluations are aimed at the interests and needs of the users; users of the evaluation must be invested in the decisions regarding the evaluation; and a variety of community, organizational, political, resource, and scientific factors affect the utilization of evaluations. Utilization-focused evaluation differs from evaluations that are focused exclusively on outcomes

Another lineage is participatory evaluation (Whitmore, 1998), in which the evaluation is merely guided by the expert and is actually generated by and conducted by those invested in the health problem. A participatory or empowerment approach invites a wide range of stakeholders into the activity of planning and evaluation, providing those participants with the skills and knowledge to contribute substantively to the activities and fostering their sense of ownership of the product (TABLE 1-1).

The fourth generation of evaluation, which emerged in the mid-1990s, seems to be meta-evaluation, that is, the evaluation of

evaluations done across similar programs. This trend in program evaluation parallels the trend in social science toward using meta-analysis of existing studies to better understand theorized relationships and the trend across the health professions toward establishing evidence-based practice guidelines. This new generation became possible because of a pervasive culture of evaluation in the health services and because of the availability of huge data sets for use in the meta-evaluations. An early example of the evaluation culture was the mandate from United Way, a major funder of community-based health programs, for their grantees to conduct outcome evaluations. To help grantees meet this mandate, United Way published a user-friendly manual (United Way of America, 1996) that could be used by nonprofessionals in the development of basic program evaluations. More broadly, the culture of evaluation can be seen in the explicit requirement of federal agencies that fund community-based health programs that

TABLE 1-1 Comparison of Outcome-Focused, Utilization-Focused, and Participatory Focused Evaluations

	Outcome-Focused Evaluations	Utilization-Focused Evaluations	Participatory Focused Evaluations
Purpose	Show program effect	Get stakeholders to use evaluation-findings for decisions regarding program improvements and future program development	Involve the stakeholders in designing programs and evaluations, and utilizing findings
Audience	Funders, researchers, other external audience	Program people (internal audience), funders	Those directly concerned with the health problem and program
Method	Research methods, external evaluators (usually)	Research methods, participatory	Research methods as implemented by the stakeholders

such programs include evaluations conducted by local evaluators.

Most people have an intuitive sense of what evaluation is. The purpose of evaluation can be to measure the effects of a program against the goals set for it and thus to contribute to subsequent decision making about the program (Weiss, 1972). Alternatively, evaluation can be defined as “the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions” (Rossi, Lipsey, & Freeman, 2004, p. 16). Others (Herman, Morris, & Fitz-Gibbon, 1987) have defined evaluation as judging how well policies and procedures are working or as assessing the quality of a program. These definitions of evaluation all remain relevant.

Inherently these definitions of evaluation have an element of being judged against some criteria. This implicit understanding of evaluation leads those involved with the health program to feel as though they will be judged or found not to meet those criteria and will subsequently experience some form of repercussions. They may fear that they as individuals or as a program will be labeled a failure, unsuccessful, or inadequate. Such feelings must be acknowledged and addressed early in the planning cycle. Throughout the planning and evaluation cycle, program planners have numerous opportunities to engage and involve program staff and stakeholders in the evaluation process. Taking advantage of these opportunities goes a long way in alleviating the concerns of program staff and stakeholders about the judgmental quality of the program evaluation.

► Evaluation as a Profession

A major development in the field of evaluation has been the professionalization of evaluators. The American Evaluation Association (AEA)

serves evaluators primarily in the United States. Several counterparts to the AEA exist, such as the Society for Evaluation in the United Kingdom and the Australian Evaluation Society. The establishment of these professional organizations, whose members are evaluators, and the presence of health-related sections within these organizations demonstrate the existence of a field of expertise and of specialized knowledge regarding the evaluation of health-related programs.

As the field of evaluation has evolved, so have the number and diversity of approaches that can guide the development of evaluations. Currently, 23 different approaches to evaluation have been identified, falling into 3 major groups (Stufflebeam & Coryn, 2014). One group of evaluations is oriented toward questions and methods such as objectives-based studies and experimental evaluations. The second group of evaluations is oriented toward improvements and accountability and includes consumer-oriented and accreditation approaches. The third group of evaluations includes those that have a social agenda or advocacy approach, such as responsive evaluations, democratic evaluations, and utilization-focused evaluation. They also acknowledge pseudo-evaluations and quasi-evaluations as distinct groups, reflecting the continuing evolution of the field of evaluation.

Several concepts are common across the types of evaluations—namely, pluralism of values, stakeholder constructions, fairness and equity regarding stakeholders, the merit and worth of the evaluation, a negotiated process and outcomes, and full collaboration. These concepts have been formalized into the standards for evaluations that were established by the Joint Commission on Standards for Educational Evaluation in 1975 (American Evaluation Association, 2011). Currently, this Joint Commission includes many organizations in its membership, such as the American Evaluation Association and the American Educational Research Association.

The five standards of evaluation established by the American Evaluation Association are utility, feasibility, propriety, accuracy, and evaluation

TABLE 1-2 Evaluation Standards Established by the Joint Commission on Standards for Educational Evaluation

Standard	Description
Utility	To increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.
Feasibility	To increase evaluation effectiveness and efficiency.
Propriety	To support what is proper, fair, legal, right, and just in evaluations.
Accuracy	To increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.
Evaluation accountability	To encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement and accountability for evaluation processes and products.

Data from American Evaluation Association (2012).

accountability (**TABLE 1-2**; American Evaluation Association, 2011).

The utility standard specifies that an evaluation must be useful to those who requested the evaluation. A useful evaluation shows ways to make improvements to the intervention, increase the efficiency of the program, or enhance the possibility of garnering financial support for the program. The feasibility standard denotes that the ideal may not be practical. Evaluations that are highly complex or costly will not be done by small programs with limited capabilities and resources. Propriety is the ethical and politically correct component of the standards. Evaluations can invade privacy or be harmful to either program participants or program staff members. The propriety standard also holds evaluators accountable for upholding all of the other standards. Accuracy is essential and is achieved through the elements that constitute scientific rigor. These established and accepted standards for evaluations reflect current norms

and values held by professional evaluators and deserve attention in health program evaluations. The existence and acceptance of standards truly indicates the professionalism of evaluators.

Achieving these standards requires that those involved in the program planning and evaluation have experience in at least one aspect of planning or evaluation, whether that is experience with the health problem; experience with epidemiological, social, or behavioral science research methods; or skill in facilitating processes that involve diverse constituents, capabilities, and interests. Program planning and evaluation can be done in innumerable ways, with no single “right way.” This degree of freedom and flexibility can feel uncomfortable for some people. As with any skill or activity, until they have experience, program planners and evaluators may feel intimidated by the size of the task or by the experience of others involved. To become a professional evaluator, therefore, requires a degree of willingness to learn, to grow, and to be flexible.

Who Does Planning and Evaluations?

Many different types of health professionals and social scientists can be involved in health program planning and evaluation. At the outset of program planning and evaluation, some trepidation revolves around who ought to be the planners and evaluators. In a sense, almost anyone with an interest and a willingness to be an active participant in the planning or evaluation process could be involved, including health professionals, businesspersons, paraprofessionals, and advocates or activists.

Planners and evaluators may be employees of the organization about to undertake the activity, or they may be external consultants hired to assist in all phases or just a specific phase of the planning and evaluation cycle. Internal and external planners and evaluators each have their advantages and disadvantages. Regardless of whether an internal or external evaluator is used, professional stakes and allegiances ought to be acknowledged and understood as factors that can affect the decision making.

Planners and evaluators from within the organization are susceptible to biases, consciously or not, in favor of the program or some aspect of the program, particularly if their involvement can positively affect their work. On the positive side, internal planners and evaluators are more likely to have insider knowledge of organizational factors that can be utilized or may have a positive effect on the delivery and success of the health program. Internal evaluators may experience divided loyalties, such as between the program and their job, between the program staff members and other staff, or between the proposed program or evaluation and their view of what would be better.

A source of internal evaluators can be members of quality improvement teams, particularly if they have received any training in program development or evaluation as they relate to quality improvement. The use of total quality management (TQM), continuous quality improvement (CQI), and other quality improvement methodologies by healthcare

organizations and public health agencies can be integral to achieving well-functioning programs.

External evaluators can bring a fresh perspective and a way of thinking that generates alternatives not currently in the agencies' repertoire of approaches to the health problem and program evaluation. Compared to internal evaluators, external evaluators are less likely to be biased in favor of one approach—unless, of course, they were chosen for their expertise in a particular area, which would naturally bias their perspective to some extent. External program planners and evaluators, however, can be expensive consultants. Some organizations that specialize in health program evaluations serve as one category of external evaluator. These research firms receive contracts to evaluate health program initiatives and conduct national evaluations that require sophisticated methodology and considerable resources.

The question of who does evaluations also can be answered by looking at who funds health program evaluations. From this perspective, organizations that do evaluations as a component of their business are the answer to the question, Who does evaluations? Although most funding agencies prefer to fund health programs rather than stand-alone program evaluations, some exceptions exist. For example, the Agency for Healthcare Research and Quality (AHRQ) funds health services research about the quality of medical care, which is essentially effect evaluation research. Other federal agencies, such as the National Institutes of Health and the bureaus within the Department of Health and Human Services, fund evaluation research of pilot health programs. However, the funding priorities of these federal agencies change to be consistent with federal health policy. This is a reminder that organizations funding and conducting health program evaluations evolve over time.

Roles of Evaluators

Evaluators may be required to take on various roles, given that they are professionals involved in a process

that very likely involves others. For example, as the evaluation takes on a sociopolitical process, the evaluators become mediators and change agents. If the evaluation is a learning–teaching process, evaluators become both teacher and student of the stakeholders. To the extent that the evaluation is a process that creates a new reality for stakeholders, program staff members, and program participants, evaluators are reality shapers. Sometimes the evaluation may have an unpredictable outcome; at such times, evaluators are human instruments that gauge what is occurring and analyze events. Ideally, evaluations are a collaborative process, and evaluators act as collaborators with the stakeholders, program staff members, and program participants. If the evaluation takes the form of a case study, the evaluators may become illustrators, historians, and storytellers.

These are but a few examples of how the roles of the professional program evaluator evolve and emerge from the situation at hand. The individual's role in the planning and evaluation activities may not be clear at the time that the project is started. Roles will develop and evolve as the planning and evaluation activities progress.

► Planning and Evaluation Cycle

Although planning and evaluation are commonly described in a linear sequential manner, they actually constitute a cyclical process. In this section, the cycle is described along with an emphasis on factors that enhance and detract from that process being effective.

Interdependent and Cyclic Nature of Planning and Evaluation

A major premise running through the current thinking about programs and evaluation is that the activities constituting program planning and program evaluation are cyclical and interdependent

(**FIGURE 1-1**) and that the activities occur more or less in stages or sets of activities. The stages are cyclical to the extent that the end of one program or stage flows almost seamlessly into the next program or planning activity. The activities are interdependent to the extent that the learning, insights, and ideas that result at one stage are likely to influence the available information and thus the decision making and actions of another stage. Interdependence of activities and stages ideally result from information and data feedback loops that connect the stages.

Naturally, not all of the possible interactions among program planning, implementation, and evaluation are shown in Figure 1-1. In reality, the cyclical or interactive nature of health program planning and evaluation exists in varying degrees. In the ideal, interactions, feedback loops, and reiterations of process would be reflected throughout this text. For the sake of clarity, however, the cycle is presented in a linear fashion in the text, with steps and sequences covered in an orderly fashion across the progression of chapters. This pedagogical approach belies the true messiness of health program planning and program evaluation. Because the planning and evaluation cycle is susceptible to and affected by external influences, to be successful as a program planner or evaluator requires a substantial degree of flexibility and creativity in recovering from these influences.

The cycle begins with a trigger event, such as awareness of a health problem; a periodic strategic planning effort; a process required by a stakeholder, such as a 5-year strategic planning process or a grant renewal; or newly available funds for a health program. An indirect trigger for planning could be information generated from an evaluation that reveals either the failure of a health program, extraordinary success of the program, or the need for additional programs. The trigger might also be a news media exposé or legal action. For those seeking to initiate the planning process, getting the attention of influential individuals requires having access to them, packaging the message about the need for planning in ways that are immediately attractive,

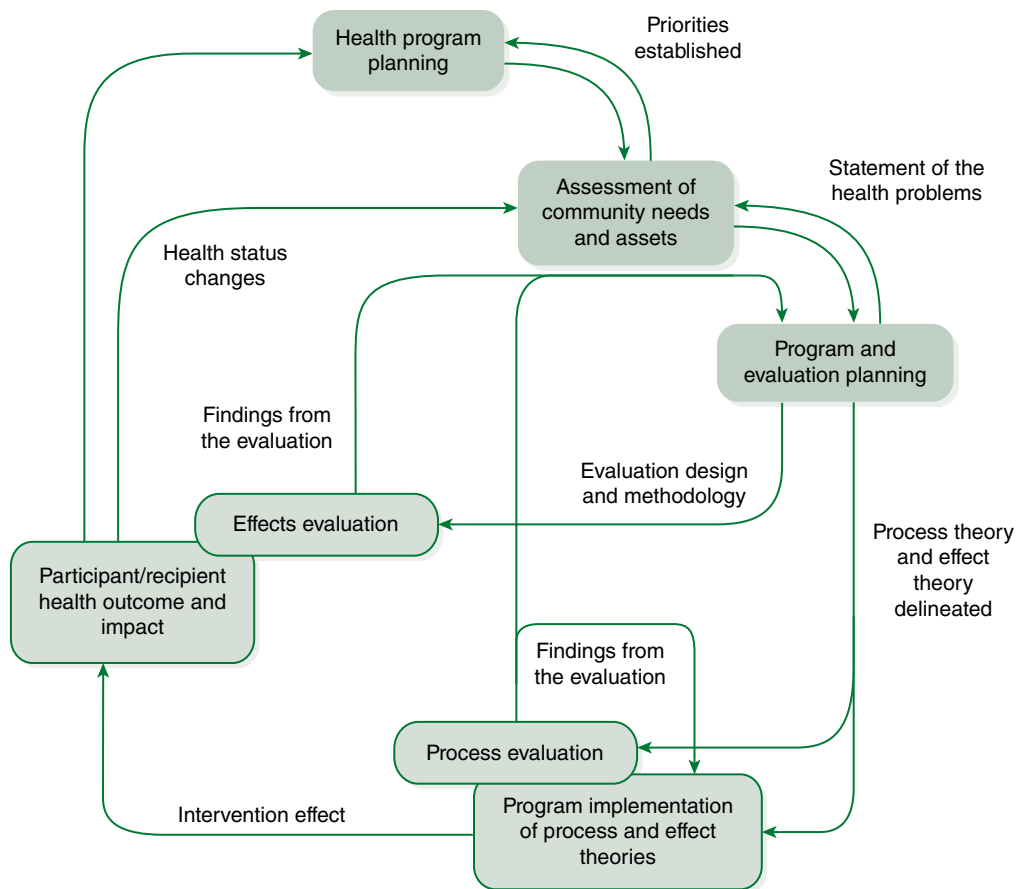


FIGURE 1-1 The Planning and Evaluation Cycle

and demonstrating the salience of the issue. Thus, to get a specific health problem or issue “on the table,” activists can use the salient events to get the attention of influential individuals. The importance of having a salient trigger event is to serve as a reminder that key individuals mentally sort through and choose among competing attention getters. This trigger event or situation leads to the collection of data about the health problem, the characteristics of the people affected, and their perceptions of the health problem. These data, along with additional data on available resources, constitute a community needs and assets assessment.

Based on the data from the needs assessment, program development begins. Problems

and their solutions are prioritized. The planning phase includes developing the program theory, which explicates the connection between what is done and the intended effects of the program. Another component of the planning phase includes assessment of organizational and infrastructure resources for implementing the program, such as garnering resources to implement and sustain the program. Yet another major component of program planning is setting goals and objectives that are derived from the program theory.

After the resources necessary to implement the program have been secured and the activities that make up the program intervention have been explicated, the program can be implemented. The logistics of implementation include marketing

the program to the target audience, training and managing program personnel, and delivering or providing the intervention as planned. During implementation of the program, it is critical to conduct an evaluation of the extent to which the program is provided as planned; this is the process evaluation. The data and findings from the process evaluation are key feedback items in the planning and evaluation cycle, and they can and ought to lead to revisions in the program delivery.

Ultimately, the health program ought to have an effect on the health of the individual program participants or on the recipients of the program intervention if provided to the community or a population. The evaluation can be an outcome evaluation of immediate and closely causally linked programmatic effects or an impact evaluation of more temporally and causally distal programmatic effects. Both types of evaluations provide information to the health program planners for use in subsequent program planning. Evaluation of the effects of the program provides data and information that can be used to alter the program intervention. These findings can also be used in subsequent assessments of the need for future or other health programs.

The model used throughout this text as a framework (Figure 1-1) generically represents the steps and processes. It is one of many possible ways to characterize the planning and evaluation cycle. As a generic representation, the planning and evaluation cycle model used in this text includes the essential elements, but it cannot provide detailed instructions on the “whens” and “hows” because each situation will be slightly different.

Using Evaluation Results as the Cyclical Link

Before embarking on either a process or an effect evaluation, it is important to consider who will use the results because, in being used, evaluation results are perpetuating the program planning and evaluation cycle. The usefulness

of an evaluation depends on the extent to which questions that need to be answered are, in fact, answered. Naturally, different stakeholder groups that are likely to use evaluation findings will be concerned with different questions.

Funding organizations, whether federal agencies or private foundations, constitute one stakeholder group. Funders may use process evaluations for program accountability and effect evaluations for determining the success of broad initiatives and individual program effectiveness. Project directors and managers, another stakeholder group, use both process and effect evaluation findings as a basis for seeking further funding as well as for making improvements to the health program. The program staff members, another stakeholder group, are likely to use both the process and the effect evaluation as a validation of their efforts and as a justification for their feelings about their success with program participants or recipients. Scholars and health professionals constitute another stakeholder group that accesses the findings of effect evaluations through the professional literature. Members of this group are likely to use effect evaluations as the basis for generating new theories about what is effective in addressing a particular health problem and why it is effective.

Policy makers are yet another stakeholder group that uses both published literature and final program reports regarding process and effect evaluation findings when formulating health policy and making decisions about program resource allocation. Community action groups, community members, and program participants and recipients form another group of stakeholders. This stakeholder group is most likely to advocate for a community health assessment and to use process evaluation results as a basis for seeking additional resources or to hold the program accountable.

Program Life Cycle

Feedback loops contribute to the overall development and evolution of a health program, giving it a life cycle. In the early stages of an idea

for a health program, the program may begin as a pilot. At this stage, program development occurs and involves use of literature and needs assessment data (Scheirer, 2012). The program may not rely on any existing format or theory, so simple trial and error is used to determine whether it is feasible as a program. It is likely to be small and somewhat experimental because a similar type of program has not been developed or previously attempted. As the program matures, it may evolve into a model program. A *model program* has interventions that are formalized, or explicit, with protocols that standardize the intervention, and the program is delivered under conditions that are controlled by the program staff members and developers. Model programs can be difficult to sustain over time because of the need to follow the protocols. Evaluations of programs at this stage focus on identifying and documenting the effects and efficacy of the program (Scheirer, 2014). Successful model programs become *institutionalized* within the organization as an ongoing part of the services provided. Successful programs can be institutionalized across a number of organizations in a community to gain wide acceptance as standard practice, with the establishment of an expectation that a “good” agency will provide the program. At this last stage, the health program has become institutionalized within health services. Evaluations tend to focus on quality and performance improvements, as well as sustainability. The last life cycle stage is the dissemination and replication of programs shown to be effective.

Regardless of the stage in a program’s life cycle, the major planning and evaluation stages of community assessment and evaluation are carried out. The precise nature and purpose of each activity vary slightly as the program matures. Being aware of the stage of the program being implemented can help tailor the community assessment and evaluation.

This life cycle of a health program is reflected in the evolution of hospice care. Hospice—care for the dying in a home and family setting—began in London in 1967 as a grassroots service that entailed trial and error about how to manage

dying patients (Kaur, 2000). As its advocates saw the need for reimbursement for the service, they began systematically to control what was done and who was “admitted” to hospice. Once evaluations of these hospice programs began to yield findings that demonstrated their positive benefits, they became the model for more widespread programs that were implemented in local agencies or by new hospice organizations. As hospice programs became accepted as a standard of care for the dying, the hospice programs became standard, institutionalized services for the organization. Today the availability and use of hospice services for terminally ill patients are accepted as standard practice, and most larger healthcare organizations or systems have established a hospice program. The evolution of hospice is but one example of how an idea for a “better” or “needed” program can gradually become widely available as routine care.

► The Fuzzy Aspects of Planning

We like to think of planning as a rational, linear process, with few ambiguities and only the rare dispute. Unfortunately, this is not the reality of health program planning. Many paradoxes inherently exist in planning as well as implicit assumptions, ambiguities, and the potential for conflict. In addition, it is important to be familiar with the key ethical principles that underlie the decision making that is part of planning.

Paradoxes

Several paradoxes pervade health planning (Porter, 2011), which may or may not be resolvable. Those involved can hold assumptions about planning that complicate the act of planning, whether for health systems or programs. Being aware of the paradoxes and assumptions can, however, help program planners understand possible sources of frustration.

One paradox is that planning is shaped by the same forces that created the problems that planning is supposed to correct. Put simply, the healthcare, sociopolitical, and cultural factors that contributed to the health problem or condition are very likely to be same factors that affect the health planning process. The interwoven relationship of health and other aspects of life affects health planning. For example, housing, employment, and social justice affect many health conditions that stimulate planning. This paradox implies that health planning itself is also affected by housing, employment, and social justice.

Another paradox is that the “good” of individuals and society experiencing the prosperity associated with health and well-being is “bad” to the extent that this prosperity also produces ill health. Prosperity in our modern world has its own associated health risks, such as higher cholesterol levels, increased stress, increased risk of cardiovascular disease, and increased levels of environmental pollutants. Also, as one group prospers, other groups often become disproportionately worse off. So, to the extent that health program planning promotes the prosperity of a society or a group of individuals, health issues for others will arise that require health program planning.

A third paradox is that what may be easier and more effective may be less acceptable. A good example of this paradox stems from decisions about active and passive protective interventions. Active protection and passive protection are both approaches to risk reduction and health promotion. *Active protection* requires that individuals actively participate in reducing their risks—for example, through diet changes or the use of motorcycle helmets. *Passive protection* occurs when individuals are protected by virtue of some factor other than their behavior—for example, water fluoridation and mandates for smoke-free workplaces. For many health programs, passive protection in the form of health policy or health regulations may be more effective and efficient. However, ethical and political issues can arise when the emphasis on passive protection, through laws

and communitywide mandates, does not take into account cultural trends or preferences.

Another paradox is that those in need ideally, but rarely, trigger the planning of health programs; rather, health professionals initiate the process. This paradox addresses the issue of who knows best and who has the best ideas for how to resolve the “real” problem. The perspective held by health professionals often does not reflect broader, more common health social values (Reinke & Hall, 1988), including the values possessed by those individuals with the “problem.” Because persons in need of health programs are most likely to know what will work for them, community and stakeholder participation becomes not just crucial but, in many instances, is actually mandated by funding agencies. This paradox also calls into question the role of health professionals in developing health programs. Their normative perspective and scientific knowledge need to be weighed against individuals’ choices that may have caused the health problem.

A corollary to the paradox dealing with the sources of the best ideas is the notion that politicians tend to prefer immediate and permanent cures, whereas health planners prefer long-term, strategic, and less visible interventions (Reinke & Hall, 1988). Generally, people want to be cured of existing problems rather than to think probabilistically about preventing problems that may or may not occur in the future. As a consequence, the prevention and long-term solutions that seem obvious to public health practitioners can conflict with the solutions identified by those with the “problem.”

One reason that the best solutions might come from those with the problem is that health professionals can be perceived as blaming those with the health problem for their problem. Blum (1981), for example, identified the practice of “blaming the victim” as a threat to effective planning. When a woman who experiences domestic violence is said to be “asking for it,” the victim is being blamed. During the planning process, blaming the victim can be implicitly and rather subtly manifested in group settings

through interpretation of data about needs, thereby affecting decisions related to those needs. Having the attitude that “the victim is to blame” can also create conflict and tension among those involved in the planning process, especially if the “victims” are included as stakeholders. The activities for which the victim is being blamed need to be reframed in terms of the causes of those activities or behaviors.

Yet another paradox is the fact that planning is intended to be successful; no one plans to fail. Because of the bias throughout the program planning cycle in favor of succeeding, unanticipated consequences may not be investigated or recognized. The unanticipated consequences of one action can lead to the need for other health decisions that were in themselves unintended (Patrick & Erickson, 1993). To overcome this paradox, brainstorming and thinking creatively at key points in the planning process ought to be fostered and appreciated.

A final paradox of planning, not included on Reinke and Hall’s (1988) list, is that most planning is for making changes, not for creating stability. Yet once a change has been achieved, whether in an individual’s health status or a community’s rates of health problems, the achievement needs to be maintained. Many health programs and health improvement initiatives are designed to be accomplished within a limited time frame, with little or no attention to what happens after the program is completed. To address this paradox requires that planning anticipate the conclusion of a health program and include a plan for sustaining the gains achieved.

Assumptions

Assumptions also influence the effectiveness of planning. The first and primary assumption underlying all planning processes is that a solution, remedy, or appropriate intervention can be identified or developed and provided. Without this assumption, planning would be pointless. It is fundamentally an optimistic assumption about the capacity of the planners, the stakeholders, and the state of the science to address the

health problem. The assumption of possibilities further presumes that the resources available, whether human or otherwise, are sufficient for the task and are suitable to address the health problem. The assumption of adequate capacity and knowledge is actually tested through the process of planning.

A companion assumption is that planning leads to the allocation of resources needed to address the health problem. This assumption is challenged by the reality that four groups of stakeholders have interests in the decision making regarding health resources (Sloan & Conover, 1996) and each group exists in all program planning. Those with the health problem and who are members of the target audience for the health program are one group. Another group of stakeholders is health payers, such as insurance companies and local, federal, and philanthropic funding agencies. The third group is individual healthcare providers and healthcare organizations and networks. Last, the general public is a stakeholder group because it is affected by how resources are allocated for health programs. This list of stakeholder groups highlights the variety of motives each group has for being involved in health program planning, such as personal gain, visibility for an organization, or acquisition of resources associated with the program.

Another assumption about those involved is that they share similar views on how to plan health programs. During the planning process, their points of view and cultural perspectives will likely come into contrast. Hoch (1994) suggested that planners need to know what is relevant and important for the problem at hand. Planners can believe in one set of community purposes and values yet still recognize the validity and merit of competing purposes. He argues that effective planning requires tolerance, freedom, and fairness and that technical and political values are two bases from which to give planning advice. In other words, stakeholders involved in the planning process need to be guided into appreciating and perhaps applying a variety of perspectives about planning.

Each stakeholder group assumes that there are limited resources to be allocated for addressing the health problem and is receptive or responsive to a different set of strategies for allocating health resources. The resulting conflicts among the stakeholders for the limited resources apply whether they are allocating resources across the healthcare system or among programs for specific health problems. Limited resources, whether real or not, raise ethical questions of what to do when possible gains from needed health programs or policies are likely to be small, especially when the health program addresses serious health problems.

It is interesting that, the assumption of limited resources parallels the paradox that planning occurs around what is limited rather than what is abundant. Rarely is there a discussion of the abundant or unlimited resources available for health planning. Particularly in the United States, we have an amazing abundance of volunteer hours and interest and of advocacy groups and energy, and recently retired equipment that may be appropriate in some situations. Such resources, while not glamorous or constituting a substantial entry on a balance sheet, deserve to be acknowledged in the planning process.

Another assumption about the planning process is that it occurs in an orderly fashion and that a rational approach is best. To understand the implications of this assumption, one must first acknowledge that four key elements are inherent in planning: uncertainty, ambiguity, risk, and control. The presence of each of these elements contradicts the assumption of a rational approach, and each generates its own paradoxes.

Uncertainty, Ambiguity, Risk, and Control

Despite the orderly approach implied by use of the term *planning*, this process is affected by the limits of both scientific rationality and the usefulness of data to cope with the uncertainties, ambiguities, and risks being addressed by the planning process (see **TABLE 1-3**).

Uncertainty is the unknown likelihood of a possible outcome. Rice, O'Connor, and Pierantozzi (2008) have identified four types of uncertainty: types and amount of resources, technological, market receptivity to the product, and organizational. Each of these uncertainties is present in planning health programs. *Ambiguity* is doubt about a course of action stemming from awareness that known and unknown factors exist that can decrease the possibility of certainty. In this sense, ambiguity results in uncertainty. Both uncertainty and ambiguity pervade the planning process because it is impossible to know and estimate the effect of all relevant factors—from all possible causes of the health problem, to all possible health effects from program interventions, to all possible acts and intentions of individuals. A rational approach to planning presumes that all relevant factors can be completely accounted for by anticipating the effect of a program, but our experiences as humans tell us otherwise.

Ambiguity is the characteristic of not having a clear or single meaning. Change, or the possibility of change, is a possible source of ambiguity. When ambiguity is ignored, the resulting differences in interpretation can lead to confusion and conflict among stakeholders and planners, among planners and those with the health problem, and among those with various health problems vying for resources. The conflict, whether subtle and friendly or openly hostile, detracts from the planning process by requiring time and personnel resources to address and resolve the conflict. Nonetheless, openly and constructively addressing the ambiguity and any associated conflict can lead to innovations in the program.

Risk is the perceived possibility or uncertain probability of an adverse outcome in a given situation. Health planners need to be aware of the community's perception and interpretation of probabilities as they relate to health and illness. Risk is not just about taking chances (e.g., bungee jumping or having unprotected sex) but is also about uncertainty and ambiguity (as is the case with estimates of cure rates and projections about future health conditions).

TABLE 1-3 Fuzzy Aspects Throughout the Planning and Evaluation Cycle

	Stages in the Planning and Evaluation Cycle			
	Community Assessment	Planning	Implementation	Effect Evaluation
Uncertainty	Unknown likelihood of finding key health determinants	Unknown likelihood of selecting an effective intervention, unknown likelihood of the intervention being effective	Unknown likelihood of the intervention being provided as designed and planned	Unknown likelihood of intervention being effective
Ambiguity	Unclear about who is being assessed or why	Unclear about the process, who is leading planning process, or what it is intended to accomplish	Unclear about the boundaries of the program, who ought to participate, or who ought to deliver the program	Unclear about meaning of the evaluation results
Risk	Unknown possibility of the assessment causing harm	Unknown possibility of planning touching on politically sensitive issues	Unknown possibility of the intervention having an adverse effect on participants	Unknown possibility of adverse effect from the evaluation design, or from misinterpretation of the findings
Control	Directing the process of gathering and interpreting data about the health problem	Directing the decisions about the program	Directing the manner in which the program is provided	Directing the process of data collection, analysis and interpretation

Risk is pervasive and inherent throughout the planning process in terms of deciding who to involve and how, which planning approach to use, which intervention to use, and in estimating which health problem deserves attention. The importance of understanding risk as an element both of the program planning process and of

the target audience provides planners with a basis from which to be flexible and speculative.

Control, as in being in charge of or managing, is a natural reaction to the presence of ambiguity, conflict, and risk. It can take the form of directing attention and allocating resources or of exerting dominance over others. Control

remains a key element of management. In other words, addressing the ambiguity, uncertainty, and risk that might have been the trigger for the planning process requires less—not more—control. Those who preside over and influence the planning process are often thought of as having control over solutions to the health problem or condition. They do not. Instead, effective guidance of the planning process limits the amount of control exerted by any one stakeholder and addresses the anxiety that often accompanies the lack of control.

► Introduction to the Types of Evaluation

Several major types of activities are classified as evaluations. Each type of activity requires a specific focus, purpose, and set of skills. The types of evaluations are introduced here as an overview of the field of planning and evaluation.

Community needs assessment (also known as community health assessment) is a type of evaluation that is performed to collect data about the health problems of a particular group. The data collected for this purpose are then used to tailor the health program to the needs and distinctive characteristics of that group. A community needs assessment is a major component of program planning because it is, done at an early stage in the program planning and evaluation cycle. In addition, the regular completion of community assessments may be required. For example, many states do 5-year planning of programs based on state needs assessments.

Another type of evaluation begins at the same time that the program starts. *Process evaluations* focus on the degree to which the program has been implemented as planned and on the quality of the program implementation. Process evaluations are known by a variety of terms, such as monitoring evaluations, depending on their focus and characteristics. The underlying framework for designing a process evaluation comes from the process theory component of

the overall program theory developed during the planning stage. The *process theory* delineates the logistical activities, resources, and interventions needed to achieve the health change in program participants or recipients. Information from the process evaluation is used to plan, revise, or improve the program.

The third type of evaluation seeks to determine the effect of the program—in other words, to demonstrate or identify the program's effect on those who participated in the program. *Effect evaluations* answer a key question: Did the program make a difference? The effect theory component of the program theory is used as the basis for designing this evaluation. Evaluators seek to use the most rigorous and robust designs, methods, and statistics possible and feasible when conducting an effect evaluation. Findings from effect evaluations are used to revise the program and may be used in subsequent initial program planning activities. Effect evaluations may be referred to as outcome or impact evaluations, terms which seem to be used interchangeably in the literature. For clarity, *outcome evaluations* focus on the more immediate effects of the program, whereas *impact evaluations* may have a more long-term focus. Program planners and evaluators must be vigilant with regard to how they and others are using terms and should clarify meanings and address misconceptions or misunderstandings.

A fourth type of evaluation focuses on efficiency and the costs associated with the program. *Cost evaluations* encompass a variety of more specific cost-related evaluations—namely, cost-effectiveness evaluations, cost-benefit evaluations, and cost-utility evaluations. For the most part, cost evaluations are done by researchers because cost-benefit and cost-utility evaluations, in particular, require expertise in economics. Nonetheless, small-scale and simplified cost-effectiveness evaluations can be done if good cost accounting has been maintained by the program and a more sophisticated outcome or impact evaluation has been conducted. The similarities and differences among these three types of cost studies are reviewed in greater detail

in the text so that program planners can be, at minimum, savvy consumers of published reports of cost evaluations. Because cost evaluations are performed late in the planning and evaluation cycle, their results are not likely to be available in time to make program improvements or revisions. Instead, such evaluations are generally used during subsequent planning stages to gather information for prioritizing program options.

Comprehensive evaluations, the fifth type of evaluation, involve analyzing needs assessment data, process evaluation data, effect evaluation data, and cost evaluation data as a set of data. Given the resources needed to integrate analysis of various types of data to draw conclusions about the effectiveness and efficiency of the program, comprehensive evaluations are relatively uncommon. A sixth type of evaluation is a *meta-evaluation*. A meta-evaluation is done by combining the findings from previous outcome evaluations of various programs for the same health problem. The purpose of a meta-evaluation is to gain insights into which of the various programmatic approaches has had the most effect and to determine the maximum effect that a particular programmatic approach has had on the health problem. This type of evaluation relies on the availability of existing information about evaluations and on the use of a specific set of methodological and statistical procedures. For these reasons, meta-evaluations are less likely to be done by program personnel; instead, they are generally carried out by evaluation researchers. Meta-evaluations that are published are extremely useful in program planning because they indicate which programmatic interventions are more likely to succeed in having an effect on the participants. Published meta-evaluations can also be valuable in influencing health policy and health funding decisions.

Summative evaluations, in the strictest sense, are done at the conclusion of a program to provide a conclusive statement regarding program effects. Unfortunately, the term *summative evaluation* is sometimes used to refer to either an outcome or impact evaluation, adding even more confusion to the evaluation terminology and vernacular language. Summative evaluations

are usually contrasted with *formative evaluations*. The term *formative evaluation* is used to refer to program assessments that are performed early in the implementation of the program and used to make changes to the program. Formative evaluations might include elements of process evaluation and preliminary effect evaluations.

Mandated and Voluntary Evaluations

Evaluations are not spontaneous events. Rather, they are either mandated or voluntary. A mandate to evaluate a program is always linked in some way to the funding agencies, whether a governmental body or a foundation. If an evaluation is mandated, then the contract for receiving the program funding will include language specifying the parameters and time line for the mandated evaluation. The mandate for an evaluation may specify whether the evaluation will be done by project staff members or external evaluators, or both. For example, the State Child Health Insurance Program (SCHIP), created in 1998, is a federally funded and mandated program to expand insurance coverage to children just above the federal poverty level. Congress has the authority to mandate evaluations of federal programs and did just that with the SCHIP. Mandated evaluations of SCHIP include an overall evaluation study by Wooldridge and associates from the Urban Institute (2003), and an evaluation specifically focused on outcomes for children with special healthcare needs (Zickafoose, Smith, & Dye, 2015).

Other evaluations may be linked to accreditation that is required for reimbursement of services provided, making them *de facto* mandated evaluations. For example, to receive accreditation from the Joint Commission, a health services organization must collect data over time on patient outcomes. These data are then used to develop ongoing quality improvement efforts. A similar process exists for mental health agencies. The Commission on Accreditation of Rehabilitation Facilities (CARF) requires that

provider organizations conduct a self-evaluation as an early step in the accreditation process. These accreditation-related evaluations apply predominantly to direct care providers rather than to specific programs.

Completely voluntary evaluations are initiated, planned, and completed by the project staff members in an effort to make improvements. However, given the relatively low reward from, and cost associated with, doing an evaluation when it is not required, these evaluations are likely to be small with low scientific rigor. Programs that engage voluntarily in evaluations may have good intentions, but they often lack the skills and knowledge required to conduct an appropriate evaluation.

When Not to Evaluate

Situations and circumstances that are not amenable to conducting an evaluation do exist, despite a request or the requirement for having an evaluation. Specifically, it is not advisable to attempt an evaluation under the following four circumstances: when there are no questions about the program, when the program has no clear direction, when stakeholders cannot agree on the program objectives, and when there is not enough money to conduct a sound evaluation (Patton, 2008). In addition to these situations, Weiss (1972) recognized that sometimes evaluations are requested and conducted for less than legitimate purposes, namely, to postpone program or policy decisions, thereby avoiding the responsibility of making the program or policy decision; to make a program look good as a public relations effort; or to fulfill program grant requirements. As these lists suggest, those engaged in program planning and evaluation need to be purposeful in what is done and should be aware that external forces can influence the planning and evaluation processes.

Since Weiss made her observation in 1972, funders have begun to require program process and effect evaluations, and conducting these evaluations to meet that requirement is considered quite legitimate. This change has occurred

as techniques for designing and conducting both program process and effect evaluations have improved, and the expectation is that even mandated evaluations will be useful in some way. Nonetheless, it remains critical to consider how to conduct evaluations legitimately, rigorously, inexpensively, and fairly. In addition, if the AEA standards of utility, feasibility, propriety, and accuracy cannot be met, it is not wise to conduct an evaluation (Patton, 2008).

Interests and the degree of influence held by stakeholders can change. Such changes affect not only how the evaluation is conceptualized but also whether evaluation findings are used. In addition, the priorities and responsibilities of the organizations and agencies providing the program can change during the course of delivering the program, which can then lead to changes in the program implementation that have not been taken into account by the evaluation. For example, if withdrawal of resources leads to a shortened or streamlined evaluation, subsequent findings may indicate a failure of the program intervention. However, it will remain unclear whether the apparently ineffective intervention was due to the design of the program or the design of the evaluation. In addition, unanticipated problems in delivering the program interventions and the evaluation will always exist. Even rigorously designed evaluations face challenges in the real world stemming from staff turnover, potential participants' noninvolvement in the program, bad weather, or any of a host of other factors that might hamper achieving the original evaluation design. Stakeholders will need to understand that the evaluator attempted to address challenges as they arose if they are to have confidence in the evaluation findings.

► The Public Health Pyramid

Pyramids tend to be easy to understand and work well to capture tiered concepts. For these reasons, pyramids have been used to depict the

tiered nature of primary healthcare, secondary healthcare, and tertiary healthcare services (U.S. Public Health Service, 1994), the inverse relationship of effort needed and health impact of different interventions (Frieden, 2010), and nutrition recommendations (Gil, Ruiz-Lopez, Fernandez-Gonzalez, & de Victoria, 2014).

The public health pyramid is divided into four sections (FIGURE 1-2). The top, or the first, section of the pyramid contains direct healthcare services, such as medical care, psychological counseling, hospital care, and pharmacy services. At this level of the pyramid, programs are delivered to individuals, whether patients, clients, or even students. Generally, programs at the direct services level have a direct, and often relatively immediate, effect on individual participants in the health program. Direct services of these types appear at the tip of the pyramid to reflect that, overall, the smallest proportion of a population receives them. These interventions, according to the Health Impact Pyramid (Frieden, 2010), require considerable effort, with minimal population effects.

At the second level of the pyramid are enabling services, which are those health and social services that support or enhance the health of aggregates. *Aggregates* are used to distinguish between individuals and populations; they are groups of individuals who share a defining characteristic, such as mental illness or a terminal disease. Examples of enabling services include

mental health drop-in centers, hospice programs, financial assistance programs that provide transportation to medical care, community-based case management for patients with acquired immune deficiency syndrome (AIDS), low-income housing, nutrition education programs provided by schools, and workplace child care centers. As this list of programs demonstrates, the services at this level may directly or indirectly contribute to the health of individuals, families, and communities and are provided to aggregates. Enabling services can also be thought of as addressing some of the consequences of social determinants of health.

The next, more encompassing level of the public health pyramid is population-based services. At the population level of the pyramid, services are delivered to an entire population, such as all persons residing in a city, state, or country. Examples of population services include immunization programs for all children in a county, newborn screening for all infants born in a state, food safety inspections carried out under the auspices of state regulations, workplace safety programs, nutrition labeling on food, and the Medicaid program for pregnant women whose incomes fall below the federal poverty guidelines. As this list reflects, the distinction between an aggregate and a population can be blurry. Programs at this level typically are intended to reach an entire population, sometimes without the conscious involvement of individuals. In this sense, individuals receive a population-based health program, such as water fluoridation, rather than participating in the program, as they would in a smoking-cessation class. Interventions and programs aimed at changing the socioeconomic context within which populations live would be included at this population level of the pyramid. Such programs are directed at changing one or more social determinants of health. Population-level programs contribute to the health of individuals and, cumulatively, to the health status of the population.

Supporting the pyramid at its base is the infrastructure of the healthcare system and the public health system. The health services at the other pyramid levels would not be possible

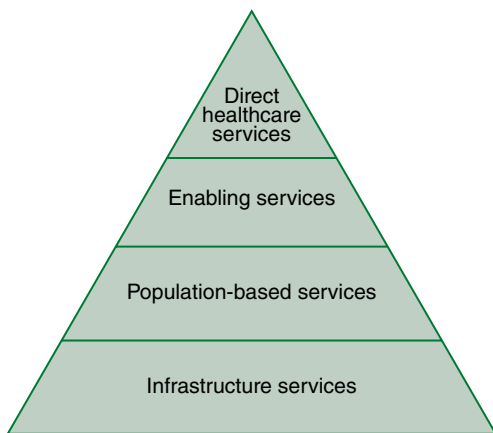


FIGURE 1-2 The Public Health Pyramid

unless there were skilled, knowledgeable health professionals; laws and regulations pertinent to the health of the people; quality assurance and improvement programs; leadership and managerial oversight; health planning and program evaluation; information systems; and technological resources. The planning and evaluation of health programs at the direct, enabling, and population services levels is itself a component of the infrastructure; these are infrastructure activities. In addition, planning programs to address problems of the infrastructure, as well as to evaluate the infrastructure itself, are needed to keep the health and public health system infrastructure strong, stable, and supportive of the myriad of health programs.

Use of the Public Health Pyramid in Program Planning and Evaluation

Health programs exist across the pyramid levels, and evaluations of these programs are needed. However, at each level of the pyramid, certain issues unique to that level must be addressed in developing health programs. Accordingly, the types of health professionals and the types of expertise needed vary by pyramid level, reinforcing the need to match program, participants, and providers appropriately. Similarly, each level of the pyramid is characterized by unique challenges for evaluating programs. For this reason, the public health pyramid, as a framework, helps illuminate those differences, issues, and challenges, as well as to reinforce that health programs are needed across the pyramid levels if the *Healthy People 2020* goals and objectives are to be achieved.

In a more general sense, the public health pyramid provides reminders that various aggregates of potential audiences exist for any health problem and program and that health programs are needed across the pyramid. Depending on the health discipline and the environment in which the planning is being done, direct service programs may be the natural or only inclination. The public health pyramid, however, provides a framework for balancing the level

of the program with meeting the needs of the broadest number of people with a given need. Reaching the same number of persons with a direct services program as with a population services program poses additional expense and logistic challenges.

The pyramid also serves as a reminder that stakeholder alignments and allegiances may be specific to a level of the pyramid. For example, a school health program (an enabling-level program) has a different set of constituents and concerned stakeholders than a highway safety program (a population-level program). The savvy program planner considers not only the potential program participants at each level of the pyramid but also the stakeholders who are likely to make themselves known during the planning process.

The public health pyramid has particular relevance for public health agencies concerned with addressing the three core functions of public health (Institute of Medicine, 1988): assessment, assurance, and policy. These core functions are evident, in varying forms, at each level of the pyramid. Similarly, the pyramid can be applied to the strategic plans of organizations in the private healthcare sector. For optimal health program planning, each health program being developed or implemented ought to be considered in terms of its relationship to services, programs, and health needs at other levels of the pyramid. For all these reasons, the public health pyramid is used throughout this text as a framework for summarizing specific issues and applications of chapter content to each level of the pyramid and to identify and discuss potential or real issues related to the topic of the chapter.

The Public Health Pyramid as an Ecological Model

Individual behavior and health are now understood to be influenced by the social and physical environment of individuals. This recognition is reflected in the growing use of the ecological approach to health services and public health

programs. The ecological approach, which stems from systems theory applied to individuals and families (Bronfenbrenner, 1970, 1989), postulates that individuals can be influenced by factors in their immediate social and physical environment. This perspective has been expanded into the social determinants perspective in public health, which has wide acceptance (Frieden, 2010). The individual is viewed as a member of an intimate social network, usually a family, which is a member of a larger social network, such as a neighborhood or community. The way in which individuals are nested within these social networks has consequences for the health of the individual.

Because it distinguishes and recognizes the importance of enabling and population services, the public health pyramid can be integrated with an ecological view of health and health problems. If one were to look down on the pyramid from above, the levels would appear as concentric circles (**FIGURE 1-3**)—direct services for individuals nested within enabling services for families, aggregates, and neighborhoods, which are in turn nested within population services for all residents of cities, states, or countries. This is similar to individuals being nested within the enabling environment of their family, workplace setting, or neighborhood, all of which are nested within

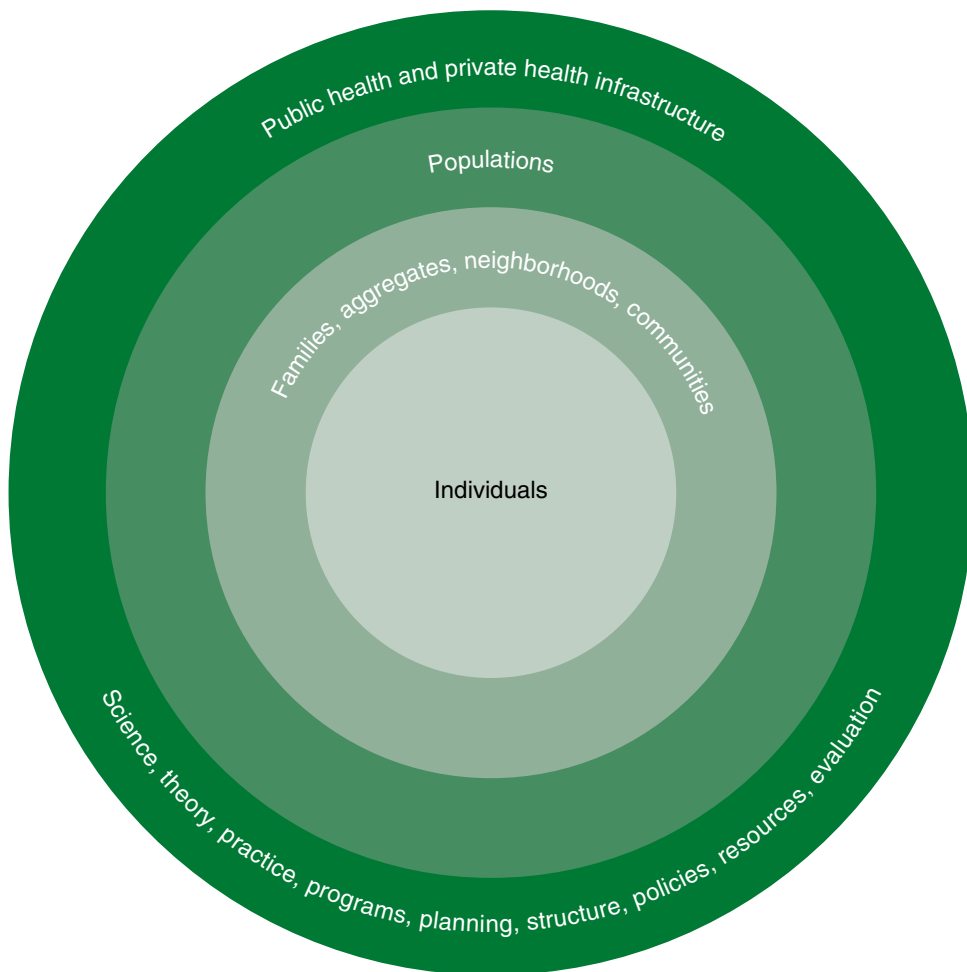


FIGURE 1-3 The Pyramid as an Ecological Model

the population environment of factors such as social norms and economic and political environments. The infrastructure of the healthcare system and public health system is the foundation and supporting environment for promoting health and preventing illnesses and diseases.

The end of the chapter presents a summary of challenges or issues related to applying the chapter content to each level of the pyramid. This feature reinforces the message that each level of the pyramid has value and importance to health program planning and evaluation. In addition, certain unique challenges are specific to each level of the pyramid. The chapter summary by levels offers an opportunity to acknowledge and address the issues related to the levels.

► The Town of Layetteville in Bowe County

As an aid to understanding and assimilating the content covered, examples from the literature are provided throughout this book. In addition, chapters include application of content to a hypothetical town (Layetteville) in an imaginary county (Bowe County). Based on a fictional community needs assessment, subsequent prioritization leads to the identification of five health problems as foci for health program planning. These health problems are used throughout the text as opportunities to demonstrate application of the chapter content. Also, some discussion questions and activities use Layetteville and Bowe County as opportunities for the reader to practice applying the chapter content. While the town and county are fictitious, the health problems around which the program planning and evaluation occur are very real and relevant.

► Across the Pyramid

At the direct services level, health program planning and evaluation focus on individual clients

or patients—that is, on developing programs that are provided to those individuals and on assessing the extent to which those programs make a difference in the health of the individuals who receive the health program. Health is defined in individual terms, and program effects are measured as individual changes. From this level of the public health pyramid, community is most likely viewed as the context affecting individual health.

At the enabling services level, health program planning and evaluation focus on the needs of aggregates of individuals and on the services that the aggregate needs to maintain health or make health improvements. Enabling services are often social, educational, or human services that have an indirect effect on health, thus warranting their inclusion in planning health programs. Health continues to be defined and measured as an individual characteristic to the extent that enabling services are provided to individual members of the aggregate. However, program planning and evaluation focus not on individuals but rather on the aggregate as a unit. At this level of the pyramid, community can be either the aggregate that is targeted for a health program or the context in which the aggregate functions and lives. How community is viewed depends on the health problem being addressed.

At the population-based services level, health program planning and evaluation focus on the needs of all members of a population. At this level of the pyramid, health programs are, at a minimum, population driven, meaning that data collected in regard to the health of the population drive the decisions about the health program. This approach results in programs that are population focused and, ideally (but not necessarily), population based. It is worth noting that population-focused programs tend to have a health promotion or health maintenance focus rather than a focus on treatment of illnesses. At a population level, health is defined in terms of population statistics, such as mortality and morbidity rates. In this regard, the *Healthy People 2020* objectives (**TABLE 1-4**) are predominantly at

TABLE 1-4 A Summary of the *Healthy People 2020* Priority Areas

1. Access to quality health services	21. Heart disease and stroke
2. Adolescent health	22. Human immunodeficiency virus infection (HIV)
3. Arthritis, osteoporosis, and chronic back conditions	23. Immunization and infectious diseases
4. Blood disorders and blood safety	24. Injury and violence prevention
5. Cancer	25. Lesbian, gay, bisexual, and transgender health
6. Chronic kidney disease	26. Maternal, infant, and child health
7. Dementias, including Alzheimer's	27. Medical product safety
8. Diabetes	28. Mental health and mental disorders
9. Disability and secondary conditions	29. Nutrition and weight status
10. Early and middle childhood	30. Occupational safety and health
11. Educational and community-based programs	31. Older adults
12. Environmental health	32. Oral health
13. Family planning	33. Physical activity
14. Food safety	34. Preparedness
15. Genomics	35. Public health infrastructure
16. Global health	36. Respiratory diseases
17. Health communication and health information technology	37. Sexually transmitted diseases
18. Healthcare-associated infections	38. Sleep health
19. Health-related quality of life and well-being	39. Social determinants of health
20. Hearing and other sensory or communication disorders	40. Substance abuse
	41. Tobacco use
	42. Vision

Source: Department of Health and Human Services. Retrieved August 20, 2016, from www.cdc.gov/nchs/healthy_people/hp2020/hp2020_topic_areas.htm.

the population level of the public health pyramid. Community is more likely to be the population targeted by the health program.

At the infrastructure level, health program planning and evaluation are infrastructure activities of both the public health system and the healthcare system. Infrastructure includes organizational management, acquisition of resources, and development of health policy. A significant document reflecting health policy is *Healthy People 2020*, which outlines the goals and objectives for the health of the people of the United States. These national objectives are considered when setting priorities and are used by many federal and nongovernmental funding agencies, which often require that a

health program identify which *Healthy People 2020* objectives are being addressed. To the extent that health planners and evaluators are familiar with these objectives, they will be better able to design appropriate programs and then to argue in favor of the relevance of each of those programs. At the infrastructure level, health can be defined in terms of the individual workers in the healthcare sector (an aggregate). More to the point, because program planning and evaluation are infrastructure activities, it is actually at the infrastructure level that the decisions are made on the definition of health to be used in the program. Similarly, the way that community is viewed is determined at the infrastructure level.

► Discussion Questions

1. When and under what conditions might it be advisable not to conduct an evaluation?
2. Oral health is a major health problem, especially for children living in poverty. Describe how an oral health program developed at each level of the public health pyramid would differ and how the considerations would differ.
3. Conduct a literature search using words such as “planning,” “evaluation,” “program,” and a health condition of interest to you. Which journals publish articles about health program planning and health program evaluations? What are the current trends in the field as reflected in the published literature that you reviewed?
4. Review the program planning and evaluation cycle in Figure 1-1. Using the literature you found for Discussion Question 3, where does most of the literature fall in that cycle?
5. Access and review the material in the following document and compare it with the perspective given in this chapter: Centers for Disease Control and Prevention. (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report*, 48(RR-11): i–41. Retrieved August 31, 2012, from www.cdc.gov/mmwr/preview/mmwrhtml/rr4811a1.htm.

► Internet Resources

American Evaluation Association (AEA)

This international, professional organization of evaluators is devoted to the application and exploration of program evaluation, personnel evaluation, technology, and many other forms of evaluation. The AEA website (www.eval.org/) includes links to professional groups, foundations,

online publications, and other resources related to evaluation.

Centers for Disease Control and Prevention

Centers for Disease Control and Prevention. (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report*, 48(RR-11): i–41. Retrieved August 31, 2012, from www.cdc.gov/mmwr/preview/mmwrhtml/rr4811a1.htm. This online textbook describes the steps involved in conducting an evaluation.

Evaluation Center of Western Michigan University

This organization focuses on advancing the theory and practice of program, personnel, and student/constituent evaluation, as applied primarily to education and human services. Its website (<http://www.wmich.edu/evaluation>) has links and a set of checklists that can be used to improve the quality of an evaluation project.

The Evaluation Exchange

Harvard Family Research Project’s evaluation periodical, *The Evaluation Exchange*, addresses current issues facing program evaluators of all levels, with articles written by the most prominent evaluators in the field. Designed as an ongoing discussion among evaluators, program practitioners, funders, and policy makers, *The Evaluation Exchange* highlights innovative methods and approaches to evaluation, emerging trends in evaluation practice, and practical applications of evaluation theory. It goes out to its subscribers free of charge four times per year. It can be accessed via the Internet at <http://www.hfrp.org/evaluation/the-evaluation-exchange>.

References

- American Evaluation Association. (2011). *The program evaluation standards: Summary form*. Retrieved August 31, 2016, from <http://www.eval.org/p/cm/ld/fid=103>
- American Public Health Association (APHA). (1991). *Healthy communities 2000: Model standards*. Washington, DC: Author.

- Blum, H. L. (1981). *Planning for health: Generics for the eighties* (2nd ed.). New York, NY: Human Sciences Press.
- Bronfenbrenner, U. (1970). *Two worlds of childhood*. New York, NY: Russell Sage Foundation.
- Bronfenbrenner, U. (1989). Ecological systems theory. *Annals of Child Development*, 16, 187–249.
- Dever, G. E. (1980). *Community health analysis: A holistic approach*. Germantown, MD: Aspen.
- Frieden, T. R. (2010). A framework for public health action: The health impact pyramid. *American Journal of Public Health*, 100, 590–595.
- Gil, A., Ruiz-Lopez, M. D., Fernandez-Gonzalez, M., & de Victoria, E. M. (2014). The FINUT Health Lifestyles Guide: Beyond the food pyramid. *Advances in Nutrition*, 5, 358S–367S.
- Guba, E. G., & Lincoln, Y. S. (1987). Fourth generation evaluation. In D. J. Palumbo (Ed.), *The politics of program evaluation* (pp. 202–204). Newbury Park, CA: Sage.
- Herman, J. L., Morris, L. L., & Fitz-Gibbon, C. T. (1987). *Evaluators' handbook*. Newbury Park, CA: Sage.
- Hoch, C. (1994). *What planners do: Power, politics and persuasion*. Chicago, IL: Planners Press.
- Institute of Medicine, National Academy of Sciences. (1988). *The future of public health*. Washington, DC: National Academy Press.
- Kaur, J. (2000). Palliative care and hospice programs. *Mayo Clinic Proceedings*, 75, 181–184.
- Patrick, D. L., & Erickson, P. (1993). *Health status and health policy: Allocating resources to health care*. Oxford, UK: Oxford University Press.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage.
- Patton, M. Q. (2008). *Utilization-focused evaluation*: Sage.
- Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Thousand Oaks, CA: Sage.
- Porter, L. (2011). Informality, the commons and the paradoxes for planning: Concepts and debates for informality and planning. *Planning Theory and Practice*, 12(1), 115–153.
- Reinke, W. A., & Hall, T. L. (1988). Political aspects of planning. In W. A. Reinke (Ed.), *Health planning for effective management* (pp. 75–85). New York, NY: Oxford University Press.
- Rice, M. P., O'Connor, G. C., & Pierantozzi, R. (2008). Implementing a learning plan to counter project uncertainty. *Sloan Management Review*, 29(2), 54–62.
- Rohrer, J. (1996). *Planning for community-oriented health systems*. Washington, DC: American Public Health Association.
- Rosen, G. (1993). *A history of public health* (expanded ed.). Baltimore, MD: Johns Hopkins University Press.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Newbury Park, CA: Sage.
- Scheirer, M. A. (2012). Planning evaluation through the program life cycle. *American Journal of Evaluation*, 33(2), 263–294. doi:10.1177/1098214011434609
- Sloan, F. A., & Conover, C. J. (1996). The use of cost-effectiveness/cost-benefit analysis in actual decision making: Current status and prospects. In F. A. Sloan (Ed.), *Valuing health care: Costs, benefits, and effectiveness of pharmaceuticals and other medical technologies* (pp. 207–232). Cambridge, UK: Cambridge University Press.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models and applications* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Swenson, M. M. (1991). Using fourth-generation evaluation. *Evaluation and Health Professions*, 14(1), 79–87.
- Turnock, B. (2004). *Public health: What it is and how it works* (3rd ed.). Sudbury, MA: Jones & Bartlett.
- United Way of America. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: Author.
- U.S. Department of Health and Human Services (DHHS). (1980). *Promoting health/preventing disease: Objectives for the nation*. Washington, DC: Author.
- U.S. Department of Health and Human Services (DHHS). (1991). *Healthy People 2000: National health promotion and disease prevention objectives*. Publication No. (PHS) 91-50212. Washington, DC: Author.
- U.S. Department of Health and Human Services (DHHS). (2000). *Healthy People 2010: Understanding and improving health* (2nd ed.). Washington, DC: U.S. Government Printing Office.
- U.S. Department of Health and Human Services (DHHS). (2011). *Healthy People 2020*. Retrieved November 9, 2011, from www.healthypeople.gov/2020/default.aspx
- U.S. Department of Health, Education, and Welfare (DHEW). (1979). *Healthy people: The surgeon general's report on health promotion and disease prevention* (DHEW, PHS Publication No. 79-55071). Washington, DC: Author.
- U.S. Public Health Service. (1994). *For a healthy nation: Return on investment in public health*. Washington, DC: Author.
- Weiss, C. (1972). *Evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Whitmore, E. (Ed.). (1998). *Understanding and practicing participatory evaluation: New directions for evaluation*. San Francisco, CA: Jossey-Bass.
- Wooldridge, J., Hill, I., Harrington, M., Kenney, G. M., Hawkes, C., & Haley, J. M. (2003). *Interim evaluation report: Congressionally mandated evaluation of the State Children's Health Insurance Program*. Washington, DC: Urban Institute. Retrieved August 31, 2011, from www.urban.org/expert.cfm?ID=JudithWooldridge.
- World Health Organization. (1947). *Constitution of the World Health Organization. Chronicle of World Health Organization*, 1, 29–43.
- Zickafoose, J. S., Smith, K. V., & Dye, C. (2015). Children with special health care needs in CHIP: Access, use, and child and family outcomes. *Academic Pediatrics*, 15(3), S85–S92. http://dx.doi.org/10.1016/j.acap.2015.02.001

CHAPTER 2

Relevance of Diversity and Disparities to Health Programs

The health status of individuals and populations is influenced by biological processes and by lifestyle behaviors and circumstances. The intersection of biology, lifestyle, and environmental circumstances leads to disparities in health status, with some groups having lower morbidity and mortality rates than other groups. At the root of health disparities is diversity in biological characteristics, as well as in social, cultural, ethnic, linguistic, and economic characteristics of individuals and populations.

In the late 1990s, President Bill Clinton put race, racism, and ethnic diversity on the public agenda. As a consequence, federal agencies, including the National Institutes of Health (NIH) and the Department of Health and Human Services (DHHS), began explicitly to fund research into understanding and eliminating racial and ethnic disparities in health status. Private foundations and other agencies funding health programs followed suit by requiring grantees to state explicitly how each program contributes to reducing racial and ethnic health disparities. The high level of attention given to health disparities means

that program planners and evaluators must appreciate the sources of disparity—notably diversity, understand what the key aspects of diversity are and how those aspects are relevant to health programs, and know which strategies can be used to address diversity so that the health program will be successful. This chapter begins to address these issues.

A current urban legend exemplifies the influence of culture on healthcare decisions and the importance of having culturally competent staff. A woman from Africa was in labor with her first child in a U.S. hospital. Her labor was not progressing, and the physician wanted to deliver the baby by cesarean section in an effort to minimize the potential brain damage that was likely to result from a vaginal delivery. The woman and her husband refused the surgery, opting for a difficult vaginal delivery. The couple explained that they needed to make their decision based on what their life would be like when they returned to Africa. In their home village, a woman with a history of a cesarean section would be in grave danger if she were to have another baby because of the lack of surgical services for delivery in her home village. The life and health of the woman

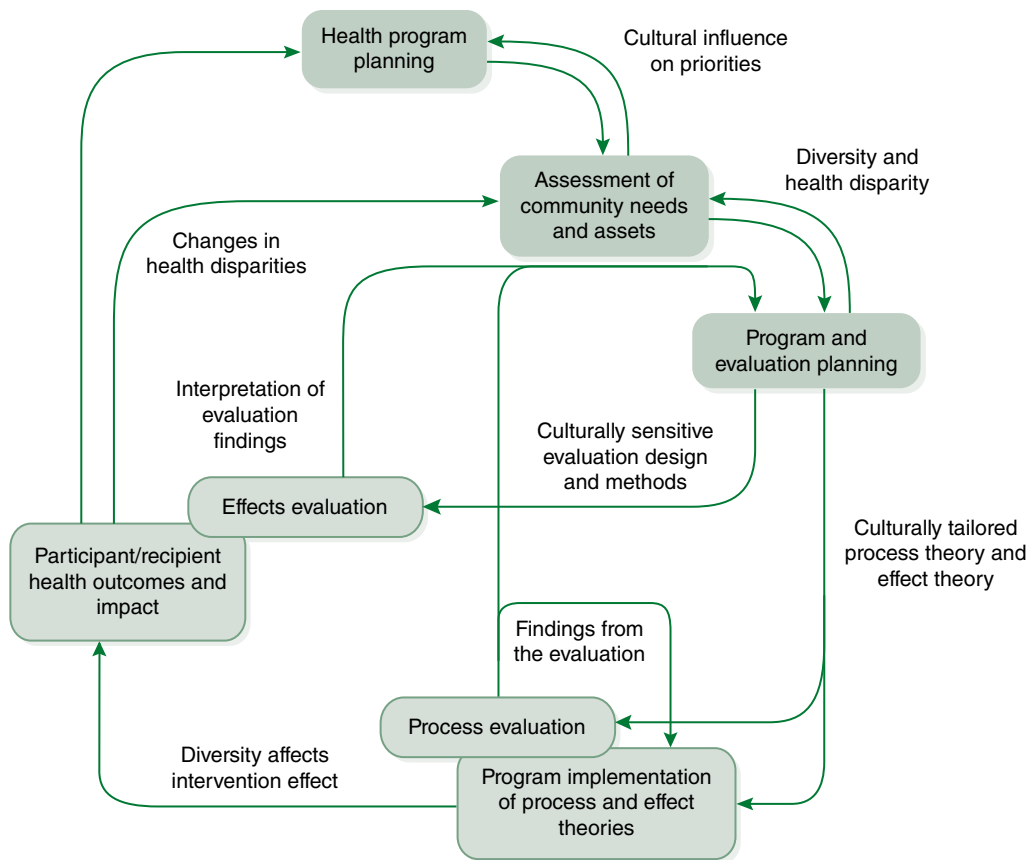


FIGURE 2-1 Effects of Diversity Throughout the Planning and Evaluation Cycle Stage in the Planning and Evaluation Cycle

were paramount. The child would be loved and cared for by the entire village, even if it were cognitively impaired from the difficult delivery. Whether the story is true has been lost in the telling. Regardless of its veracity, it highlights the influence of cultural values and norms on behavior and demonstrates the vital roles that culture, diversity, and life circumstances play in health discussions.

The topic of diversity is addressed early in this text because of its relevance throughout the planning and evaluation cycle (FIGURE 2-1). Diversity is relevant with regard to assessment of the health disparities to be addressed. It also affects the intervention choice and delivery, a

component of which is the issue of diversity of health providers. TABLE 2-1 provides examples of considerations that need to be weighed throughout the health program planning and evaluation cycle. The culture of the healthcare organization and the cultural competency of the program staff are directly related to the ability to tailor programs culturally, as is the formation of coalitions.

► Health Disparities

Health disparities and *health inequities* are terms denoting important differences in health status

TABLE 2-1 Examples of Cultural Tailoring Throughout the Program Planning and Evaluation Cycle

Stage in the Planning and Evaluation Cycle	Examples of Tailoring for Cultural and Ethnic Diversity
Community needs assessment	Definitions of health and illness; willingness to reveal needs or wants; self-definition in terms of culture, race, or ethnicity; health disparities; experience of disparities in access to or quality of health care
Program theory and development	Identification of contributing and determinant factors of the health disparities; role of discrimination and culturally bound health behaviors in the disparities; culturally acceptable and appropriate interventions
Process or program implementation	Culturally and ethnically adjusted program objective targets; cultural, racial, and ethnic representations and appropriateness of materials developed or chosen, such as visual representations, colors used, language or languages, location, media used, modality of distribution, and enticement used
Program intervention delivery	Type of intervention; length of time participants receive intervention (e.g., session length); amount of intervention (e.g., number of sessions)
Program effect evaluation	Language or languages of survey questionnaires; culturally appropriate enticements to participate; access to culturally and ethnically equivalent control groups

among socioeconomic, racial, and ethnic groups. *Disparities* in health care are defined as differences by race or ethnicity in access to or the quality of health care that are not due to the health or clinical needs or preferences of the person. The intuitive understanding belies the challenges in defining health disparities in a way that addresses the complexity of the problem (Braveman, 2006).

Well-documented health disparities exist. For example, blacks have nearly twice the rates of low birthweight infants and infant mortality as whites (National Center for Health Statistics [NCHS], 2016). Unintentional injury mortalities for American Indian children ages 1 to 4 years (11.7 per 100,000) is higher than for black (10.7 per 100,000) or non-Hispanic whites (7.2 per 100,000) (Hearon, 2016). Disparities also exist for

chronic illnesses: American Indians and Alaska Natives are 1.7 times more likely to have diabetes than non-Hispanic whites of a similar age, a notable improvement in narrowing that disparity since 2000 (NCHS, 2016). Black women have higher mortality rates from breast cancer than any other racial or ethnic group in the United States (National Cancer Institute, 2014). Some health status disparities might be explained by disparities in physical activity and consumption of fruits and vegetables (Gavin et al., 2011). These are a few examples of health disparities that could be addressed by individual practitioners but are perhaps more appropriately targeted by health programs across the public health pyramid.

The causes of health disparities remain the subject of research, but current theories regarding

health disparities posit that they have multiple, interactive (i.e., not mutually exclusive) causes that are biological, socioeconomic, and cultural in nature. For instance, among black mothers in particular, the biological effect of heightened levels of cortisol due to perceived discrimination has been associated with adverse birth outcomes (e.g., preterm birth, low birthweight). Among American Indians in the Southwest, the prevalence of diabetes was minimal until the 20th century, when water restrictions transformed an agricultural society into one dependent on government subsidies, largely comprised of processed commodities (Satterfield, DeBruyn, Francis, & Allen, 2014). The interactive causes of health disparities either can be primary targets for health programs or can constitute a contextual environment for the health program. In either case, at the heart of addressing health disparities in a practical manner and developing successful health programs lies the need to understand the relationship of diversity to health disparities.

Diversity and Health Disparities

Diversity, in the context of health, refers to the numerous ways in which individuals and groups differ in their beliefs, behaviors, values, backgrounds, preferences, and biology. Diversity is most often described in terms of language, culture, ethnicity, and race. Each of these aspects, along with biological diversity within the human population, has health implications.

Culture is a learned set of beliefs, values, and norms that are shared by a group of people; it is a design for how to live (Spector, 1991). As a set of behavioral norms or expectations, cultural beliefs influence a wide range of behaviors, including dietary choices, hygiene practices, sexual practices, and illness behaviors. Through such behaviors, culture has an effect on health and therefore is relevant to health programs. Cultures can be difficult to define and distinguish, particularly when subcultures rather than the dominant culture are the target of a program.

Assigning a label to a culture is less important than seeking information about unique or distinct culturally bound patterns of behavior that have health implications. For example, it is not as important to be able to identify a person as being from Hopi culture versus Navajo culture as it is to ask about daily consumption of meats and fresh vegetables and the ways in which those foods are prepared, and to understand the historical context that has influenced changes in dietary patterns. Culture, as the sharing of similar beliefs, values, and norms, contributes to a sense of unity among the members of the culture. The cultural cohesion and sense of belonging to a cultural group is a powerful force in creating conflicts as well as in creating opportunities. Both the Hopi and the Navajo have strong cultural identities that present an opportunity for health program planners to build that cultural identity into a program. The strong cultural identity can also create conflicts, however, between program planners and people from the Hopi or Navajo nations if the program is perceived as threatening their culture or being inconsistent with their cultural beliefs.

The relationship between culture and illness is recognized as having distinct manifestations, especially in mental health. The American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (DSM-5), has replaced the diagnostic category of "culture-bound syndromes" (Tseng, 2006) with "cultural concepts of distress" (Ventriglio, Ayonrinde, & Bhugra, 2016). *Cultural-concepts of distress* is a broad classification that considers the ways that cultural groups experience and frame symptoms, ways of communicating the distress, and explanations of the symptoms. The new diagnostic category stems from a growing understanding that illness is, at least in part, socially and culturally constructed. The interaction of culture and illness extends into physical illnesses: A study of pain found different levels of pain tolerance between Italians, Swedes, and Saudis (Al-Harthy, Ohrback, Michelotti, & List, 2016), suggesting the possible influence of cultural

factors. The message is that diversity in culture is related to diversity in illness manifestations and responses to illnesses (Edwards, Moric, Husfeldt, Buvanendran, & Ivankovich, 2005; Hastie, Riley, & Fillingim, 2005), even within ethnic subgroups (Tan et al., 2008). Such differences can only be identified through cross-cultural comparisons and astute observations. If program planners lack direct personal knowledge of the culture, they will need to rely on key informants and published reports of cultural influences on illness manifestations that are specific to the target audience.

Diversity also exists with regard to the economic well-being of individuals, as measured through socioeconomic status (SES). The relationship between SES and a wide variety of health status indicators has a long, well-documented history (Kosa & Zola, 1975; Polednak, 1997). Research continues to unravel the relationships among education, income, and health. For example, Russell and colleagues (2016) found that mothers with less than high school education had 1.7 times increase in the likelihood of having a child with attention deficit-hyperactivity disorder and that coming from a low income family increased that risk to more 2.2 times. The correlation between SES and health status applies both across racial groups and within racial groups. For example, Bruce and colleagues (2010) found that, among African Americans, those with a higher SES had a lower risk of chronic kidney disease. Other research has found that the interaction among gender, income, and race/ethnicity has varying effects on self-rated health (Brown, Richardson, Hargrove, & Thomas, 2016). The fact that individuals in lower SES groups, regardless of other characteristics, have poorer health suggests that health programs may need to target specific SES groups, not just specific cultural, racial, or ethnic groups.

The attention given to cultural and ethnic diversity is driven, in part, by the numbers. In 2014, over one million individuals newly immigrated to the United States. Of those, 13% were from Mexico, 8% from India, 8% from China,

and 13% were refugees (Mossaad, 2016). By 2060, the population of the United States will be 43% non-Hispanic white, 29% Hispanic, 13% African American, and 9% Asian and Pacific Islander (Colby & Ortman, 2014).

Diversity and Health Programs

As this very brief introduction to health disparities suggests, the extent of diversity within a target population can have various effects on how health programs are developed and provided. Diversity has particular implications for measurement done during planning and evaluating the health program; for the design and implementation of the health program intervention; and for the healthcare organization and program itself, including cultural competency and coalition formation. Each of these is addressed in some detail in the sections that follow.

► Measurement

Measurement occurs throughout the planning and evaluation cycle. Measurement of health status and of factors contributing to the health problem occurs during the community needs assessment phase. Program delivery and participation measurement occurs during the process evaluation phase. Measurement of program effects occurs during program evaluation. At each of these points in the planning and evaluation cycle, diversity in the target audience and in program participants or recipients has ramifications regarding what is measured, which data are collected, and how data are collected.

The first consideration is always the purpose of measuring an aspect of diversity. This purpose is paramount in deciding how diversity will be measured. Imagine that in a Bowe County community assessment, an atheist born in Layetteville and a Muslim born in a neighboring town were grouped into the same ethnic category. Stated in this way, it seems strange to

assign these two individuals the same ethnicity. But grouping these individuals together makes sense if the purpose of the assessment is to have data on Mexican immigrant culture. Given that *ethnicity* denotes a set of religious, racial, national, linguistic, or cultural characteristics that define a group, the ethnicity measure in this community needs assessment was based on religion as Catholic or not and on birthplace as Bove County or not. Thus, non-Bove County-born Catholics were assigned a Mexican ethnicity. This example was intentionally contrived to demonstrate the importance of purpose in developing indicators of diversity and the profound effect the variables used have on the indicator and subsequent findings.

Culture is often implicit, tacit, and not expressed as a distinct factor, making it difficult to measure. In addition, because a dominant culture exists at a societal level, measures of culture are less useful in health programs than indicators of more discrete, smaller subpopulations, such as those that might be defined by ethnicity or nationality. For these reasons, ethnicity is used as a proxy for cultural identity. Typically, ethnicity is measured with a single item; however, using a valid and reliable measure of ethnicity is key to having good data for planning and evaluating health programs. The extent of language diversity and religious diversity makes constructing a comprehensive measure of ethnicity very difficult. For example, the large number of religions, religious sects, and churches listed in the U.S. military's *Ministry Team Handbook* reflects wide religious diversity, with each having specific dietary practices; clothing; health practices; religious practices; and birth, marriage, and death rituals. Health researchers are attempting to understand the relationship between health status and mainstream religious beliefs and practices (Baetz & Toews, 2009; Masters & Spielmans, 2007; Krause, Emmons, & Ironson, 2015).

Nationality, which identifies the place of birth of the individual or the parents, is a more straightforward measure. Because cultural identity and ethnicity can be difficult to measure,

nationality by birth or birthplace of the parents is sometimes used as an indicator of culture and ethnicity. Many countries have multiple ethnic groups, however, making it problematic to equate nationality with ethnicity or culture. Thus, if nationality is measured, another measure, such as primary language, may be needed to have a more accurate measure of ethnicity and culture.

The following example demonstrates the importance of carefully choosing indicators of diversity, such as measures of ethnicity or culture, for planning health programs. In one neighborhood of Chicago, a large percentage of the residents belong to a specific sect of Judaism. In this neighborhood, the food stores are kosher, the women's clothing is consistent with their religion, and friendships are built around synagogue membership. Less than half a mile away is another neighborhood with a large percentage of residents with ties to the Indian subcontinent. In this neighborhood, the food stores stock food for their cuisine, the women wear the traditional sari, and the social structure is built around the dominance of the male head of the household. The health statistics for the Jewish neighborhood are relatively good, but the health statistics for the East Indian neighborhood reveal women's health problems due to high rates of domestic violence and chronic illness related to alcoholism. Unless the data from the two neighborhoods are separated, the health statistics for the area as a whole will mask some of the women's health problems and understate the males' health problems related to alcoholism.

This description of two actual neighboring ethnic groups shows the extent to which program planners need to be familiar not only with the data but, more important, with the community characteristics. These characteristics include the cultural beliefs of the residents and the degree to which ethnic and religious diversity coexist rather than overlap. Having this level of understanding about the cultural and ethnic diversity of a community facilitates appropriate interpretation of community health status data.

Race has long been considered a physical characteristic. From a biological perspective, race

has historically been associated with specific genetic diseases, including sickle cell anemia, thalassemia, and some forms of lactose intolerance. Race has also been used as a proxy measure of culture, ethnicity, and SES. To the extent that race can be used as a risk factor for specific genetically

transmitted health conditions, it has some medical value. The categories used to measure race remain somewhat inconsistent across key national health data sources, as do indicators of ethnicity (TABLE 2-2). Use of race and ethnicity indicators deserves careful attention (Bhopal, 2006).

TABLE 2-2 Indicators Used to Measure Race in Different Surveys

Category Used	2010 U.S. Census ¹	2003 U. S. Standard Certificate of Live Birth ²	2015–2016–NHANES ³	2012 National Hospital Ambulatory Care Survey ⁴
Race				
White	X		X	X
Non-Hispanic white		X		
Non-Hispanic black		X		
Black, African American	X		X	X
American Indian or Alaska Native	X	X	X	X
Other Pacific Islander	X			
Native Hawaiian or Pacific Islander	X		X	X
Asian Indian	X			
Asian or Pacific Islander (API)		X	X	
Asian				X
Ethnicity				
Hispanic or Latino	X	X	X	X
Mexican, Mexican American, Chicano	X		X	

(continues)

TABLE 2-2 Indicators Used to Measure Race in Different Surveys (*continued*)

Category Used	2010 U.S. Census ¹	2003 U. S. Standard Certificate of Live Birth ²	2015–2016–NHANES ³	2012 National Hospital Ambulatory Care Survey ⁴
Puerto Rican	X	X	X	
Cuban	X		X	
Other Spanish, Hispanic, Latino	X			
Other Spanish or Hispanic			X	
Chinese	X			
Filipino	X			
Japanese	X			
Korean	X			
Vietnamese	X			
Other Asian	X			
Guamanian or Chamorro	X	X		
Samoan	X	X		
Other Pacific Islander	X			
Central American, South American, other Hispanic or Latino			X	
Other race	X		X	X

Sources: ¹U.S. Census. Retrieved June 30, 2016, from http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_DP_DPDP1&src=pt.

²U.S. Standard Certificate of Live Birth. Retrieved June 30, 2016, from <http://www.cdc.gov/nchs/nvss/births.htm>.

³National Health and Nutrition Examination Survey (NHANES), Screener Module 1. Retrieved June 30, 2016, from http://www.cdc.gov/nchs/nhanes/nhanes2015-2016/questionnaires15_16.htm.

⁴Survey Content for the National Ambulatory Medical Care Survey and Hospital Ambulatory Medical Care Survey. [Revised Nov. 2012] Retrieved June 30, 2016, from http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm.

Researchers have made progress in identifying genetic markers for diseases that are more specific than self-reported race. As tests for genetic markers have become more widely available and affordable, race, as currently measured, has lost some medical value. It is easy to imagine a future when the current self-report measures of race will no longer be medically relevant. Until the future arrives, however, race will continue to be used as an indicator in planning health promotion and disease prevention programs.

The cultural and ethnic background of program participants affects the development or choice of questionnaires as well as the interpretation of results. In the development of scientifically sound and rigorous data collection tools, the language and culture of the intended respondents must be considered. To ensure that a questionnaire is culturally and linguistically appropriate and understood requires that the questionnaire be translated from the primary language into the second language and then translated back into the primary language. The back-translated version is then compared with the original version of the questionnaire to determine the accuracy of the translation. In addition, the translation in each direction ought to be done with input from several fully bilingual experts in the content of the questionnaire. Translation of the words is not sufficient; both the ideas embodied in the questionnaire and the wording of each item need to be translated (Epstein, Santo, & Guillemin, 2015; Willis & Zahnd, 2007). The questionnaire needs to be culturally equivalent so that the ideas and the expressions are the same, not just the words.

The following is an example of the measurement challenges that program planners and evaluators could face. A questionnaire developed in the United States in English regarding group functioning in a work unit was chosen for use with Taiwanese employees. The questionnaire was translated into Chinese by three Taiwanese researchers and then translated back into English by three other Taiwanese researchers.

The back-translated version was considerably different from the original English version.

In trying to understand what had happened, the researchers found that two factors had come into play. First, the lack of a future tense in some Chinese languages made it impossible to translate directly the English items that asked about the future actions of the respondent. Second, the questionnaire had been designed to measure the degree of individual and group functioning based on the American value of individualism. Thus, the questionnaire was difficult to translate both linguistically because of the future tenses and conceptually because of the individualist versus collectivist values of the two cultures.

This example hints at the potential complexity of using a survey questionnaire designed for one culture with a second culture. It also highlights the potential ethnocentrism involved in thinking that what is valued in the American culture—in this example, individualism—would be relevant in other cultures. This translation story helps explain why such extensive publications exist for the SF-12, a 12-item measure of overall health that is one of the most widely translated health questionnaires and now is also available with only 8 items. The various publications document the SF-12's psychometric properties when translated and used in different countries, demonstrating that even a widely used and thoroughly researched questionnaire requires a considerable amount of work to ensure that it is culturally and linguistically appropriate with each culture.

Cultural diversity also affects the interpretation of findings based on the data collected. Stakeholders involved in the health program who come from different backgrounds and cultures will often hold different values and ideas. Their culturally based interpretations may be quite different from the interpretations of health professionals, who have their own professional culture.

Culture influences how meaning is attributed to findings and how data are collected for program evaluation. For example, a violence prevention

program measured the program's effectiveness in terms of the lack of gang tags spray-painted on walls in a neighborhood that was next to a city park. When residents of the neighborhood were presented with the findings, they interpreted the findings in a skeptical manner. They explained that, for them, the lack of gang tags did not mean the lack of gangs, just that they no longer knew where the gang boundaries were and therefore where it was safe to go, including whether it was safe to go to the park for exercise. This actual example exemplifies both the powerful influences of culture on interpreting data and the value of involving stakeholders in any data interpretation.

► Interventions

Program *interventions* are the actions done intentionally to have a direct effect on program participants or recipients. The interventions used in health programs must be tailored to the intended audience if the program is to be successful in achieving the desired health effects. The choice of interventions and manner of intervention delivery ought to be based on both the sociocultural diversity of the target audience and the biological diversity within the target audience. Three approaches are evident in how culture is addressed during the development of program interventions. In addition, the diversity of the health professionals and health sectors plays a role in the effectiveness of program interventions.

Influences of Sociocultural Diversity on Interventions

Fisher, Burnet, Huang, Chin, and Cagney (2007) conducted a literature review of interventions focused on culture as means of improving health. They argue that *cultural leverage* is a strategy used to improve the "health of racial and ethnic communities by using their cultural practices,

products, philosophies, or environments as vehicles that facilitate behavior change" (p. 245) of individuals and healthcare providers. Cultural leverage, therefore, encompasses culturally tailoring interventions to specific ethnic or cultural groups as well as culturally targeting specific ethnic or cultural groups. In addition, the interventions developed for cultural leverage are culturally competent (as discussed later in this chapter). The key point is that there is increasing emphasis on, and more sophisticated approaches to, addressing culture in ways that are appropriate and beneficial to improving health and decreasing health disparities.

Understanding how to tailor the program given cultural differences begins with having or collecting information about differences across and within cultural groups. Navarro, Wilson, Berger, and Taylor (1997), in providing Native American students with a program to prevent alcohol and substance abuse, found that tribal differences and conflicting religious themes among tribes were important to individuals participating in the program. This is not surprising given that more than 500 Native American languages exist, each associated with a different tribal culture. Among low-income, urban, African American women, Beckjord and Klassen (2008) found variations in cultural values such that the women with more traditional values were less likely to seek and receive breast cancer screening. Sometimes, however, cultural tailoring may not be necessary. For example, Bond et al. (2016) found similar knowledge and perceptions about human papillomavirus (HPV) among African Americans, Caucasians, and Hispanics living in the same town in South Carolina. A health information campaign in this area intended to increase HPV vaccination thus would not need to culturally tailored with regard to race or ethnicity.

A common practice is to incorporate faith into health programs and to design programs that are church or parish based. In a review of the literature on church-based health promotion programs, Campbell and colleagues (2007) found that this approach is effective for African

Americans. The practice of collaborating with spiritual leaders and basing health programs in places of worship is likely to continue as an approach to reducing disparities. Developing faith-based programs may or may not require first understanding the health disparity in terms of religion, but it certainly does require understanding ways to collaborate effectively with church leadership and members for the delivery of a health program.

Interventions may result in immediate or permanent changes, but most health behavior interventions are intended to change behaviors that must be sustained over time. Culture can affect whether behaviors are sustained. Potentially, one type of culturally tailored intervention might be needed to initiate change and another type of culturally appropriate and tailored intervention might be needed to maintain the change or program effects. Also, program outcome objectives need to be culturally appropriate, with correspondingly appropriate target levels.

Sociocultural influences on intervention may emerge in unanticipated ways, such as through program participants themselves. For example, program participants bring their culture to the program in ways that can affect the intervention and its effectiveness. Higginson (1998) studied adolescent mothers in a high school program to examine their competitive culture, concluding that it was shaped by the mothers' social class, age, and race. Their competitive culture also pervaded the beliefs and norms of the health program in which they participated. These adolescents socialized new program participants into the competitive culture, thereby creating a "program culture."

One way to understand the culture of a program is to analyze the ceremonies and rituals associated with it. Intentionally building ceremonies and rituals into programs can be important for high-risk groups whose members need the sense of belonging that comes with having a program culture. This can help to retain them as active program participants. These are some

examples of how diversity can affect program interventions.

Influences of Biological Diversity on Interventions

For some health conditions, physiological responses may vary by race, gender, or age, which in turn affects decisions about the type and intensity of interventions used in the health program. Generational differences in values, norms, beliefs, and health problems all contribute to diversity. From the perspective of health program planning, age distribution is an important factor in reaching the intended audiences of a program. Gender and sexual orientation are other dimensions of physical diversity that have ramifications for program development. Disability—whether physical, mental, or developmental—is another dimension of diversity but is less often mentioned. Nonetheless, it may be extremely relevant for some health programs.

The distribution of physical characteristics within a population or community influences decisions during health program planning and later during program evaluation. Take age as an example. Imagine that the Bowe County Board wants to increase the physical activity of all county residents. The age distribution across the community and within its towns will affect the nature and content of the countywide media messages. Messages that relate to the physical abilities of the elderly will need to be quite different from messages that address the physical abilities of adolescents. Similar considerations would be needed for the other types of physical diversity.

Approaches to Developing Programs

Various perspectives exist in regard to explaining patterns of health behavioral differences by culture, ethnicity, and race. Kim, McLeod, and Shantzis (1992) suggest that three approaches are used in health-related programs: cultural content

approaches, cultural integration approaches, and cultural conflict approaches.

In the cultural content approach, cultural backgrounds and norms are viewed as leading to behaviors and illnesses. For example, Kleinman (1980), a medical anthropologist, explains that illness is cultural in that sickness and symptoms are saturated with specific meaning and are given patterns of human behavior. The notion that illness is cultural, and not just biological, affects the degree to which individuals accept professional explanations of health and illness.

Cultural integration approaches to developing health programs focus on acculturation. *Acculturation*, the adoption and assimilation of another culture, affects behavior in that the less dominant group takes on behaviors of the dominant group. When planning programs, planners need to consider the degree of acculturation because it affects health beliefs and behaviors. Behavior is also affected when individuals identify with more than one culture to varying degrees so that bicultural individuals have health beliefs and behaviors that are a blend of the dominant and less dominant cultures. When targeting groups or individuals who identify with more than one culture, planners need to understand their health beliefs and behaviors as a “new” culture. This is particularly relevant for health programs targeting immigrants or first-generation U.S. citizens.

Cultural conflict approaches underscore conflict as the genesis of behaviors. Several areas of potential cultural conflict exist. One area stems from the generation gap, which leads to family conflict and unhealthy behaviors and illnesses. Differences between the role expectations of different cultures are another source of cultural conflict and unhealthy behaviors. Racism, oppression, and lack of political power lead to alienation and identity conflict, and subsequently to unhealthy behaviors and illnesses. From a psychological perspective, individuals who are experiencing these kinds of conflicts are more likely to experience stress and therefore have less attention and energy to engage in health-promoting behaviors or may be less receptive to making change. Thus an assessment

of the target population ought to assess the degree of cultural conflict. Program planners need to address the immediate, root causes of the cultural conflict if they are to develop appropriate interventions for the health program.

Profession and Provider Diversity

Health program planning and evaluation draw on the expertise of individuals from a multitude of health disciplines, including medicine, nursing, pharmacy, social work, nutrition, physical therapy, and dentistry, as well as social science disciplines, including health education, health psychology, social demography, and medical sociology. Each discipline has its own specialized knowledge, values, and professional norms. Successful planning, implementation, and evaluation of health programs require working on teams that bring together the strengths of the various professions and that respect the different educational backgrounds of team members (TABLE 2-3). Each health discipline speaks a slightly different professional language, holds different beliefs about how to identify and address health problems, and adopts a different perspective on what constitutes a health outcome. To tap into the wealth of information and experience available through professional diversity requires that the team develop a common language and shared goals for the health program.

Health professionals do not reflect the diversity profile of the population of the United States in terms of cultural, racial, and ethnic diversity. For example, African Americans accounted for 10.7% of registered nurses in 2010, yet African Americans made up 13.6% of the overall population. Similarly, 5.4% of registered nurses were Hispanic, compared to 15.5% of the total population (U.S. Department of Health and Human Services [U.S. DHHS], 2015). This same pattern of underrepresentation of minorities exists across all health professions. The ensuing lack of racial and ethnic diversity among health professionals creates a cultural gap between professionals and patients, clients, and program participants. The extent of the cultural gap between planners and

TABLE 2-3 Professional Diversity Among Health Professions

Health Discipline	Average Education	Primary Focus	Licensure/ Certificate	Programmatic Contribution	Estimated Number in United States
Acupuncturist	Minimum 3 years of Asian medicine education	Restore balance between the physical, emotional and spiritual aspects of the person	Licensure based on certification	Alternative approach to understanding health and illness based on oriental medicine	27,835 in 2012 ³
Community health worker	High school or baccalaureate	Informal education, advocacy, assistance of community members	Varies by state for certification	Shares ethnic, linguistic, socioeconomic status, and life experiences with community members	48,130 in 2015 ¹
Dentistry	Dental-doctorate (3–4 years after baccalaureate)	Diagnosis and treatment of conditions of the teeth and gums	Licensure	Oral health knowledge	195,722 in 2015 ²
Dietitian, nutritionist	Baccalaureate	Dietary and nutritional elements necessary for health	Licensure and certificate	Nutritional knowledge, influence of nutrition on health	59,740 in 2015 ¹
Health administration	Master's degree	Leadership and management of healthcare organizations	Certificate	Management and administration	166,990 in 2015 ¹
Health education	Baccalaureate	Development and delivery of materials and curriculum designed to impart health knowledge and change behavior	Certificate	Social and behavioral knowledge	57,570 in 2015 ¹

(continues)

TABLE 2-3 Professional Diversity Among Health Professions (*continued*)

Health Discipline	Average Education	Primary Focus	Licensure/ Certificate	Programmatic Contribution	Estimated Number in United States
Massage therapist	Certificate	Massage of soft tissues and joints	Licensure varies by state, certificate		92,090 in 2015 ¹
Medicine	Medical doctorate (4 years after baccalaureate), plus residency of 3–4 years	Differential diagnosis and treatment of illnesses	Licensure and certificate	Medical, pathological, and treatment knowledge	621,066 in 2015 ¹
Nursing	Baccalaureate (associate degree minimum requirement)	Promotion of health and well-being based on scientific knowledge	Licensure and certificate	Integration of behavioral and medical knowledge	2,745,910 in 2015 ¹
Physical therapy	Baccalaureate	Restoration and maintenance of body strength and flexibility for the purpose of maximizing physical capabilities	Licensure	Focus on enhancing capability within limitations	209,690 in 2015 ¹
Social work	Master's degree	Address basic needs; help people manage environmental forces that create problems in living	Licensure	Focus on family and psychological factors	619,310 in 2015 ¹

Sources: ¹Community Health Workers National Workforce Survey. Retrieved September 11, 2012, from <http://bhpr.hrsa.gov/healthworkforce/reports/chwstudy2007.pdf>;

²Bureau of Labor Statistics (BLS). (2011). Occupational Outlook Handbook (OOH), 2010–2011 Edition. Retrieved September 11, 2012, from www.bls.gov/oco/home.htm;

³Data not available on these health disciplines.

⁴Bureau of Labor Statistics (BLS). (2011). Occupational Outlook Handbook (OOH), 2010–2011 Edition. Retrieved September 7, 2011, from <http://www.bls.gov/oco/home.htm>

a health program's target audience contributes to a reduced understanding of the target audience, a greater need to become informed about the target audience, and—potentially—tensions between the planners and advocates for or from the target audience. The more comprehensive the health program and the greater the cultural diversity of the target population, the greater the need to have parallel diversity among those planning, providing, and receiving the program.

The Three Health Provider Sectors

From an anthropological perspective, the effects of health provider diversity (or lack thereof) can be understood by considering the three sectors of the health–illness system from which individuals seek help when experiencing illness (Kleinman, 1980). Each sector has direct implications for planning, implementing, and evaluating health programs.

One sector consists of allopathic, naturopathic, and other formally trained health professionals who make up the medical healthcare system. Professionals from this sector have legally sanctioned practice parameters. Traditional, Western medical professions include, among others, physicians, nurses, pharmacists, psychologists and physical therapists, whereas complementary health professionals include, for example, naturopathic physicians, chiropractors, acupuncturists, homeopathy practitioners, licensed massage therapists, and mind–body therapists. The notion of health program planning falls within this sector, as do the methods and knowledge about health program planning and evaluation. In addition, the preponderance of health programs are designed in accordance with theories and knowledge generated from this sector.

A second sector from which individuals might seek help is the folk healthcare sector, which comprises nonprofessional, secular, or sacred healers who have not received formal education but who are very likely to have received training through some type of apprenticeship. A wide variety of traditional healers makes up

this sector—curanderos, espiritistas, santeros, singers, shamans, and root-workers, among others. Evidence of the presence of folk healers can be found when visiting neighborhoods that are ethnically isolated or that maintain folkloric traditions. This would be case with enclaves of recent immigrants. Individuals may consult folk healers while also receiving modern or Western health care. The theories of illnesses and diseases that are the basis of folk health practices can conflict with allopathic theories and thus may diminish the effectiveness of interventions based on an allopathic frame of reference. The role of folk healers in community health behaviors and in addressing health problems can be central for some health programs, especially those targeting individuals who have maintained “the old ways.”

The third (and largest) sector of health providers is the popular or lay sector, consisting of family and friends. Undoubtedly, most of us talk to a family member or friend about our illness before seeking either professional or folk health care. This sector is the most relied upon, from receiving the latest news disseminated through the mass or social media to getting a mother's recipe for chicken soup. Health information is spread through the lay sector through social networks, making it a powerful factor in influencing health knowledge and behavior. Health programs that seek to change social norms or population-level behaviors are essentially seeking to change the lay healthcare sector.

► Diversity Within Healthcare Organizations and Programs

From a systems theory perspective, an organization that is internally diverse will be better able to respond to externally diverse needs and demands. This concept has been formalized into the concept of requisite variety (Weick, 1979). The concept of requisite variety suggests that

healthcare organizations with a culturally diverse and culturally competent workforce are better suited to provide services that meet culturally diverse health needs. The need for requisite variety is a fundamental reason for having a culturally and ethnically diverse health professions sector. The need for a diverse workforce was recognized in a report to the Bureau of Health Professions, within the Health Resources and Services Administration (U.S. DHHS, 2015), especially to benefit underserved and minority populations.

Organizational Culture

Many different types of organizations offer health programs, including state or local health agencies, for-profit acute care networks, nonprofit community-based agencies, and academic institutions. Each organization has a unique set of values, norms, and beliefs that are collectively held by its members and that are passed on to new employees; this constitutes the organizational culture (Deal & Kennedy, 1982; Schein, 1995). Well-known examples of organizational culture are the norms about starting meetings on time and the willingness to help other employees accomplish tasks.

Program managers need to be sensitive to the degree of fit between the organizational culture and the goals of the health program. Not all good ideas for programs are good for the organization. A good match or fit between the organization's view of its mission and philosophy—in other words, its beliefs and values—and the purpose of the health program may be important to the success of the health program in terms of financial, personnel, and other organizational support. In a similar vein, the integration and sustainability of a program within an organization are affected by organizational culture.

Another implication of organizational culture for program managers is that staff members with work experience hold some of the values and norms of their prior organizational culture. These values and norms can be shaped; in other words, new employees need to become acculturated into the new organization, a process that begins with their initial orientation. Cox (2001), an expert on multicultural organizations, defined diversity

within an organization as the variation in the social and cultural identities of people existing together. For organizations, diversity provides added value because it increases respect, improves problem solving, increases creativity and ideas, increases organizational flexibility, improves the quality of employees, and improves marketing strategies. Diversity within organizations does not just create benefits, however; it also poses challenges for managing and enhancing that diversity.

An essential element contributing to a healthcare organization's cultural competency is its ability to engage in self-assessment of its cultural competency. Understanding the cultural competency continuum puts the self-assessment into perspective.

Cultural Competency Continuum

Accompanying the emphasis on diversity and health disparities is the emphasis on *cultural competency*, the extent to which individuals are able to live or work in a culture other than their own. Cultural competency, by its very nature, has shades of less and more that extend along a continuum (Cross, Bazron, Dennis, & Issacs, 1989; Orlandi, 1992; **TABLE 2-4**), an idea that has gained wide acceptance (e.g., Knibb-Lamouche, 2012). It is possible for health professionals and program staff members to reside at different points along the continuum, depending on a variety of factors, such as the specific circumstances and the individuals' experiences with cultures other than their own. While the prevailing norm and politically correct stance is to be as culturally sensitive and as competent as possible, acceptance of different values and beliefs can be difficult, particularly those of cultures that are dramatically different from one's own.

Cultural Destructiveness

At the least tolerant end of the continuum is cultural destructiveness (Orlandi, 1992), which includes a set of attitudes and practices that explicitly promote one culture over another based on the notion of one culture being superior to the

TABLE 2-4 Cultural Continuum with Examples of the Distinguishing Features of Each Stage						
	Cultural Destructiveness	Cultural Incapacity	Cultural Blindness	Cultural Openness	Cultural Competence	Cultural Proficiency
Attitude toward other cultures	Hostility	Dislike, separate but equal	Ambivalence, treat all alike	Curious, cultural awareness	Respect and tolerance, cultural sensitivity	Fully comfortable, cultural attunement
Knowledge of other cultures	Active avoidance of knowledge	None	Little or none	Some	Fair amount	Extensive
Degree of integration across cultures	None	None	None	Contemplation of potential benefits of integration	Some integration, some elements of multicultural integration	Extensive integration, fully multicultural, fusion of cultures
Implications for health program of participants at each stage	Programs address consequences of cultural destructiveness	Need to have programs provided to separate groups	If have multicultural elements, may need to justify and explain	Can provide program to participants from multiple cultures but will need to provide information and role modeling of competence	Can provide program to participants from multiple cultures with minimal adjustments	Can provide multilingual, multicultural interventions in one program

other. The attitude of superiority of one's culture over the inferior culture stems from the notion of the other being different or distasteful. Often physical (visible) characteristics, especially race, gender, sexual orientation, and age, are used as the basis for cultural destructiveness. Although staff members of a health program are not likely to be at this end of the continuum, health programs might be needed by and planned for individuals with attitudes reflective of cultural destructiveness. In fact, many of the global conflicts that lead to humanitarian crises and refugees have their roots in cultural destructiveness. International health programs are likely to deal directly with the consequences of cultural destructiveness. For programs within the United States, program planners will need to have an "insider" understanding of factors that would make the health program acceptable to culturally destructive groups.

Cultural Incapacity

Individuals at the next step, cultural incapacity, also promote one culture over another, albeit more implicitly than individuals at the cultural destructiveness stage. Cultural incapacity is manifested in the doctrine of "separate but equal," with the accompanying segregation and discrimination. In the United States, both cultural incapacity and cultural destructiveness have been made illegal through constitutional, federal, and various state statutes.

Cultural Blindness

Cultural blindness is a perspective of being unbiased, such that people are viewed as being alike and consequently are treated alike. At this point, the definition of "alike" is based on the dominant culture, giving cultural blindness ethnocentric overtones. Historically, health programs sought and delivered universal solutions without regard to different communication patterns of different cultures (Airhihenbuwa, 1994). Treating everyone in an unbiased manner would seem to be a reasonable premise for a health program. Cultural blindness, however, does not lead to effective programs.

One explanation for this phenomenon, taken from educational psychology, centers on the role of the dominant culture. Boekaerts (1998) suggests that because culture affects self-constructs, it also affects key features of how individuals learn and process information. As a result, what may be an effective learning environment for members of the dominant culture may not be effective for members of the less-dominant culture, who are being treated like members of the dominant culture. This theory implies that health programs, especially those with education or learning components that are based on a cultural blindness perspective, are not likely to be effective for individuals who are not from the dominant culture.

Another way of thinking about the consequences of cultural blindness is by acknowledging its failure to recognize that ideas and concepts are not the same across cultures due to the differences in self-constructs and learning. From this perspective, the earlier discussion of the need to translate concepts used in questionnaires is another example of how to overcome cultural blindness and its potential consequences for health program planning and evaluation.

Cultural Openness

Cultural openness is the attitude of being receptive to a different culture and to active learning about other cultures. Although other cultures are valued and some knowledge of other cultures exists, cultural openness does not include any integration of cultures or cross-pollination of cultural ideas. In this regard, cultural openness is similar to cultural awareness. Each culture is valued and understood as separate and distinct.

An example of being culturally open is someone from a dominant white culture going to a local Native American powwow or to an inner-city black evangelical church service simply to observe what happens. Cultural openness in health programs would be evident in having minority representation on community or advisory boards for the health program, using consultants with expertise in cultural awareness, and providing

cultural sensitivity training for staff members. Such culturally open practices increase the likelihood that the health program will be culturally appropriate, but they do not ensure its appropriateness. To ensure that the health program is culturally appropriate requires actively seeking information and integrating that information into the design, delivery, and evaluation of the health program. This process requires cultural competence.

Cultural Competence

Cultural competence encompasses not only demonstrating respect for other cultures but also actively seeking advice and consultation from members of the less dominant cultural group about what is culturally appropriate from their perspective. Acting in a culturally competent manner requires various skills that one needs to acquire intentionally. These skills are more specific than listening and being respectful. Continuing with the Native American example, if a tribal healer is consulted and included as a full member in the planning team for a health program intended for members of his tribe, then the health planning team is exhibiting culturally competent behaviors, especially if the healer's approach to healing is included in the program. Generally, cultural competence is understood as an individual characteristic of providers. For example, in a study of medical clinics, Paez, Allen, Carson, and Cooper (2007) found that more culturally competent provider behavior was associated with the clinic having more nonwhite staff members and more culturally adapted patient education materials.

One challenge to understanding what constitutes cultural competence is that other terms may be used to describe it, such as "cultural sensitivity" and "cultural attunement." Both sensitivity and attunement can be viewed as elements of cultural competence. Hoskins (1999) has proposed five principles of cultural attunement: acknowledging the pain of oppression by the dominant culture, engaging in acts of humility, acting with reverence, engaging in mutuality, and coming from a place of "not

knowing." Her work has influenced thinking across health disciplines (e.g., Jackson & Samuels, 2011). Hoskins's principles are notably developed for members of the dominant culture, with the implicit expectation that the member of the dominant culture needs to become culturally competent. In other words, it is incumbent upon the member of the dominant culture to strive for cultural competence. These principles also reveal that cultural competence, as a set of behaviors, may be difficult to attain or maintain over time.

The Lewin Group (2001), writing in a report for the Health Resources and Services Administration (HRSA) on cultural competence, listed nine domains for measuring cultural competence for healthcare organizations: values and attitudes of mutual respect and regard; cultural sensitivity; communication; policies and procedures (i.e., hiring staff members who reflect the linguistic and cultural diversity of the community); training and staff development; facility characteristics; capacity and infrastructure (i.e., cultural appropriateness of the physical environment, materials, and resources; use of posters and brochures with representatives from different races and ethnicities); intervention and treatment features; and community and consumer involvement and participation in decision making, and monitoring and evaluation of research. These domains continue to be relevant and used as the basis for organizational self-assessments.

This list of domains hints at the corresponding amount of work needed to achieve and maintain a culturally competent organization and workforce. These same domains clearly apply to programs.

Cultural Proficiency

At the most culturally capable end of the cultural competency continuum is cultural proficiency, which involves proactively seeking knowledge and information about other cultures, as well as educating others about other cultures. Cultural proficiency, as with any end point on a continuum, is difficult to achieve and may not be sustained for a long period of time. Those rare individuals who can move seamlessly among

cultures, be accepted in those cultures, and act as ambassadors of multiple cultures would be considered culturally proficient.

Being *multicultural*—that is, fully accepting and integrating two or more sets of cultural values and beliefs—is a manifestation of cultural proficiency. Multiculturalism in an organization or program (Cox, 1991) is the extent to which different cultures are fully integrated. It is manifested in programs that integrate folk or professional practitioners and treatment options, have predominantly bicultural staff, celebrate holidays important to cultural groups involved in the program, and synthesize different cultural beliefs into the program plan and implementation.

Enhancing Cultural Competency

Program managers can enhance cultural sensitivity, cultural awareness, and cultural competencies

through several strategies other than hiring consultants or sending staff members for cultural competency training. Cox (2001) has stressed that to have a diverse, friendly organization, workplace, or program requires making system-wide changes, affecting everything from hiring policies to the physical structure of the workplace, that are aligned with valuing and respecting the diversity of personnel.

For example, before making plans for organizational system changes, an organizational or program self-assessment of cultural competency is warranted. A variety of assessment tools have been developed and validated for assessing health-care employees (e.g., Loftin, Hartin, Branson, & Reyes, 2013). In addition, the National Center for Cultural Competence (Cohen & Goode, 1999) has developed a simple checklist (**EXHIBIT 2-1**) for use by program planners as well as by other individuals who have roles in shaping policy

EXHIBIT 2-1 Checklist to Facilitate Development of Cultural and Linguistic Competence Within Healthcare Organizations

Does the healthcare organization, primary healthcare system, or program have:

- A mission statement that articulates its principles, rationale, and values for culturally and linguistically competent healthcare service delivery?
- Policies and procedures for staff recruitment, hiring, and retention that will achieve the goal of a diverse and culturally competent workforce?
- Position descriptions and personnel performance measures that include skill sets related to linguistic competence?
- Policies and resources to support ongoing professional development and in-service training (at all levels) related to linguistic competence?
- Policies, procedures, and fiscal planning to ensure the provision of translation and interpretation services?
- Policies and procedures regarding the translation of patient consent forms, educational materials, and other information in formats that meet the literacy needs of patients?
- Policies and procedures to evaluate the quality and appropriateness of interpretation and translation services?
- Policies and procedures to periodically evaluate consumer and personnel satisfaction with interpretation and translation services that are provided?
- Policies and resources that support community outreach initiatives to persons with limited English proficiency?
- Policies and procedures to periodically review the current and emergent demographic trends for the geographic area served in order to determine interpretation and translation services needs?

at the federal, state, or local levels. Using this checklist can help determine which areas are in need of attention (Goode, Jones, & Mason, 2002), with actions subsequently being taken to enhance the cultural competency of staff members and the program as a whole.

Enhancing the cultural competency of program staff members begins with recruiting minorities that mirror the composition of potential program participants. Staffing pattern includes having a diverse workforce as a venue for staff members to learn from each culture—that is, from other staff members. The beliefs, values, and goals of individual staff members need to be congruent with those of the organization, which results in the second factor, fit between an individual and the organization is well accepted as an appropriate criterion for hiring decisions (Cable & Judge, 1997; McCulloch & Turban, 2007) and may entail sensitive hiring decisions. Program personnel must believe in achieving cultural competency for all program staff members, not just themselves or other staff.

Cultural competency training sessions ought to be designed to overcome learner resistance and avoid creating stereotypes (Boutin-Foster, Foster, & Konopasek, 2008). One strategy to use with individual program staff members is to make it acceptable to ask questions about cultural beliefs, practices, and norms so that staff members can acquire the information necessary to become more culturally competent. Program personnel need to be able to express both their comfort and their discomfort with other cultures as a step toward receiving whatever information or counseling is needed to overcome the discomfort. Out of respect, cultural labels ought to be avoided, using instead objective descriptors or names of individuals.

Not all staff members will be equally accepting and competent with all other cultures, depending on their cultural background. Some cultures are more accepting and seeking of new experiences than others. Being alert to cultural differences within program staff members is an important step toward developing and ensuring organizational and program

cultural competency. Ignoring the difficulties inherent in having diversity can lead to further problems; therefore, the challenges inherent in moving an organization, a program, or an individual toward cultural competency need to be acknowledged and addressed in a forthright yet sensitive manner.

Another strategy for enhancing the cultural competency of program personnel is to make diversity visible. This effort might include displaying posters or cultural artifacts. It may also include making available to staff members professional journals with a health and culture focus, such as *American Indian Culture and Research Journal*, *Ethnicity and Disease*, *International Journal of Intercultural Relations*, *Journal of Black Psychology*, *Journal of Cross-Cultural Psychology*, *Journal of Health Care for the Poor and Underserved*, and *Journal of Multicultural Counseling and Development*. The high visibility of diversity in the workplace becomes a symbol that reflects the organizational culture of valuing and respecting cultural diversity.

Rea, Martin, & Wright (2002) offer realistic suggestions for addressing cultural issues in the workplace. They recommended avoiding open conflicts over cultural issues, especially given that no one “right way” exists. They also recommend working through informal communication channels when cultural issues need to be addressed or to achieve changes in organizational culture. Another realistic suggestion is for managers to focus on reinforcing those new behaviors that promote cultural competency and sensitivity. Acknowledging that individuals have personalities and thus rotating staff members to other work units or programs may be the best approach in some situations. The positive aspect of this last suggestion recognizes a hard truth: When a fit between the program and staff members does not exist, both parties may benefit from a change in the relationship. The trick to addressing this type of situation in a culturally and legally competent manner is for both parties to understand the issue as one of fit and not as a personal judgment.

► Stakeholders and Coalitions

Another key approach to achieving requisite variety is through the inclusion of diverse stakeholders in the process of planning and evaluating the health program, which is often accomplished through the development of coalitions. Several federal agencies, such as the Office of Minority Health and the Centers for Disease Control and Prevention (CDC), and private foundations, such as the W. K. Kellogg Foundation and the Robert Wood Johnson Foundation, have funding priorities related to health disparities that require programs to engage in coalition development, often in the form of community engagement. The emphasis on developing coalitions parallels the emphasis on health disparities and diversity. Coalitions, partnerships, alliances, consortia, and collaborative linkages are some of the structural forms that result when stakeholders, interested parties, members of the target audience, and professionals with expertise agree to work together toward the common goals of community and health improvements for common constituents. The term *coalition* is used as the umbrella term for such agreements.

Coalitions, in whatever form, can be viewed as potentially having power and being power brokers (Braithwaite, Taylor, & Austin, 2000). Underlying the emphasis on coalition initiatives is growing evidence that collaboration among stakeholders is key to ensuring effective community involvement and to decreasing health disparities. For example, coalitions have proved effective in reducing the number of uninsured children (Stevens, Rice, & Cousineau, 2007). Coalitions for health programs may be developed for a variety of reasons, such as creating a power base from which to gain attention for the health problem or resources to address the problem, or to achieve long-term sustainability of the health program. To achieve this purpose, of course, the coalition must be effective (i.e., successful). In a review of the literature, Zakocs and Edwards (2006) found some evidence that the following

characteristics were associated with greater coalition effectiveness: use of formal rules and procedures, an inclusive leadership style, participation by members, a diverse membership, collaboration with agencies, and group cohesion.

The process of forming a coalition follows commonsense, deceptively simple steps. At the core of a coalition is attention to group process, as the following discussion suggests. The initial step in forming a coalition is to identify potential coalition members who are either individual stakeholders or representatives of organizations with a potential stake in the healthcare program. Naturally, the potential members ought to reflect the diversity being addressed by the health program.

An early step is the task of articulating the common goal for the coalition. Coalitions are more likely to succeed if they have a defined goal with specific tasks that can be realistically accomplished with minimal expense. As coalition members, funding priorities, and leadership changes, and as time passes, the goal for which the coalition was established will need to be reiterated as a sounding board for decisions and directions. It is also worth noting that coalitions have a life cycle, which may begin with a programmatic focus but evolve to have a policy focus (Hill et al., 2007).

Also early in the formation of the coalition, program staff members must build credibility and trust both within the coalition and with stakeholders in its work. It takes time to build trusting and credible relationships, which are inevitably tested over time. Credibility and trust are extremely difficult to recover if lost. The credibility and trustworthiness of organizers are especially important considerations when working with culturally and ethnically diverse groups whose members have had negative experiences with coalitions or health programs in the past.

Rose (2000) suggested two strategies for building relationships in the coalition. One approach is to adopt issues of the coalition members as issues for the coalition. This strategy would be feasible when issues overlap—say, housing affordability and health programs for the homeless. The other strategy is to promote

honest dialogue, in which members can be frank without feeling threatened by retribution for ideas. Complementing this strategy is the adoption of a policy of “agree to disagree.” This ground rule for interactions tends to foster cooperation as well as trust. Rose reminded us that humor is a very effective tool for unifying members and for relieving tensions. It is always healthy to laugh at situations, to find the bright side, and to be amused. This need transcends cultures, despite cultural differences in what makes something humorous.

Throughout the process of forming and working with a coalition, attention to cultural competency is crucial. One aspect of being culturally competent involves conducting a self-assessment that assesses the values and principles that govern participation in coalitions. The National Center for Cultural Competence has developed a checklist that can be used to assess cultural competency in community engagement (Goode, 2001; **EXHIBIT 2-2**). The health program planners could use this tool—after substituting

“program” for “organization”—as a means of gauging the cultural competency of the health program to engage the community in health program development.

► Across the Pyramid

At the direct services level of the public health pyramid, disparities are seen as affecting individuals and their health status. As individuals from diverse cultures, ethnicities, races, and SES backgrounds interact with health professionals and the health program staff, the training in cultural sensitivity and competency is put into practice. If the professionals and staff members have not received or integrated this knowledge into their practice, the potential for continued healthcare disparities is present.

Health programs designed for the direct services level of the pyramid will need to verify that the interventions included in the program match the culture, language, and norms of the

EXHIBIT 2-2 Checklist to Facilitate Cultural Competence in Community Engagement

Does the healthcare organization, primary healthcare system, or program have:

- A mission that values communities as essential allies in achieving its overall goals?
- A policy and structures that delineate community and consumer participation in planning, implementing, and evaluating the delivery of services and supports?
- A policy that facilitates employment and the exchange of goods and services from local communities?
- A policy and structures that provide a mechanism for the provision of fiscal resources and in-kind contributions to community partners, agencies, or organizations?
- Position descriptions and personnel performance measures that include areas of knowledge and skill sets related to community engagement?
- A policy, structures, and resources for in-service training, continuing education, and professional development that increase capacity for collaboration and partnerships within culturally and linguistically diverse communities?
- A policy that supports the use of diverse communication modalities and technologies for sharing information with communities?
- A policy and structures to periodically review current and emergent demographic trends?
- Community partners who are representative of the diverse population in the geographic or service area?
- Ways to identify new collaborators and potential opportunities for community engagement?
- A policy, structures, and resources to support community engagement in languages other than English?

Included with permission of the Georgetown University National Center for Cultural Competence, Georgetown University Center for Child & Human Development, Georgetown University Medical Center.

program recipients. It may also be necessary for the health program to be designed so that the intervention can be culturally, ethnically, and linguistically tailored to those participating in the program at the moment. In terms of measurement considerations at this level of the pyramid, the direct interaction with program participants allows for needs assessment, program process, and program effect data to be collected from individuals, through either quantitative questionnaires or qualitative interviews.

At the enabling services level of the pyramid, disparities are seen as they affect aggregates and families. Diversity is manifested in subcultures or enclave ethnicity, as well as in the larger cultural context. The interpersonal interaction between the program staff members and the program recipients remains an essential element of services at this level. As a consequence, the cultural competency of individual program staff members continues to be important as they implement the program interventions.

The interventions provided as enabling services will need to be tailored to the specific sociocultural characteristics and preferences of the target aggregate. For example, an existing enabling service may be planned for a new target audience. This endeavor would result in fairly specific changes, modifications, or additions to the existing program in an effort to make it culturally and linguistically acceptable to the new target audience. In terms of measurement, data are likely to be collected from individuals, allowing for tailoring the data collection to the characteristics of the aggregate.

At the population-based services level of the public health pyramid, disparities within a population are revealed through the collection of data related to that population, such as vital statistics and healthcare utilization. For all practical purposes, disparities are most easily identified by examining differences within a population, although they can also be identified within large aggregates, such as schools. Because health programs designed for the population level of the pyramid are delivered or provided to the population, interpersonal interaction between program staff members and program recipients

will vary from minimal (e.g., in an immunization campaign) to none (e.g., in a media campaign). Thus, issues of cultural competency for program staff members are lessened.

The need for the intervention itself to reflect cultural competency remains at the population-based service level. Health programs targeted at populations face the challenge of deciding whether to make the program generically acceptable for most members of the population or whether to develop different versions of the intervention tailored to known, culturally distinct subpopulations or aggregates. This challenge, while similar to the need for flexibility in direct services programs, is complicated by the inability to tailor the intervention during a program encounter.

With regard to measurement, most data collected at the population-based services level of the public health pyramid will be on such a scale that simple, generic data collection methods will be needed. This will result in having data that offer less detail but cover more program recipients. Unlike programs at the direct services or enabling services levels, a population-based program may not be able to gather data on actual program recipients. This fact creates a situation in which program planners may need to work more closely with the organizations and agencies responsible for collecting population-level data to ensure that the measures employed are as relevant to the program as possible.

At the infrastructure level, personnel diversity, organizational culture, and program culture all play roles in program planning and delivery. Overall, diversity and disparities are visible through their effects on existing and new health policy and priorities and on organizational processes and culture. Interpersonal interactions among program planners, staff members, stakeholders, and policy makers are the focus of efforts to address health disparities and cultural issues. Programs at the infrastructure level aim to change the cultural competency of the workforce and the capacity of the workforce to address health disparities and cultural diversity.

As with programs for the other levels of the pyramid, interventions implemented at the

infrastructure level need to be tailored to the sociocultural characteristics of the target audience within the infrastructure. In addition, they need to address the professional diversity that exists within the infrastructure, within specific healthcare organizations or agencies, and within the healthcare system as a whole.

With regard to measurement at the infrastructure level, the availability of individual data versus aggregate data will depend on the nature of the health program. Health programs provided to groups of workers, such as cultural competency training, make it possible to measure specific attributes of program participants. Health program interventions designed to change health policy are not amenable to direct data collection but would rely on population-level data, especially for program effects.

One other infrastructure issue that warrants mentioning is the legal implications of diversity. For example, the Americans with Disabilities Act (ADA) of 1990 requires that planning for programs take into account issues of accessibility for disabled persons. Another legal issue relates to antidiscrimination laws, which affect both the management of program personnel and the process by which program participants are recruited and accepted into the health program. State laws and local ordinances regarding same-sex marriage and civil unions may also affect reimbursement for programs, responses to survey questions about marriage, and recruitment of family members into programs. All these factors influence the planning and evaluation of the health program and therefore fall within the purview of the infrastructure level of the public health pyramid.

► Discussion Questions

1. Discuss the ways in which the linguistic diversity of a target audience affects programs being planned at each level of the public health pyramid.
2. Think of a specific health program provided by a specific healthcare organization with which you are familiar. Complete

either of the cultural competency self-assessments included in this chapter (Exhibits 2-1 and 2-2). What surprised you about taking the self-assessment? Which recommendations would you make based on the results of the self-assessment?

3. Identify one health-related questionnaire that has been used with more than one cultural or linguistic group. Discuss the adequacy of the linguistic and conceptual translations of the questionnaire.
4. List four health programs in your community. Are they supported by coalitions? What is the composition of each coalition? Does there appear to be a relationship between coalition diversity and health program success?

► Internet Resources

Centers for Disease Control and Prevention

The Centers for Disease Control and Prevention's National Center for Health Statistics is a good resource for health statistics. It allows users to manipulate data on a specific health indicator by variables such as race or income. Find it at www.cdc.gov/nchs/hdi.htm.

National Center on Cultural Competence

Georgetown University's National Center on Cultural Competence (NCCC) website (<http://nccc.georgetown.edu/>) has a wealth of resources related to cultural competency. The mission of the NCCC is to increase the capacity of health and mental health programs to design, implement, and evaluate culturally and linguistically competent service delivery systems.

Bureau of Primary Health Care

The Bureau of Primary Health Care, which is part of the Health Resources and Services Administration (HRSA), has a website devoted

to creating centers of excellence. The relevant document is entitled *Transforming the Face of Health Professions Through Cultural and Linguistic Competence Education: The Role of the HRSA Centers of Excellence*; it can be found at www.hrsa.gov/culturalcompetence/cultcompedu.pdf.

Public Health Services, Office of Minority Health

The 2013 document entitled *National Standards for Culturally and Linguistically Appropriate Services in Health Care*, by the Office of Minority Health, sets out the standards for linguistically appropriate health care. It and related resources can be found at <https://www.thinkculturalhealth.hhs.gov/Content/clas.asp>.

Medical Anthropology

This medical anthropology website includes a page of culture-specific syndromes, with some explanation and maps. Visit it at http://anthro.palomar.edu/medical/med_4.htm.

The Joint Commission

The Joint Commission. (2010). *Advancing effective communication, cultural competence, and patient and family-centered care: A roadmap for hospitals*. Oakbrook Terrace, IL: Author. Although this workbook is written for hospitals, the content provides a comprehensive overview of the considerations involved in effectively communicating. It can be found at https://www.jointcommission.org/roadmap_for_hospitals/.

References

- Airhihenbuwa, C. O. (1994). Health promotion and the discourse on culture: Implications for empowerment. *Health Education Quarterly*, 21, 345–353.
- Al-Harthy, M., Ohrbach, R., Michelotti, A., & List, T. (2016). The effect of culture on pain sensitivity. *Journal of Oral Rehabilitation*, 43(2), 81–88. doi:10.1111/joor.12346
- Baetz, M., & Toews, J. (2009). Clinical implications of research on religion, spirituality, and mental health. *Canadian Journal of Psychiatry*, 54(5), 292–301.
- Beckjord, E. B., & Klassen, A. C. (2008). Cultural values and secondary prevention of breast cancer in African American women. *Cancer Control*, 15, 63–71.
- Bhopal, R. (2006). Race and ethnicity: Responsible use from epidemiological and public health perspectives. *Journal of Law, Medicine, and Ethics*, 34, 500–507.
- Boekaerts, M. (1998). Do culturally rooted self-construals affect students' conceptualization of control over learning? *Educational Psychologist*, 33, 88–108.
- Bond, S. M., Cartmell, K. B., Lopez, C. M., Ford, M. E., Brandt, H. M., Gore, E. I., & Alberg, A. J. (2016). Racial and ethnic group knowledge, perceptions and behaviors about HPV, HPV vaccination, and cervical cancer among adolescent females. *Journal of Pediatric and Adolescent Gynecology*. doi:<http://dx.doi.org/10.1016/j.jpjag.2016.02.005>
- Boutin-Foster, C., Foster, J. C., & Konopasek, L. (2008). Viewpoint: Physician, know thyself: The professional culture of medicine as a framework for teaching cultural competence. *Academic Medicine*, 83(1), 106–111.
- Braithwaite, R. L., Taylor, S. E., & Austin, J. N. (2000). *Building health coalitions in the black community*. Thousand Oaks, CA: Sage.
- Braveman, P. (2006). Health disparities and health equity: Concepts and measurement. *Annual Review of Public Health*, 27, 167–94.
- Brown, T. H., Richardson, L. J., Hargrove, T. W., & Thomas, C. S. (2016). Using multiple-hierarchy stratification and life course approaches to understand health inequalities: The intersecting consequences of race, gender, SES, and age. *Journal of Health and Social Behavior*, 57, 200–222.
- Bruce, M. A., Beech, B. M., Crook, E. D., Sims, M., Wyatt, S. B., Flessner, M. F., et al. (2010). Association of socioeconomic status and CKD among African Americans: The Jackson Heart Study. *American Journal of Kidney Diseases*, 55(6), 1001–1008.
- Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person–organization fit and organizational selection. *Journal of Applied Psychology*, 82, 546–561.
- Campbell, M. K., Hudson, M. A., Resnicow, K., Blakeney, N., Paxton, A., & Baskin, M. (2007). Church-based health promotion interventions: Evidence and lessons learned. *Annual Review of Public Health*, 28, 213–234.
- Cohen, E., & Goode, T. (1999). *Policy brief 1: Rationale for cultural competence in primary health care*. Washington, DC: National Center for Cultural Competence, Georgetown University Child Development Center. Retrieved July 27, 2016, from <http://nccc.georgetown.edu/information/organizations.html>
- Colby, S. L., & Ortman, J. M. (2014). *Projections of the size and composition of the U.S. population: 2014 to 2060*, Current Population Reports, P25-1143. Washington, DC: U.S. Census Bureau.
- Cox, T. (1991). The multicultural organization. *Academy of Management Executive*, 5, 34–47.
- Cox, T. (2001). *Creating the multicultural organization*. San Francisco, CA: Jossey-Bass.
- Cross, T. L., Bazron, B. J., Dennis, K. W., & Isaacs, M. R. (1989). *Towards a culturally competent system of care:*

- A monograph on effective services for minority children who are severely emotionally disturbed: Volume 1. Washington, DC: Georgetown University Child Development Center.
- Deal, T. E., & Kennedy, A. A. (1982). *Corporate cultures: The rites and rituals of corporate life*. Reading, MA: Addison-Wesley.
- Edwards, R. R., Moric, M., Husfeldt, B., Buvanendran, A., & Ivankovich, O. (2005). Ethnic similarities and differences in the chronic pain experience: A comparison of African American, Hispanic, and white patients. *Pain Medicine*, 6, 88–98.
- Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology*, 68(4), 435–441. doi:10.1016/j.jclinepi.2014.11.021
- Fisher, T. L., Burnet, D. L., Huang, E. S., Chin, M. H., & Cagney, K. A. (2007). Cultural leverage interventions using culture to narrow racial disparities in health care. *Medical Care Research and Review*, 64, 243s–282s.
- Gavin, A. R., Melville, J. L., Rue, T., Guo, Y., Dina, K. T., & Katon, W. J. (2011). Racial differences in the prevalence of antenatal depression. *General Hospital Psychiatry*, 33(2), 87–93.
- Goode, T. (2001). *Policy brief 4: Engaging communities to realize the vision of one hundred percent access and zero health disparities: A culturally competent approach*. Washington, DC: National Center for Cultural Competence, Georgetown University Child Development Center. Retrieved July 8, 2016, from <http://nccc.georgetown.edu/information/organizations.html>
- Goode, T., Jones, W., & Mason, J. (2002). *A guide to planning and implementing cultural competence organization self-assessment*. Washington, DC: National Center for Cultural Competence, Georgetown University Child Development Center.
- Hastie, B. A., Riley, J. L., & Fillingim, R. B. (2005). Ethnic differences and responses to pain in healthy young adults. *Pain Medicine*, 6, 61–71.
- Hearon, M. (2016). *Deaths: Leading causes for 2014*. National Vital Statistics Report. Hyattsville, MD.
- Higginson, J. G. (1998). Competitive parenting: The culture of teen mothers. *Journal of Marriage and Family*, 60, 135–149.
- Hill, A., De Zapien, J. G., Staten, L. K., McClelland, D. J., Garza, R., Moore-Monroy, M., et al. (2007). From program to policy: Expanding the role of community coalitions. *Preventing Chronic Disease*, 4, A103.
- Hoskins, M. L. (1999). Worlds apart and lives together: Developing cultural attunement. *Child and Youth Care Forum*, 28, 73–84.
- Jackson, K. F., & Samuels, G. M. (2011). Multiracial competence in social work: Recommendations for culturally attuned work with multiracial people. *Social Work*, 56, 235–245.
- Kim, S., McLeod, J. H., & Shantzis, C. (1992). Cultural competence for evaluators working with Asian-American communities: Some practical considerations. In M. A. Orlandi, R. Weston, & L. G. Epstein (Eds.), *Cultural competence for evaluators: A guide for alcohol and other drug abuse prevention practitioners working with ethical/racial communities* (pp. 203–260). [DHHS Publication No. (ADM) 92–188]. Washington, DC: U.S. Government Printing Office.
- Kleinman, A. (1980). *Patients and healers in the context of culture*. Berkeley, CA: University of California Press.
- Knibb-Lamouche, J. (2012). Culture as a social determinant of health: Examples from native communities. Institute on Medicine: Roundtable on the Promotion of Health Equity and the Elimination of Health Disparities, Seattle.
- Kosa, J., & Zola, I. K. (1975). *Poverty and health: A sociological analysis* (rev. ed.). Cambridge, MA: Harvard University Press.
- Krause, N., Emmons, R., & Ironson, G. (2015). Benevolent images of god, gratitude, and physical health status. *Journal of Religion & Health*, 54(4), 1503–1519. doi:10.1007/s10943-015-0063
- Lewin Group. (2001). Health Resources and Services Administration study on measuring cultural competence in health care delivery settings: A review of the literature. Retrieved July 27, 2016, from <https://archive.org/stream/healthresources00lewi#page/n0/mode/2up>
- Loftin, C., Hartin, V., Branson, M., & Reyes, H. (2013). Measures of cultural competence in nurses: An integrative review. *The Scientific World Journal*, Volume 2013. Retrieved from <http://dx.doi.org/10.1155/2013/289101>
- Masters, K. S., & Spielmann, G. I. (2007). Prayer and health: Review, meta-analysis, and research agenda. *Journal of Behavioral Medicine*, 30, 329–338.
- McCulloch, M. C., & Turban, D. B. (2007). Using person-organization fit to select employees for high-turnover jobs. *International Journal of Selection and Assessment*, 15, 63.
- Mossaad, N. (2016). *U.S., Lawful permanent residents: 2014. Annual flow report*. Washington, DC: Office of Immigration Statistics, Department of Homeland Security.
- National Cancer Institute. (2014). *A snapshot of breast cancer*. Retrieved June 30, 2016, from <http://www.cancer.gov/research/progress/snapshots/breast>
- National Center for Health Statistics [NCHS]. (2016). *National Center for Health Statistics. Health, United States, 2015: With Special Feature on Racial and Ethnic Health Disparities*. Hyattsville, MD.
- Navarro, J., Wilson, S., Berger, L. R., & Taylor, T. (1997). Substance abuse and spirituality: A program for Native American students. *American Journal of Health Behavior*, 21, 3–11.
- Orlandi, M. A. (1992). Defining cultural competence: An organizing framework. In M. A. Orlandi, R. Weston, & L. G. Epstein (Eds.), *Cultural competence for evaluators: A guide for alcohol and other drug abuse prevention practitioners working with ethical/racial communities* (pp. 293–299). [DHHS Publication. No. (ADM) 92–1884]. Washington, DC: U.S. Government Printing Office.

- Paez, K. A., Allen, J. K., Carson, K. A., & Cooper, L. A. (2007). Provider and clinic cultural competence in a primary care setting. *Social Science and Medicine*, 66, 1204–1216.
- Polednak, A. P. (1997). *Segregation, poverty, and mortality in urban African-Americans*. Oxford, UK: Oxford University Press.
- Rea, J., Martin, C., & Wright, K. (2002). Using person-centered supports to change the culture of large intermediate care facilities. *Person-Centered Planning: Research, Practice and Future Directions*. Baltimore: Paul H Brookes.
- Rose, F. (2000). *Coalitions across the class divide: Lessons from the labor, peace and environmental movements*. Ithaca, NY: Cornell University Press.
- Russell, A. E., Ford, T., Williams, R., & Russell, G. (2016). The association between socioeconomic disadvantage and attention deficit/hyperactivity disorder (ADHD): A systematic review. *Child Psychiatry & Human Development*, 47, 440–458.
- Satterfield, D., DeBruyn, L., Francis, C., & Allen, A. (2014). A stream is always giving life: Communities reclaim native science and traditional ways to prevent diabetes and promote health. *American Indian Culture and Research Journal*, 38(1), 157–190.
- Schein, V. E. (1995). *Working from the margins: Voices of mothers in poverty*. Ithaca, NY: ILR Press, Cornell University Press.
- Spector, R. E. (1991). *Cultural diversity in health and illness* (3rd ed.). Norwalk, CT: Appleton & Lang.
- Stevens, G. D., Rice, K., & Cousineau, M. R. (2007). Children's health initiatives in California: The experiences of local coalitions pursuing universal coverage for children. *American Journal of Public Health*, 97, 738–743.
- Tan, E., Lim, Y., Teo, Y., Goh, R., Law, H., & Sia, A. T. (2008, September). Ethnic differences in pain perception and patient-controlled analgesia usage for postoperative pain. *The Journal of Pain*, 9(9), 849–855.
- Tseng, W. (2006). From peculiar psychiatric disorders through culture-bound syndromes to culture-related specific syndromes. *Transcultural Psychiatry*, 43, 554–576.
- U.S. Department of Health and Human Services, Health Resources and Services Administration. (2015). Sex, race, and ethnic diversity of U.S. health occupations (2010–2012), Rockville, MD: National Center for Health Workforce Analysis. Retrieved July 8, 2016, from <http://bhpr.hrsa.gov/healthworkforce/index.html>
- Ventriglio, A., Ayonrinde, O. & Bhugra, D. (2016). Relevance of culture-bound syndromes in the 21st century. *Psychiatry and Clinical Neuroscience*, 70, 3–6.
- Weick, K. (1979). *Social psychology of organizing* (2nd ed.). Reading, MA: Addison-Wesley.
- Willis, G., & Zahnd, E. (2007). Questionnaire design from a cross-cultural perspective: An empirical investigation of Koreans and non-Koreans. *Journal of Health Care for the Poor and Underserved*, 18(4S), 197–217.
- Zakocs, R. C., & Edwards, E. M. (2006). What explains community coalition effectiveness? A review of the literature. *American Journal of Preventative Medicine*, 30, 351–361.

SECTION II

Defining the Health Problem

CHAPTER 3

Community Health Assessment for Program Planning

This chapter begins with a definition of community and types of needs. It then provides an overview of different approaches to planning. Then types of needs assessment are reviewed, with full acknowledgment that, as with many aspects of program planning and evaluation, this field is an area of specialization. The various models that underlie an assessment are reviewed, as are the prototypical types of assessment. The distinctions made among the models and the types of assessment do become blurred in practice. Nonetheless, having a clear sense of what could be involved allows for more thoughtful and intentional decisions on how best to design and carry out the community health assessment.

► Defining Community

The use of the ecological model and the public health pyramid leads naturally to considering community and its role in health program planning and evaluation. Ambiguity about the

meaning of community can lead to conflicts and confusion throughout the planning and evaluation cycle. Thus, it is worth considering what is and is not a community, as a prelude to clarify thinking and as a foundation for better planning of health programs. A *community* encompasses people and some form of proximity or place that enables interaction, and that interaction leads to shared values or culture (Bell & Newby, 1971; **TABLE 3-1**). A defining characteristic of a community is a group of people with the potential for interaction. Without the potential for interaction, sharing of values and norms cannot occur. Interactions can be virtual as well as person-to-person. To the extent that a large number of interacting individuals share values and culture, a community can exist. Of course, virtual communities that exist via electronic media extend the traditional, anthropological notions of community that grew from the study of tribes and villages.

Another defining characteristic of a community is that its members have shared values and norms of behavior. The prerequisite that a community

TABLE 3-1 Three Elements of Community, with Their Characteristics

Element	Characteristics of the Community Elements
People	Values, beliefs, behaviors, size, membership, demographic characteristics, social and economic status, sense of power or influence, sense of belonging
Place	Geography, boundaries, housing, industry, air, water, land, virtual presence
Interaction	Interpersonal communication, family based, education related, religion based, politically related, recreationally focused, virtual communication

have commonly held values precludes, in many instances, a census tract, a ZIP code, a telephone area code, a consortium of health agencies, or a catchment area for a health service from being a community in the more pure sense. In contrast, active members of a church or residents of a small and homogeneous neighborhood might be a community.

This distinction between a convenient geographic designation and an actual community is important for planning how to have participation by community members in the planning process. Nilsen (2006) reminds us that the major problem with defining a community based on geographic location is that the geography does not ensure that a sense of community exists. Shared values form the basis for the cultural unity of a community, which in turn is the basis for the perception of being connected and belonging to a community. From that sense of belonging stems the subsequent behaviors that might be attributed to members of a particular community.

The concept of a “sense of community” has been studied and found to include aspects of membership, influence over what occurs within the community, shared values and needs fulfillment, and a shared emotional connection. Perhaps not surprisingly, the breadth of this list of the elements means that community and its associated emotional and cultural components create challenges for health program planning and evaluation. Individuals’ sense of community varies according to their race/ethnicity

and income (Coffman & BeLue, 2009; Kono, Tadaka, Kanaya, Dai, Itoi, & Imamatsu, 2012; Belue, Taylor-Richardson, Lin, McClellan, & Hargreaves, 2006). To achieve participatory planning, it may first be necessary to understand the local variations in sense of community. Regardless of the size of the population, the sense of community can be a key factor in gaining support for the program and for maximizing the health effects of some programs.

Community as Context and Intended Recipient

Program planners must clarify the purpose for which “the community” is being delineated. A community, as a unit of individuals with some degree of cultural cohesion, can be both the target of a health program and the context in which a health program is provided and evaluated.

When a health program targets a community for the intervention, the health program may intend some or all of the community members to receive the health program. Thus, to establish the size of the health problem within that community, planners must delineate the community boundaries. In this sense, a community is akin to a population. Community assessments or community needs assessments are processes by which planners seek to understand and describe the health problem more fully. The word *community* is used whether or not a community

or a population is being assessed. However, if an actual community—in contrast to a population or an aggregate—is the focus, planners must clarify the boundaries of the community membership if they are truly to understand the specific health and social conditions within that community.

The community may be viewed as the context of the program in two different ways. First, the notion of community embodies a myriad of sociopolitical and economic factors that can influence the program plan and implementation. It may be possible only to acknowledge, articulate, and take into account these influences as contextual to the program (if resources will not address these influences directly). Second, from a different perspective, community members who are invited to participate in the planning process become an immediate, intimate context of the program intervention.

The concepts of community as target and community as context are not mutually exclusive. An example helps demonstrate their interactive nature. At an elementary school in Toronto, students from fifth and sixth grades, with assistance from pediatric residents, developed a 20-page health magazine for their classmates (Porepa et al., 2016). They were a community as target to the extent that all the children at the school were the intended recipients of the health magazine and were assessed as a unit to identify health topics. They were community as context to the extent that their level of knowledge, norms, social structure, and school structure were influences on developing the magazine.

In this case, to address the contextual influences, the program initiators included and promoted participation by members of the university community in planning the wellness program. In so doing, they overcame institutional barriers and mobilized resources for the program. The synergies achieved by involving community members became possible because the university community was understood not as a single thing or a simple geographic location but rather as a group of individuals. In other words, to involve “the community” in program planning requires having influential, energetic, devoted actual



FIGURE 3-1 Connections Among Program, Agency, and Community

individuals who serve as representatives of or who come from that group actively participate in planning activities.

Collaboration between health program planners and formal or informal leaders from a community, and the participation of those community leaders in the development and evaluation of a health program are increasingly valued. In fact, funding agencies often mandate collaboration as a prerequisite to being considered for funding. Collaboration creates interaction, which can intensify the sense of community and promote synergy among the community representatives, the agency sponsoring the health program, and the program staff (**FIGURE 3-1**). These interactions and influences move in two directions, with ideas and energy flowing toward the health program, and results and respect flowing from the health program.

Defining Terms: Based, Focused, and Driven

Three terms related to community need to be defined in any discussion of community health planning: community based, community focused, and community driven.

Community-based is an adjective describing where a program or service is provided. A health program is community-based if it is delivered at locations considered within the boundaries of the community rather than at a centralized location outside the community boundaries. Generally, this understanding translates into a program being delivered in local churches, schools, recreation centers, local clinics, or libraries.

Community-focused refers to the way in which the program is designed. Health programs that seek to affect the community as a whole, as a unit, are best described as community-focused. A community-focused program may seek to change the norms or behaviors of the members of a community that contribute to the health problem, or it may seek to reach all members of the community.

In contrast, health programs that result from the involvement of community members and their preferences and needs are referred to as *community-driven*. A community-driven program has its genesis—that is, its design and implementation—in the involvement, persistence, and passion of key representatives of or members from the community.

Three points are worth noting about these qualifiers. First, the terms *based*, *focused*, and *driven* are not mutually exclusive (FIGURE 3-2). In other words, a program can be all three: based in, focused on, and driven by a community.

Second, the degree to which a program is community-based, community-focused, or community-driven can vary. Each can vary independently of the other. For example, it is

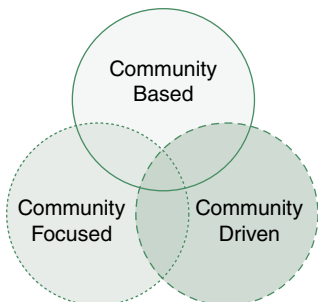


FIGURE 3-2 Venn Diagram of Community-Based, Community-Focused, and Community-Driven

easy to imagine a community-based city health clinic that is moderately community-focused and not at all community-driven.

Third, the designations of “based,” “focused,” and “driven” can also apply to families, populations, or other aggregates. Thus, family-based programs might be provided to individuals, but within the theoretical or physical presence of families, family-focused programs would be designed to enhance the family as a unit, and family-driven programs would be the result of a group of families advocating for or demanding the programs.

These distinctions can be important in terms of describing health programs and conceptualizing the nature of a program. Whether a health program is based or focused on a unit ought to flow from an understanding of the health problem and the best strategy for addressing it.

► Types of Needs

Four types of needs (Bradshaw, 1972) ought to be considered in a needs assessment. This classic typology of needs continues to be useful in planning the community assessment. Understanding the characteristics of each type of need as it relates to the population of interest is critical to successful health program planning.

Expressed need is the problem revealed through health care-seeking behavior. In other words, expressed need is manifested as the demand for services and the market behavior of the target audience. Measures of expressed need include the number of people who request services, the types of services sought, and utilization rates.

Normative need is a lack, deficit, inadequacy, or excess as defined by experts and health professionals, usually based on a scientific notion of what ought to be or what the ideal is from a health perspective. A norm or normal value is used as the gauge for determining if a need exists. For example, a community with an infant mortality rate above the national average would have a normative need related to causes of infant mortality. At an individual level, having a body

mass index (BMI) greater than 29.9 indicates, normatively, a need for weight reduction. Given that the health professional is an outside observer, normative need reflects norms through the eyes of an observer.

The third type of need is the *perceived or felt need*, which is the sense of lack as experienced by the target audience. Perceived needs are demonstrated in what members of the target audience say that they say they want, and in their stated deficits and inadequacies. For example, parents in a community may demand a new school based on their perception that their children have too far to travel to go to school. Perceived need is the view through the eyes of the person having the experience.

The *relative or comparative need* is the identified gap or deficit as identified through a contrast between advantaged and disadvantaged groups. Relative need entails a comparison that demonstrates a difference that is interpreted as one group having a need relative to the other group. Most health disparities are stated as relative need. For example, the black infant mortality rate is twice that of the white infant mortality rate. Health professionals or community members can choose the point of comparison, but each is likely to choose a different point of comparison. The relative needs thus would also be different.

The ways in which the interaction of these needs plays out in day-to-day situations can be seen in the experiences of a group of health planners who conducted a community needs assessment in a fictitious town called Layetteville. In conducting their assessment, they identified a neighborhood as having a relative need. The planners found that the neighborhood had higher rates of adolescent pregnancy, deaths due to gunshot injuries, birth defects, and diabetes than other neighborhoods in the city and that these rates were two to three times higher than those set out in the *Healthy People 2020* objectives. When the group explored the healthcare utilization patterns of the neighborhood residents, they found that the residents rarely used a primary healthcare clinic; instead, they used the local emergency department.

Confident in their understanding of what was needed to improve the health of the neighborhood, the group of planners approached the residents with a plan to establish a primary care clinic in the neighborhood that would provide prenatal care and diabetes management. Bluntly, the residents indicated that they were not interested: They wanted a community swimming pool in their neighborhood. They felt a strong need for recreation, a community meeting place, and an equal opportunity to engage in the healthy behavior of swimming as an alternative to the gang activities that were contributing to the shootings. Only after those conducting the assessment agreed to address the community's perceived need for a community swimming pool did the residents then consider how to address the normative, comparative, and relative needs that had been identified by the health professionals.

► Types of Strengths

A counterpoint to types of needs is types of community strengths that exist as resources available for use. Briefly, these three community strengths—community competence, community resilience and community assets—remind us to acquire a balanced view and understanding of communities. These concepts help focus the assessment on the glass as half full.

Community competence has long been considered a key strength (Cottrell, 1976; Goepfinger, Lassiter, & Wilcox, 1982). *Community competence* is the process whereby a community is able to identify problems and take actions to address those problems, and it is increased by community organizing (Denham, Quinn, & Gamble, 1998; Norris, Stevens, Pfefferbaum, Wyche, & Pfefferbaum, 2008). Greater community competence has been associated with both better health of the community and larger amounts of social capital (Lochner, Kawachi, & Kennedy, 1999) *Community resilience*, the ability of a community to withstand and recover from adverse situations or conditions (Plough et al., 2013),

has been largely applied to disasters and preparedness. Community resilience encompasses community capital, along with social, economic, and institutional factors (Ostadtaghizadeh, Ardalan, Paton, Jabbari, & Khankeh, 2015). Resilience contributes to the overall strength of a community. In contrast, community assets embody a breadth of tangible and intangible building blocks that exist as resources within the community and community members. Schools can be a community asset (Caan, Cassidy, Coverdate, Ha, Nicholson, & Rao, 2014). In one study of Chicago, in a 62-square-mile area, over 8,000 assets were identified across 23 sectors, such as food, trade services, and religious worship (Lindau et al., 2016).

► Approaches to Planning

With the basic understanding of needs and strengths, the health program planner next needs to engage more intentionally in planning. *Planning*, in the context of the planning and evaluation cycle, is the process of gathering information and making the subsequent decision regarding how something (e.g., the community assessment or the health program) will be done. A variety of definitions of “planning” have been suggested (Hoch, 1994; Alexander, 1992), but all share the elements of using a rational approach, making change, and using a democratic or participatory process. In terms of programs, planning is the set of activities in which key individuals define a set of desired improvements, develop an approach to achieve those desired improvements, and establish a means to measure the attainment of those desired improvements.

The wide range of different methods for planning can be categorized broadly into six approaches. An awareness of the various approaches used to accomplish planning helps planners interpret events and guides others through the planning process. Each approach provides a basis for assessing possible points of contention or agreement regarding health and

program priorities. Each of the six different perspectives on planning provides a different lens through which the program planning process can be viewed, understood, and then used to address the health problem. Beneveniste (1989) identified the processes underlying each approach. These approaches remain reflective of current approaches (Oliveira & Pinho, 2010; Guyadeen & Seasons, 2016). Each approach has various typical planning activities (**TABLE 3-2**). None of the approaches is inherently better or worse than any of the others. Rather, the purpose of becoming aware of all six approaches is to select an approach that matches the situation and to use the strengths of the approaches to arrive at an optimal process for developing a health program and its evaluation.

Incremental Approach

The incremental approach to planning does not attempt to address the problem in any particular context or across a specific time span. Instead, the approach is one of addressing immediate concerns and, to some extent, having faith that the small, rather disconnected plans and actions will have a cumulative effect on the problem. Incrementalism, by its very nature, focuses only on the immediate, without attempting to see the big picture or implement a long-term plan.

In the very early days of the human immunodeficiency virus, acquired immunodeficiency syndrome (HIV/AIDS) epidemic, before the causative virus had been identified or named, the only health planning options were incremental: shut down bath houses, use infectious disease precautions, and seek funding to study the health problem. These actions were isolated, disjointed efforts, but under the circumstances, the incremental approach was the only available option. Advocates for infected individuals had not yet emerged, nor was there a scientific basis for making rational, apolitical decisions.

As this example points out, incrementalism, while not the most effective planning approach, may be the only option in some circumstances.

TABLE 3-2 Summary of the Six Approaches to Planning, with Public Health Examples

Approach	Underlying Assumptions	Consequences of Use	Public Health Examples
Incremental	Not feasible to do more than small portions at a time; the parts are greater than the whole	Plans may fail because of unforeseen or unaccounted for factors	Specific programs implemented that reflect discrete, categorical funding despite potential overlap or existing similar programs
Apolitical	Options are known; makes the problem technical; the means to the ends are known; can anticipate all caveats	Experts may not accurately speak for others; media attention is likely to focus on the spokesperson rather than the issue	Evidence-based practice
Advocacy	An external expert can accurately speak for those with less power	An increased sense of confidence and an increased ability to solve one's own problems; potential for conflict	Activists for marginalized groups and environmental issues
Communicative action	Language is powerful; those with the problem have the capability to enact a solution	Takes considerable time and effort to implement; likely to have more dissent to overcome; results in an encompassing, intertwined set of actions	Community coalitions that take on a program or become nonprofit organizations
Comprehensive rational	System feedback loops are contextual and can be known; rational choices are preferred	Lacks flexibility to respond to emerging issues; a costly process to arrive at a plan	Community-focused initiatives
Strategic planning	Can anticipate and predict the future; stability is more pervasive than change	Circumscribed focus may miss opportunities or emergent needs	<i>Healthy People</i> series; states' 2- to 5-year plans; Title V 2-year plans

In addition, when resources are limited, incrementalism can lead to small gains related to immediate problems. The major disadvantage of incrementalism is that the myriad small

planning efforts may lead to conflicting plans, confusing programs, programs or services that are not integrated, or personnel redundancy or mismatch with the “new” program.

Apolitical Approach

The apolitical approach to planning relies solely on technical knowledge to arrive at a solution and assumes that technical knowledge makes it possible to achieve compromises among those involved in the health problem and the planning process. In a sense, the apolitical approach is fundamentally a problem-solving approach that relies on current knowledge about the problem and known alternatives to address the problem. The name of this approach reflects the fact that its focus on the technical aspects ignores the political aspects inherent in any problem. This approach is implicitly a gold standard of planning, particularly when those involved in the planning process are more technically inclined and focused.

To the extent that the apolitical approach relies on objective information for decision making, the application of evidence-based practice (EBP) guidelines can be viewed as essentially apolitical approaches to planning. In EBP, guidelines for practice by individual practitioners and complex health promotion programs are developed solely based on the best available scientific knowledge, without consideration of the context of the practice or the preferences of those experiencing the health problem. When planners use EBP guidelines as the basis for health program planning without taking other factors into account, they are engaged in apolitical planning.

One criticism of the apolitical problem-solving approach is that it does not account for interpersonal dynamics and possible struggles for control (Forester, 1993). This approach also neglects cultural issues involving the potential program participants and program staff members, which can be substantial stumbling blocks in applying the health programs. Nonetheless, the apolitical approach has the advantage of being—or at least providing—the appearance of being logical and rational and of specifying solutions with the documented highest efficacy.

Advocacy Approach

The advocacy perspective on planning focuses on the client and mandates citizen participation in

planning activities. Beneveniste (1989) described advocacy planning as a bottom-up form of comprehensive rational planning. Planners using the advocacy approach, however, would be likely to speak for or on behalf of those with the health problem.

For example, experts in environmental hazards may testify before city or county elected officials to plead the case for people living in areas where hazardous waste is generated or stored. When those who live in the area are unaware of a problem but an expert is nevertheless safeguarding their best interest, advocacy planning is occurring.

As Guerin, Allotey, Elmi, & Baho (2006) discovered, advocacy is personal because of the intense involvement between the advocate and those receiving advocacy. They also see political advocacy as a final stage of community-based participatory research. This perception is similar to the findings of Hill and colleagues (2007), who noted that community planning groups that begin with a focus on program development often evolve into policy planning and advocacy. These reports suggest that an advocacy approach to planning is more likely when the group has been established long enough to acquire an in-depth understanding of the health problem and a deep appreciation that an advocacy planning approach is needed to address the health problem further.

The advantages of the advocacy approach are most readily evident in situations in which clients or citizens are not empowered, for whatever reasons, to convey their own preferences or concerns. In such situations, having an advocate may be the only option for planning a needed health program. The disadvantages are that the clients or citizens may not agree with the opinions or views of the advocate. In fact, their “advocate” may not even be representing those he or she claims to speak for because this individual implicitly holds a normative view of the needs of the clients or citizens. The advocacy approach also implicitly entails some degree of conflict or confrontation, which may have negative repercussions over the long term. Social irresponsibility arises when the solution ignores important social or cultural

factors (Blum, 1982). Strong advocates and users of the apolitical approach to planning may be prone to this pitfall simply because a scientific basis may not take into account social realities and needs.

Communication Action Approach

Communication action planning is concerned with the distribution of power and communication. From this perspective, those involved in the planning make efforts to empower those with the problem through sharing of information. Whereas the advocacy approach does not enable those with the problem to participate as equals with the “experts” in the planning process, the communication action approach is predicated on making those with the problem equals in the planning process. According to Forester (1993), this perspective leads planners to think of planning as shaping attention, changing beliefs, gaining consent, and engendering trust and understanding among those involved.

One of the many examples of this approach to planning is evident in mental health. The National Alliance on Mental Illness (NAMI) is a nonprofit organization whose purpose is to support those with mental illness, and it is run by individuals with mental illness. Individuals with mental health problems are taught and guided in the process of developing small-scale, community-based programs and services for those with mental illnesses. This approach exemplifies the communication action approach.

A major advantage of the communication action approach is that members of the target audience gain skills, knowledge, and confidence in addressing their own problems. However, the health planner who is involved in critical planning needs to have a set of skills that are different from those needed to do rational or incremental planning. Also, because time and effort are needed to enable those with the problem to participate fully in the planning process, planning may proceed more slowly. The time needed to implement communication action lengthens the planning time line such that it is not useful in emergencies.

Comprehensive Rational Approach

The comprehensive rational perspective on planning is fundamentally a systems approach. It involves analyzing the problem by drawing on ideas from systems theory—namely, feedback loops, input and output, systems, and subsystems. The systematic, logical sequence of thought processes and actions employed explains why this approach is termed “rational.” Assumptions are made that the factors affecting the problem (the elements of the systems that contribute to the goals) are knowable and that virtually all contingencies can be anticipated. In this sense, this perspective is rational and logical. The approach is comprehensive in the sense that planners can take into account those contingencies and peripheral influences.

In the comprehensive rational approach, the planners set goals, identify alternatives, implement programs, and monitor results. This approach is clearly the dominant perspective of this text and of most courses in planning and evaluation. Health program planning—particularly of national initiatives—often reflects the effort to use a comprehensive rational approach. In fact, one of the planning principles outlined by Reinke and Hall (1988) is to be as objective as possible, given the context, and to use rationality rather than status or position as much as possible as a basis for power.

One benefit of this approach is that it facilitates obtaining information from stakeholders who might otherwise be reluctant to share information because it diffuses authority in favor of an information and rational base. The comprehensive rational approach allows planners to address issues faced by the entire system rather than just by subsystems; in this respect, it resembles quality improvement methodologies. Another benefit of the comprehensive rational approach is that it yields more information for decision making than does an incremental approach.

The comprehensive rational approach nonetheless has flaws. Forester (1993), for example, critiqued the cybernetic (systems) perspective that underlies this approach for

its failure to take into account the norms and values of individuals either involved in the planning process or affected by the planning process. Beneveniste (1989) acknowledged that the comprehensive rational approach separates planners from the political realities of the health situation. Forester (1993) also pointed out that the systems approach assumes that the means and the ends are known, which may not be the case. Unsubstantiated claims about the relationship of means (actions, interventions, process) to ends (effects, outcomes) adversely affect planning. The idea that planners know best in terms of which means are optimal reveals a normative perspective on planning that may be unacceptable to stakeholders with less expertise.

Strategic Planning Approach

The strategic planning perspective focuses on the organization and its ability to accomplish its mission in a fiscally responsible manner. While this approach is rarely used to address specific health conditions, it is particularly applicable to the infrastructure level of the public health pyramid. Through strategic planning, resources needed to address the health problem are identified and considered in terms of the mission of the organization. This widely used approach often affects program choices, as in healthcare organizations whose services center on one area of health, such as Planned Parenthood or the March of Dimes, or on one aggregate, such as the Boys and Girls Club or the American Association of Retired Persons. To some extent, the national goals and objectives set forth in *Healthy People 2020* are also examples of strategic planning.

This approach, when used by one member of a planning team, can affect the planning of the whole team. For example, the Centers for Disease Control and Prevention (CDC) awarded a large city a grant for a 1-year planning phase of a major community-based health promotion initiative. The funded agency involved numerous local community agencies in the planning

process. At one planning meeting, the chief executive officer of one of the community health agencies asked what his organization could do, what was being asked of his organization, and how his organization would benefit from participating in the health programs that would result from the planning process. His questions reflect implicit thinking about how to place his organization strategically within the field of contenders for grant money. His participation in the planning process also implied that his organization's mission was compatible with the general direction being taken to address the health problem.

The strategic planning approach has some clear advantages: It takes into account the context, whether competition or policy, and it has a slightly longer-term focus. Most strategic planning scholars recommend a time frame of approximately 5 years when planning because it typically takes that long to make strategic changes in programs and services. Strategic planning is a rational model or systematic approach to decision making, so decision points can be quantified, weighted, and sequenced, and then programmed into computer software that shows which option is the “best.”

Despite the capability to quantify mathematically the decision-making process, identifying the best option does not guarantee that the best decision option or program plan will be adopted. Human beings are irrational, with biases in how they think about probabilities and possibilities (Tversky & Kahneman, 1974). These human characteristics are usually not quantified in the decision models, but they are powerful forces in interpreting information and then shaping the actions taken. This irrationality can lead to situations in which the broad goals developed during the planning process are not, in fact, acted upon. Another disadvantage of the strategic planning approach is its lack of flexibility to respond to newly emerging environmental opportunities or threats (Egger, 1999). In addition, strategic planning, if done properly, is time and resource intensive.

Summary of Approaches

To some extent, all the approaches to health planning discussed in this section are likely to occur during the planning phase of addressing a health problem. One example is the efforts to increase adult immunization in Bowe County. The need to track immunizations necessitated the involvement of both state health officials and representatives from large health maintenance organizations (HMOs) in the planning process. A lobbyist used advocacy to gain passage of state legislation supportive of financing immunizations across all age groups, particularly for underinsured individuals from minority groups. The choice of vaccines was an apolitical decision based on research indicating which combinations were likely to result in the highest levels of protection for the population. Senior citizens were involved as informed consumers in shaping policy and tracking procedures, so communication action was also part of the planning process.

It is quite unlikely that all the individuals involved in the immunization planning efforts recognized the mixture of approaches being used in this community-based effort. Had they been aware of the approaches being used by various constituents, additional strategies could have been developed that would have made the planning process more effective, efficient, and palatable to parents, providers, and policy makers alike. Undeniably, a blend of approaches is typically needed, particularly in health program planning that aims to address more recalcitrant or population-based health problems.

Each of the six approaches to health planning represents a way to identify a problem, identify options, and make a choice—the classic definition of decision making. In other words, planning is decision making. From the perspective of an organization or agency engaged in health planning, health-planning activities can be framed in terms of managerial and organizational decision making. This fact

is important when assessing constraints for implementing the health program and developing the organizational resources.

► Models for Planning Public Health Programs

Although planning is generally described as a linear process, program planning is a cyclical activity, with recursive events requiring additional or revised courses of action for the life of a health program. Advances in defining planning in health care parallel advances in models for use by practitioners. Beginning in the mid-1980s, the CDC began to develop and promote methodologies for systematic approaches to health planning for those working in public health. These models are important for their structured approach to planning health programs and for synthesizing the knowledge available at the time about health and program planning. Internet resources related to Mobilizing for Action through Planning and Partnership (MAPP) and Protocol for Assessing Community Excellence in Environmental Health (PACE-EH) are listed at the end of this chapter.

Two events in the United States spurred development of a new type of planning: the September 2001 attacks and the 2005 Katrina hurricane disaster in the Gulf Coast, particularly in New Orleans. In the years following these events, greater emphasis was placed on preparedness planning. According to the U.S. Department of Homeland Security, planning for disasters ought to result in “plans that incorporate an accurate threat analysis and risk assessment and ensure that capabilities required to prevent, protect against, respond to, and recover from all-hazards events are available when and where they are needed” (U.S. Department of Homeland Security, 2007, p. 6). Although this definition is specific to disasters, the notion that planning results in a plan that takes into account data from an assessment can be generalized to health program planning efforts at the local and systems levels.

Mobilizing for Action through Planning and Partnership (MAPP)

Mobilizing for Action through Planning and Partnership (MAPP) is a strategic planning tool that helps public health leaders facilitate community prioritization of public health issues and identify resources for addressing them. The first phase of MAPP is to mobilize community members and organizations under the leadership of public health agencies. The second phase is to generate a shared vision and common values that provide a framework for long-range planning. The third step of MAPP involves conducting four assessments of four areas: community strengths, the local public health system, community health status, and the forces of change. The final step is implementation. MAPP materials can be ordered from the MAPP website, which can be accessed through the National Association of County and City Health Officials (NACCHO) website (NACCHO, n.d.).

Application of the MAPP process at a county level is possible and can result in a user-friendly, public document outlining a long-term health improvement plan at both the city and county levels. Since its release, MAPP has been established as a standard approach to planning and is widely used by local health departments (NACCHO, 2011).

Community Health Improvement Process (CHIP)

CHIP, developed in 1997 (Durch, Bailey, & Stoto, 1997), incorporates organizational performance monitoring and community assets, followed by strategic planning, implementation, and evaluation. Like the other models, it involves two phases: (1) problem identification and prioritization and (2) implementation. CHIP was the first model that linked the community assessment and planning to the performance measures of *Healthy People 2000* and the *Healthy Communities 2000* model standards. The principles and recommended processes for conducting CHIP are similar to those of MAPP and APEXPH—specifically, collection

of data, involvement and engagement of the community, and priority setting. CHIP continues to be used by both state and local health governments (Oconto County Department of Health and Human Services, 2008; Wisconsin Department of Health Services, 2010).

Protocol for Assessing Community Excellence in Environmental Health (PACE-EH)

The Protocol for Assessing Community Excellence in Environmental Health (PACE-EH), developed by CDC and NACCHO, focuses the community assessment on evaluation of environmental health conditions (CDC, n.d.). Thus, PACE-EH is a community environmental health assessment tool. The data gathered via this tool are then used to identify populations at risk and set priorities. A key feature of PACE-EH is the emphasis on health equity and social justice. As with the other public health planning models, PACE-EH focuses on building relationships with constituents and sharing the power and responsibilities with the community. PACE-EH has been used by county governments to improve the environment and thereby the health of their citizens.

In March 2006, a summit was held to address the barriers to the use of PACE-EH and the applications of this tool. The summary of the summit (Hubbard, 2006) highlighted the fact that more than 60 communities had used PACE-EH, primarily with a focus on improving air and water quality. Given the overall importance of air and water quality to individual and population health, using PACE-EH can be an effective approach to planning at local, county, and state levels. CDC continues to support use of PACE-EH, citing its contribution to many of the essential public health services (CDC, 2011; CDC, 2014).

In Summary

This brief overview of the models that have been used in planning public health programs highlights the ongoing evolution of thinking in this area.

TABLE 3-3 Comparison of Models Developed for Public Health Planning

	MAPP	CHIP	PACE-EH
Developers	CDC and NACCHO	NACCHO	CDC and NACCHO
Appropriate for which unit/level	Broadly, community health at city, county, and state levels	Local health departments and county level	City, county, and state levels
Distinguishing emphasis	Strategic planning, community involvement in planning process	Performance of local health departments, data used for planning	Environmental health, legal advocacy on environmental issues

Evident in this evolution is the development of tools for designing health promotion programs, particularly those focused at the community or population level. Each model has a slightly different strength and emphasis (TABLE 3-3). Many of these materials are designed for public health agencies and leaders as well as for community members. It is also worth noting that the content of the materials and the underlying philosophical perspectives are applicable to other types of health agencies that provide health programs to across the public health pyramid.

► Perspectives on Assessment

In the planning and evaluation model (FIGURE 3-3), the transition from planning to assessment occurs somewhat iteratively, with planning needs and approaches being informed by the data about needs and strengths. In turn, the assessment becomes directed by the planning efforts. The goal of a needs assessment is to guide and inform decisions related to problem prioritization and program development. Basically, a needs assessment is a procedure used to collect data that describe the needs and strengths of a specific group, community,

or population. To simplify the language, *needs assessment* is the term used in this chapter to encompass broadly both the deficit and the asset perspectives. Also, the term intended audience is used as a way to denote those for whom a program is intended. Only if the program is targeted at a true population is the term *intended population* used. The intended audiences can be a group (a relatively small set of individuals who interact), a community, a neighborhood, an *aggregate* (a set of individuals who share one characteristic in common, such as a school or a health condition), or a complete population. One of the first tasks in planning and conducting a needs assessment is to determine who is likely to make up the target audience or target population and in what larger unit they are situated.

Five types of models exist for conducting a needs assessment: the epidemiological model, the public health model, the social model, the asset model, and the rapid perspective model. Each one has its own intellectual perspective, as well as both advantages and disadvantages (TABLE 3-4). The distinctions among these types have been accentuated for the purpose of highlighting the possible different emphases. In practice, a model may be chosen but supplemented with elements of other models in accordance with resources and the purpose of the assessment.

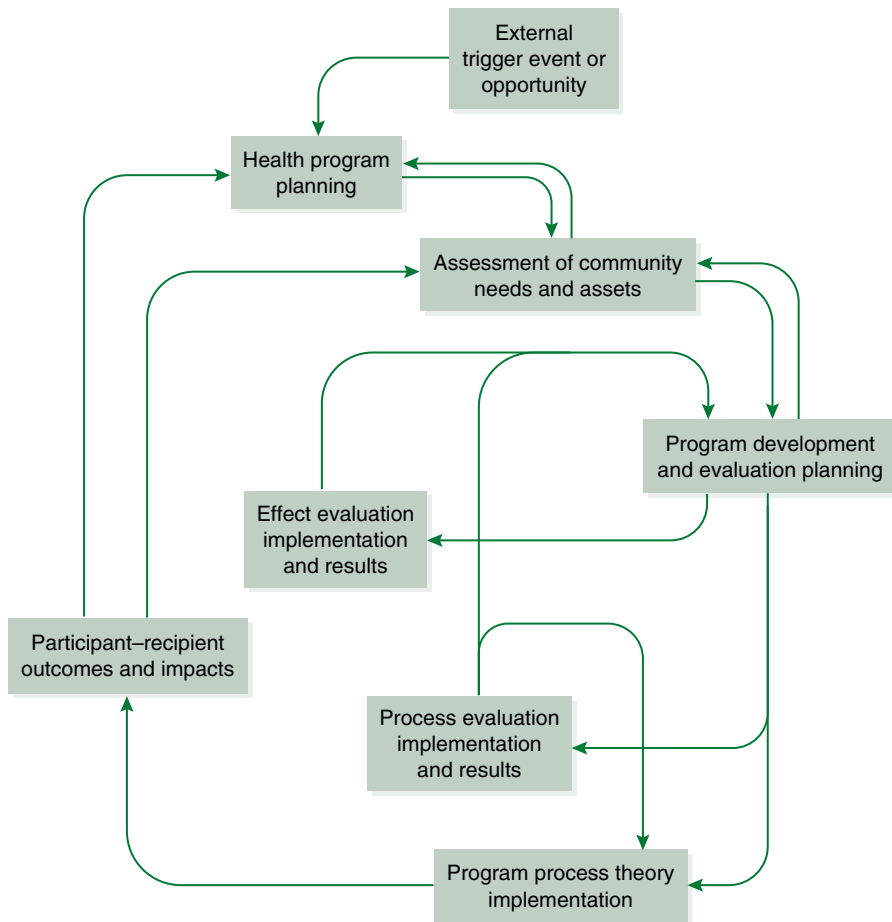


FIGURE 3-3 The Planning and Evaluation Cycle

Epidemiological Perspective

The epidemiological perspective on needs assessment focuses on quantifying health problems, using national data sets, and applying epidemiological methods and statistics. This model seeks to answer questions with an epidemiological focus, such as, “What is the magnitude of the problem?” or “What illness and disease trends are evident?” or “What patterns of selectivity are exhibited in the distribution of the problem?” Other questions stemming from this perspective would be, “Is the problem preventable?” “How treatable is the problem?” and “What is currently being done?” As these questions suggest, epidemiological models

often include a focus on identifying hazards, risks, and precursors to the health problem.

Examples of tools used in the epidemiological model are disease and death registries and national probability sample surveys such as the National Health Information Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES). An advantage of epidemiological models is that they provide data for assigning relative weights to the seriousness of a health problem, the importance of that health problem, and its prevalence. However, these models do not provide a breadth of data that might also be key in prioritizing health problems.

TABLE 3-4 A Comparison of the Five Perspectives on Community Health and Needs Assessment

	Epidemiological Perspective	Public Health Perspective	Social Perspective	Asset Perspective	Rapid Perspective
Population assessed	Populations	State and communities or counties	Populations, selected aggregates	Community, neighborhoods	Community, neighborhoods
Data sources	Registries, national probability sample surveys, existing national databases	State and local agencies, vital records	Individual surveys, national surveys	Rosters of agencies, focus groups, maps	Windshield surveys, existing data, interviews
Examples	National Health Interview Survey (NHIS), Healthcare Cost Utilization Profile (HCUP)	MAPP, CHIP, PACE-EH	U.S. Census, Bureau of Labor Statistics	Assets-Based Community Development Institute	Rapid Assessment and Response (RAR), Rapid Assessment and Response and Evaluation (RARE)
Types of needs assessed	Normative, expressed, and relative needs can be estimated	Normative and relative needs can be estimated	Relative needs can be estimated, perceived needs are directly determined	Perceived needs, perceived strengths	Normative and perceived needs
Advantages	Statistically sound and generalizable findings	Administratively sound; includes focus on constituent concerns	Statistically sound; provides information on factors contributing to health problem	Existing resources are identified	Quickly completed and provides basic information
Disadvantages	No information on perceived needs; local variations may not be captured or described	Relies on other data sources; perceived needs are not directly determined	Does not directly measure the extent of the health problem	Does not measure the extent of the health problem	Does not measure the extent of the health problem; may miss problems or causes

Public Health Perspective

Public health perspective focuses on quantifying health problems for the purpose of prioritizing the identified health problems being addressed with limited resources. From this perspective, the questions reflect an interest in the health of a population. Needs assessment questions from this perspective might be, “What is the seriousness of the problem?” or “What is the distribution of the problem?” or “What factors are contributing to the problem?” Other questions relate to the prioritization aspect, such as, “What is the perceived importance of the problem?” or “What resources are available to address the problem?”

Public health approaches to needs assessment typically rely on existing data and epidemiological data. Specific tools or models are used in the public health approach to needs assessment, such as MAPP and PACE-EH. These models provide a framework for determining which types of data ought to be collected as part of the needs assessment. There is considerable overlap between public health models and the epidemiological approach to needs assessment. Although these models are fairly comprehensive, they are somewhat weak in terms of their ability to account for the sociocultural aspects of health.

Social Perspective

The social perspective on needs assessment focuses on quantifying characteristics that contribute to the sociocultural, economic, and political context that may affect the health of individuals. This approach leads to questions focused on social ecological determinants of health. For example, a social perspective question might be, “What is the relationship among health problems and social characteristics?” or “Which social trends are evident in health behaviors?” Other questions stemming from this approach are, “What is the relationship between the use of social and health resources and the problem?” and “How have social and health policies affected the magnitude, distribution, or trends of the problem?” Key characteristics of a social perspective are a focus on

collecting data regarding social characteristics, such as income, and on collecting data from specific aggregates about specific social and economic topics, as is done by the U.S. Census. The social approach reflects a social determinant perspective on health and healthcare needs. In health care, planning based solely on social indicators is considered incomplete. Without the health indicators, the community needs assessment is incomplete. Nevertheless, assessments that include social ecological data do generate crucial information that helps identify antecedents or prior conditions leading to the health problems.

Asset Perspective

A fourth perspective on needs assessment is based on asset models, which focus on the strengths, assets, abilities, and resources that exist and are available rather than on the needs, deficits, lacks, and gaps between the healthy and ill. Assessments from this perspective seek to answer questions such as, “Which social and health resources exist within the community experiencing the health problem?” or “What do community members view as strengths and resources within their community?” Another question might be, “To what extent are the resources mobilized or able to be mobilized to address health problems?”

Asset-based community development (ABCD) has been recognized as potentially useful for creating healthy communities for children (Baker et al., 2007) and assessing mental health care resources in Ethiopia (Selamu et al., 2015). Indeed, asset models represent an important counterbalance to deficit models (Morgan & Ziglio, 2007) in efforts to overcome barriers to improving community health. Inherent in the asset perspective is a focus on the collective resources of individuals, particularly in the form of their social networks. In this regard, the asset perspective is aligned with the emphasis on social capital. An asset assessment could reasonably include an inventory of the social capital within the community being assessed for health problems.

The asset perspective on community assessment seeks to identify and then build on the capabilities of a community and thus resolve health issues. Although the asset models have some appeal, especially to community stakeholders, gathering asset data can be challenging. No generally accepted set of asset indicators exists. And rarely does asset information exist at the time of the assessment, making data collection necessary. These disadvantages contribute to asset models being less widely used as a sole approach to needs assessment, and they are poorly integrated into the more widely used models of needs assessment.

Rapid Perspective

In some circumstances, expediency is needed or desirable when conducting an assessment. Over the past 20 years, a consensus has emerged that a rigorous, reliable approach to rapidly assessing the health of a population or community is both possible and of value. Essentially, the rapid assessment approach uses multiple methods—such as focus groups, existing data, interviews, and mapping—to involve the community in rapidly developing and implementing needed health interventions. A rapid assessment seeks to answer the key question, “What are the most immediate and pressing needs that can be addressed with readily available resources?” As this question implies, the focus is on obtaining a quick response rather than ensuring the depth or breadth of the assessment.

Various terms are used as names for this type of assessment: rapid assessment and response (RAR) and rapid assessment and response and evaluation (RARE). These names denote the process by which an assessment is rapidly conducted and used to develop interventions. The RARE model consists of community participation, use of multiple methods and triangulation, and evaluation of both short- and long-term outcomes (Kamineni, Turk, Wilson, Satyanarayana, & Chauhan, 2011; McNall & Foster-Fishman, 2007). These basic steps are also part of the RAR model, which has been used as part of the response to natural

disasters such as Hurricane Katrina (Rogers, Guerra, Suchdev, & Chapman, 2006).

Stimson, Fitch, and Don Des Poznyak (2006), in a review of RAR studies conducted in several countries, found that RAR was effective in linking the assessment findings to the development and implementation of new or modified interventions.

► Types of Assessments

Before initiating a full-scale needs assessment, planners and planning groups need to be familiar with assessment as a process. This process begins by studying available data in order to gain a working knowledge of the community and the prevalent health problems. The familiarization assessment, as it is called, is a starting point from which to consider whether more data are needed and whether to proceed with conducting a larger-scale needs assessment. It is possible that a local agency has already done a needs assessment that might be adequate for the task at hand. Thus, becoming familiar with the community, the health problem, and existing assessments can save time and effort during the planning process. If it appears that an assessment is needed or is mandated by a regulating body, five types of assessments listed in Table 3-4 may need to be performed. Again, the distinctions are accentuated for the purpose of highlighting the possibilities, but the distinctions are blurred in practice.

Organizational Assessment

An organizational assessment determines the strengths, weaknesses, opportunities, and threats to the organization providing the health program. While an organizational assessment can be thought of as part of the logistics planning for a health program, it is critical to have a good sense of the organizational willingness and capabilities to provide a program to address the health problem under consideration *before* planning proceeds. In the PRECEDE-PROCEED model (Green & Kreuter, 2005), which is widely used by health educators, an organizational assessment is viewed

as a key component in planning health educational interventions. The organizational assessment seeks to answer the key question: “What is the capability and willingness of the organization to provide the health program?” The breadth of the organizational assessment is akin to the infrastructure assessment addressed in MAPP.

Data for an organizational assessment are gathered from members of the organization as well as from existing organizational records and documents. These data help determine the organizational feasibility of developing and implementing a health program—that is, whether adequate and appropriate resources are available and whether the health program fits with the organization’s mission and goals. One key aspect of the organizational assessment is the assessment of human resources within the organization, with particular attention being paid to their ability to meet the needs identified in the community needs assessment. This type of assessment can also identify changes needed within the organization as a prerequisite to providing the health program. In this way, the organizational assessment provides critical information for developing internal strategies to ensure the success of the health program.

The extent to which an organizational assessment is critical to the success of a health program is reflected in the organizational capacity study carried out by Roberts-Gray, Gingiss, and Boerm (2007). Their measures of organizational capacity predicted both the quantity of implementation activities and the quality of the implementation of a schoolwide intervention.

Marketing Assessment

Just as understanding the needs and assets of a target audience is key in program planning, it is equally important to understand the extent to which the target audience would be interested in the health program. The data for a market assessment can be collected at the same time as the needs assessment data are gathered. This type of assessment seeks to answer questions such as, “What will draw the target audience into the

program?” or “Which other health organizations are providing what types of services?”

Typical market analyses, such as those conducted in businesses, differ from marketing assessments for health programs in several ways. In health programs, marketing concepts are adapted to reflect the programs’ social and behavioral intervention focus. In addition, the price and packaging aspects that are addressed in marketing assessments play different roles in health programs.

Key data to be collected in a marketing assessment deal with competitive programs (available community resources) and overall interest in the intended program. Incorporating elements of the marketing assessment into assessment activities minimizes lost opportunities to collect key data and helps to provide a more complete assessment of the conditions that might affect both services utilization and health outcomes.

Needs Assessment

A needs assessment, in the more narrowly defined, traditional sense, is a means by which to determine the gaps, lacks, and wants relative to a defined population and to define a specific health problem. Data from a needs assessment are used to identify health problems or conditions that can be addressed by a health program. In this way, the assessment serves as a starting point for planning a program, such as was done by Norris and colleagues (2011) and then subsequently for planning, implementing, and evaluating the program (Palmer, Samson, Batra, Triantis, & Mullan, 2011).

A needs assessment provides health-related information that enables planners to gauge the priorities to be given to specific health problems, and it helps identify the trade-offs that inevitably arise in addressing one health problem rather than another. Often a needs assessment is done to answer the question, “What health problems exist, and to what extent?” Fundamentally, a needs assessment identifies normative needs. A needs assessment can be roughly equated to the “community health status assessment” in MAPP.

Typically a needs assessment is problem-oriented; thus, it tends to begin with a stated health

problem about which more information is wanted or needed. Another facet of needs assessment is the delineation of the community as a subsystem deserving of specific assessment. Sometimes more detailed information is needed about one aspect of a community, and a needs assessment focused on the community can provide such information.

Community Health Assessment

A community health assessment is used to establish the magnitude of selected health problems in a selected community, neighborhood, or other designated locality relative to the strengths and resources within that community, and to determine the priority the community gives to addressing the health problem. A community health assessment casts a broad net, encompassing all aspects of the community. It examines health and human service resources and assets, as well as the health problem and other community weaknesses. This type of assessment seeks to answer the question, “What are the key health problems, and which resources are available to address those health problems?” In this sense, a community health assessment encompasses and integrates each of the four assessment models described previously. From this integrative perspective, the chapter provides details on conducting a community health assessment.

The Affordable Care Act (ACA) includes a mandate for nonprofit healthcare organizations to conduct a community health needs assessment. According to the regulations, *community* is defined geographically, but it must also include the target population served by the healthcare organization. The healthcare organization must implement a strategy to address the gaps identified through the community health needs assessment. A greater emphasis on community health needs assessment is likely to generate new approaches to meeting needs as well as potentially new awareness of the community health problems.

Workforce Assessment

A workforce assessment is not commonly thought of as part of a community health assessment.

At the infrastructure level of the public health pyramid, however, workforce assessments are particularly germane. A workforce assessment seeks to answer the question, “Which human resources exist at which level of expertise to address the health needs?” A workforce assessment examines the current competencies among the workforce, trends and drivers of change related to the quantity and quality of the workforce, and building scenarios to understand the potential size of the gap between projected needs and the projected available workforce.

Across the health professions, workforce assessments conducted by scholars as rigorous research projects have revealed a dire situation for health care. In the foreseeable future, predictions of shortages continue for personnel in various health-related fields, such as nursing (American Association of Colleges of Nursing [AACN], 2014), occupational health (Lin, Zhang, & Dixon, 2015), and medicine (Dall, West, Chakrabrth, & Iacobucci, 2016). These predictions make it necessary that, before undertaking the development of any health program, a local assessment be conducted to identify the current and future workforces that will be tapped into to support the anticipated program. After all, there is no point in developing a great program on paper if it will not be possible to hire health professionals with the qualifications needed to make the program a success in the real world. Just as the organizational assessment is critical to determining the support of potential programs, so the workforce assessment is critical to determining the feasibility of attracting program personnel.

► Steps in Planning and Conducting the Assessment

No one right way exists for conducting a community health assessment. Nevertheless, certain basic steps exist across any approach to community health assessment. First, involve community

members in the development and execution of the community assessment. Second, define the community or population to be assessed, followed by making decisions about which data to collect regarding the nature of the health problem, such as the magnitude of the problem, precursors to the health problem, and demographic and behavioral characteristics. Third, collect these data using a variety of data sources and approaches. Once the data have been collected, the assessment and planning team must analyze the data using statistical procedures to arrive at statistical statements about the health problems in the population. Based on these data and the statistics obtained via their analysis, the last step seeks to develop a summary statement of the need or the problem that ties together the antecedent and causal factors of the health problem, along with the asset factors that counter the existence of the health problem.

Form and Develop the Team

Planning is a collective activity. The individuals involved at the various stages of the planning cycle easily influence the directions and decisions made. A key strategy for achieving successful planning is to have a visible, powerful sponsor. Given that politics of one form or another are inherently part of the planning process, having a backer who is recognized, respected, and influential becomes an essential element to successfully planning and implementing a health program.

Many factors influence the selection of the planning team members, including legal considerations (i.e., antidiscrimination laws and municipal mandates for advisory boards), the reasons for wanting to participate, the level and type of expertise that an individual can contribute to the process, the amount and type of resources that an individual can contribute to the process, the person's status as a current or potential user or client of the health program, and the person's role as an advocate for a group likely to be affected by the health program. Group size is another consideration, with groups of 10 to 15 being acceptable for the planning process

while also recognizing that an optimal size for a task force or work group is 5 to 7 persons.

One "law" of groups points out the challenges in working collectively: There will always be one person who is not a team player. Given this likelihood, knowing the strengths and weaknesses of the individuals involved helps to understand why each individual has been selected as a member. Having such information helps ensure that a planning group possesses a balance of strengths that will contribute to an efficient and effectively functioning group. Attention to the composition of the planning group also ensures that a breadth of knowledge and concerns is represented while eliminating disruptive individuals. Reinke and Hall (1988) also remind us that it is critical to have trained, skilled, and knowledgeable planning staff. Similarly, a growing consensus suggests that successful planning processes begin with developing the planning group's awareness, concern, and skills to address the problem at hand.

Different types of public health planning groups exist, but the most common type is a consortium. A *consortium* is a quasi-temporary body that is formed for a specific programmatic purpose and that has an independent sponsor, broad representation, and experts as members. Consortia are popular means to increase involvement of community members and to address (implicitly) the paradox of professionals not having the "right" solution.

Overall, the literature on participation in decision making reveals a pattern in favor of involving those affected by the decision in the decision-making process (e.g., Kegler, Painter, Twiss, Aronson, & Norton, 2009; Quick & Feldman, 2011). For this reason, the involvement of stakeholders throughout the planning, implementation, and evaluation cycle is highly recommended. It is also important to educate those involved in the planning process, using a communication action approach (as discussed earlier in this chapter). When those individuals who will be affected by the decision are involved in making it, their resistance to the change is likely to be diminished and they will begin to

“own” the program or plan, although participation does not guarantee ownership (Goodman, Steckler, Hoover, & Schwartz, 1993).

The last consideration in the team development stage is the selection of a leader for the planning group. Duhl (2000) argues that many types of leaders may be necessary for an effective planning process. A leader can emerge or be appointed based on his or her capability to function as an educator, a doer, or even a social entrepreneur. At any point in the planning process, or even during oversight of the program implementation, different individuals may be better suited to play a leadership role. Recognizing the fluidity of the leadership situation and acting on that recognition are both healthy and useful.

What may be less fully articulated, especially during the earliest stages, is the degree of formalization of leader selection. In other words, the planning group needs an acknowledged and standard process for designating a legitimate leader. This process can and does vary, ranging from the ad hoc emergence of a natural leader to the election of an individual from a slate of candidates according to formalized bylaws. Regardless of where along that continuum the group wants to be, the key will be to have an articulated and accepted process that facilitates the planning process rather than hindering team members' creativity and commitment. Ideally, each member of the planning group needs to be actively involved in ensuring that the process is open and agreed upon.

Create a Vision

The first step in planning, according to the American Planning Association, is to create a vision. Development of a vision is also one of the first steps in the MAPP models of assessment.

The nature of the mission statement appears to have a bearing on the success of healthcare organizations (Bart, 2007). For example, *Healthy People 2020* outlines a national vision of eliminating racial and ethnic disparities in health and increasing the quality of life. Whatever the trigger event, health program planners must create a vision with which existing and future

stakeholders can identify and to which they can devote attention and energy. A vision frames information for the stakeholders and helps identify economic assumptions that may affect the overall health program.

Part of the process of creating a vision of the final “product” is reaching a consensus on how to arrive at that final ideal. In this regard, one element of creating a vision is deciding on a system for prioritizing both problems and possible solutions to the highest-priority problems. How decisions are made—whether by voting, consensus, or complex algorithm—should be one of the first decisions of the planning group.

Involve Community Members

Before starting a community health assessment, the planners will ideally devote time to developing a strategy for involving members of the community to be assessed. The rationale behind involvement of those likely to be targeted by a program stems from a philosophy of empowerment as well as a practical concern with stakeholder and consumer reactions to the data. From the philosophical perspective of empowerment, involvement by community members enhances both their capacity to assist in the assessment and their ownership of the data gathered and results produced by the assessment. This theme of involvement is carried throughout the phases of planning and evaluating a health program.

From a practical perspective, involving those likely to be affected by the assessment has immediate and direct consequences for how the community health assessment evolves. The involvement of community members can even shape the questions addressed by the needs assessment. Of course, the strong views and bias of any one group can also become evident during the planning of the assessment. Program planners can use these revelations to begin to anticipate how those views might influence the interpretation of the data. By involving community stakeholder groups in the community health assessment, planners can uncover, acknowledge, and hopefully address their concerns.

Involving community members is rarely an easy venture. Numerous barriers to their involvement must be overcome: time constraints on busy individuals, competing interests for available time, parking problems, limited accessibility of the meeting location, lack of awareness of the opportunity for involvement, feelings of inadequacy or insecurity about being involved, and lack of day care for members' children. There is no one best way to increase involvement of community members. Instead, multiple strategies are needed, and the ones used are likely to evolve as the community health assessment proceeds. In addition to strategies that specifically address barriers to community participation, other strategies to increase involvement can include obtaining names of key individuals from agency personnel, providing food as an incentive, providing informal training or skills related to being involved, having specific tasks in which individuals can be involved, or having regularly scheduled meeting dates and times.

Sometimes, however, it may not be wise to focus on involving community members: when severe time constraints on completing the community health assessment exist, when severe fiscal constraints limit the scope of the assessment, when profound allegiances might affect the quality of interactions among community members, or when insufficient leadership skills exist to initiate and sustain community involvement. At other times, community members simply must be involved in the community health assessment: when there is a mandate from a funding agency for community involvement, when doing so will reduce the perception of being excluded, when insiders' connections and perceptions are needed to ensure a complete community health assessment, or when the goal is to have the community take responsibility for sustained implementation of the health program.

Define the Population

Delineating who is to be assessed is an important early step in conducting a community health assessment. The question of who to assess is often

influenced by who is doing the assessment. That "who" can be defined geographically, enabling the population of interest to be delineated by a site, such as workplace, location of residence, or school. Using locality as the defining characteristic is common, and ZIP code areas, census tracts, community areas, or legal boundaries are often used to define who is assessed. A state health department or a state health program will focus on the state population, whereas a small, local, nonprofit agency is likely to focus only on individuals who are potential customers. For example, the Traditional Indian Alliance in Tucson serves only Tucson's Native American population. Not surprisingly, its needs assessment was very limited in terms of both geography and population segment (Evaneshko, 1999). In contrast, a United Way organization in a large metropolitan area will assess the health and social needs of the population in its catchment area.

Using highly specific parameters to define "who" allows the assessment to be more focused and detailed. For example, Dickerson, Smith, Ahn, and Ory (2011) used the needs assessment data from a seven-county region of Texas to focus on the needs of individuals with diabetes. The data from these authors' work-site assessment enabled them to develop work-site health promotion programs that addressed both work-site and other health risks. This example shows that defining "who" based on a narrowly defined location may be a convenient means to access an aggregate, obtain detailed information, and tailor a health program very specifically.

In program planning and evaluation, the term *intended audience* refers to those for whom the program or intervention is designed and intended—in other words, those who are targeted by the program. The term *target population* is used if the program is intended for an entire population rather than a subpopulation. Put simply, the target audience includes all potential participants. Those who actually receive the program or intervention are referred to as the *recipient audience*. Thinking about this distinction between targets and recipients helps clarify who ought to be included in the community health assessment: Basically,

both groups should be included, and the target audience encompasses the potential recipients.

The parameters used to distinguish individuals for whom the program or intervention is intended from individuals for whom it is not intended become the boundaries of the target population. The target audience is usually some portion of the *population at risk*—that is, those individuals who have some social, physical, or other condition that increases their likelihood of an undesirable health problem or state. The term *at high risk* is usually reserved for those individuals with the highest probability of having an undesirable health state or outcome.

When conducting a community health assessment, planners must recognize that the boundaries of the target audience are likely to change with the collection and analysis of the data. For example, imagine that when a community health assessment is begun, an entire neighborhood or community area is viewed as the target audience. As epidemiological data and asset data are analyzed and interpreted, the planners may realize that only the black elderly residents or white adolescent residents or working mothers are at high risk for a health problem that can be addressed by the organization. This evolution of “who” from the broad boundaries to a more refined definition of the target audience is what ought to occur as a result of the community health assessment.

Define the Problem to Be Assessed

Just as the “who” of a community health assessment evolves with the collection of data and synthesis into information, so, too the “what” is likely to evolve as the assessment process unfolds.

Community health assessments are undertaken to address a purpose—and that purpose is never an altruistic desire to identify the breadth of the health problems that exist within a community. More likely, the community health assessment will be performed in response to a specific mandate. For example, the federal Maternal and Child Health Bureau mandates that all states conduct a

statewide needs assessment every 5 years as part of the states’ application for Title V block grant funding. Similarly, local jurisdictions may request a needs assessment as part of a strategic plan to be more responsive to changing health and social needs of their constituents and to address challenges created by budgetary constraints. If township or county officials face pressure to address the health or social problems of a particular group, such as adolescents, immigrants, or the disabled, those officials may request a community assessment to substantiate or discredit the needs of the group. This may be one way that the political aspect of program planning is played out.

For larger health agencies, organizations, or jurisdictions, a community health assessment might be conducted in response to calls for grant proposals that specify health problems or conditions as a high priority for being addressed. For example, the CDC has funded health promotion programs designed to reduce racial and ethnic disparities with regard to diabetes, cardiovascular diseases, infant mortality, HIV/AIDS, and immunizations. An agency wishing to compete for these CDC funds needed to identify within its community the specific racial or ethnic disparities as well as needs and assets relative to one of those five health problems.

Investigate

During the investigation phase of planning, data are gathered that will be used first to prioritize health problems and then to prioritize possible programmatic solutions. Generally, data relevant to planning health programs come from community assessment, population preferences, previous program evaluations, and research on possible interventions. The importance and possible scope of data collection carried out through community assessment are very critical factors that are addressed in subsequent chapters.

Two elements of the nonlinear nature of health program planning are worth introducing during this stage of planning. One is the need to focus on future considerations—specifically, interventions—even before the program direction

has been decided. The other is the need to be aware of the willingness of key individuals to support the planning and program process and to understand the quantification of health problems in terms of quality of life.

Interventions are actions that are done intentionally to have a direct effect on the health problem or condition. This broad definition of intervention includes medical treatments, pharmacological treatments, behavioral treatments, and health policy development, as well as education and skill enhancement, social support, and financial aid. In the healthcare realm, research resulting in EBP can be used to determine the effectiveness of potential interventions. Of particular concern are the sensitivity of the health problem to the intervention and the specificity of the intervention with regard to the health problem it addresses. Another consideration in determining the effectiveness of an intervention is the theoretical or conceptual logic underlying the way in which the intervention alters the health problem or condition. This issue is discussed in more detail in subsequent chapters.

Solutions, whether programmatic interventions or other ideas, often exist even before the problem is formally identified. Proponents of the solution might, for example, be waiting for a window of opportunity for “their” idea to be applied. Although having ready-made solutions available can certainly be helpful, too many individuals are inclined to jump on a particular solution bandwagon before the planning process has fully explicated the problem and all of its potential solutions. To the extent that any intervention or solution is well suited to a clearly defined problem, the planning process is effective.

According to Blum (1982), one factor that adversely affects planning is wishful thinking. In other words, solutions are sometimes based on idealistic and overly optimistic hopes rather than on scientific knowledge. This factor leads to the failure to examine the range of possible effective interventions or solutions to the problem. A key to avoiding this pitfall is the use of evidence-based medicine (EBM) and EBP, even though facts may not convince individuals with

strongly held beliefs about what is scientifically the right thing to do.

An additional benefit of focusing on interventions is that this tactic helps avoid an undue emphasis on needs assessment and data collection. Goodman et al. (1993) found that planning groups have a tendency to “frontload” the planning cycle, by devoting considerable time and effort to collecting risk and health problem data and conducting data analysis. Focusing on identifying realistic interventions balances the early planning stages with the later stages of implementation.

Prioritize

During the prioritization stage, data gathered during the community health assessment, along with the information on preferences and interventions, are integrated into a decision about what to address and how. During the establishment of priorities, the planning group is likely to face conflicts stemming from the group members’ different philosophies about how to establish priorities. As discussed earlier with regard to ethics, no single ethical approach to prioritizing issues is inherently right or wrong. Centering the team’s debates on what ought to be the guiding philosophical and ethical framework for making prioritization decisions thus can be an important step toward building consensus, trust, and mutual respect among stakeholders. Once these underlying principles have been agreed upon, a systematic, quantitative approach can be applied to determine health priorities. These priority-establishing techniques are covered in more detail in subsequent chapters.

Make a Decision

Inevitably, the priority ranking of health problems will not be acceptable to some stakeholders. As a consequence, the rankings may need to be revised by seeking stakeholder input or until a consensus is gained. Such an activity reflects the reality of blending the rational and political approaches to program planning. These decisions regarding which health problems to address serve

as the starting point for program development and then implementation.

Decisions about which health problems to address can fail for two major reasons. First, the organizational norms and institutionalized objectives may support conflicting priorities regarding health problems, limiting which interventions are acceptable. Second, the experts conducting the community health needs assessment may be biased, which will shape their findings. In other words, data from the community health assessment that are made available for planning may reflect the views of those who conducted the assessment rather than revealing the full scope of what exists as both strengths and problems in the community.

Once a health problem or condition has been chosen as the focus of a health program or service, a detailed implementation plan needs to be developed, along with a plan for conducting the evaluation. Planners should be aware that once a health problem has been identified, the composition of the planning group is likely to change. Members with vested interests will remain part of the team, while those with little expertise or interest in the chosen priority will fade away. At this point in the cycle, it may be important to revisit the group's composition and address why potential key stakeholders should become involved.

Implement and Continue

The planning cycle is complete after one full iteration, when the program is implemented, monitored to determine the extent of the implementation, and assessed for its effectiveness. For some health programs, implementation includes a termination phase or phase-out period, as happens with health programs funded for a limited time. Evaluation, whether of immediate effects or long-term outcomes, provides a basis for further program planning and completes one cycle.

Throughout the planning process, multiple foci are useful as evaluation end points: an epidemiological focus on the characteristics of the health problem; a scientific focus on identifying the best possible, feasible programmatic interventions; and a managerial focus on the planning

cycle and program implementation. If these foci are maintained, then feedback loops will develop more quickly, new triggers to additional planning activities will be perceived, and the evolution of involvement of stakeholders will be more rapid.

► Anticipate Data-Related and Methodological Issues

Data-related issues associated with conducting a needs assessment are well recognized and documented (Altschuld & Witkin, 2000; Lee, Altschuld, & White, 2007). Such issues must be addressed to enhance the quality of the data collected, thereby improving the accuracy of the community health assessment. Several types of methodological issues are discussed here in the context of conducting a community health assessment rather than within the more typical research framework.

First, when one is attempting to uncover what is occurring, there is a temptation to ask those experiencing the problem to provide information about the problem. As mentioned earlier, the trouble with this approach is that those receiving services may be systematically different from those not receiving services. Also, this approach is unlikely to uncover latent needs, meaning that some needs may not be manifested in an easily recognizable form. In the earlier example near the beginning of the chapter of the community that wanted a swimming pool, a latent need was to have an inequity addressed as manifested in community members' perceived need for recreational opportunities.

Another methodological problem is that asking potential consumers of the program about their needs has the potential to bias the answer. In other words, when asked about their needs, community members may take the opportunity to express all kinds of frustrations, wants, and needs. In addition, asking about needs, problems, and deficits does not allow for understanding the

community's assets, strengths, potential, resources, and capabilities. Thus, data collection methods are best designed to enable the collection of data that would fall on both sides of the equation.

Community health assessments can take as long as a year to accomplish, particularly if the assessment is comprehensive in scope and involves community members in the process. Unfortunately, time constraints are a reality that can heavily influence both the quality and the quantity of data collected as part of the assessment. Realistic strategies and designs for collecting data must match the time lines; otherwise, only partial data will be collected and will most likely be imbalanced in nature, leading to faulty conclusions.

In addition, the measures used to collect data must adhere to scientifically rigorous standards. Most important, the instruments used must have both validity and reliability. *Validity* is the degree to which that instrument measures what it is intended to measure. *Reliability* is the degree to which the instrument will yield the same results with different samples. Epidemiological measures, such as mortality, have high validity; death is rarely misdiagnosed. By contrast, the underlying causes of death as reported on death certificates are prone to both validity and reliability problems. The validity problems stem from conceptual issues of whether the cause of death ought to be the immediate cause or the underlying cause. The reliability problems relate to how each death certificate is completed and coded. Similarly, other epidemiological measures, such as adequacy of prenatal care, have been questioned with regard to validity and reliability. In terms of conducting a community health assessment, the point being made here is that no data are perfect, and these imperfections can lead to inaccurate numbers and hence faulty program planning decisions. Given this potential for error, planners should openly discuss the limits of the data and take reasonable scientific steps to obtain the best data possible.

The issue of determining from whom to collect the community health assessment data will always be important. This is a sampling problem. Sampling is a science, with numerous

sampling strategies being possible. In terms of conducting a needs assessment, the sampling decision depends on the degree to which individuals providing community health assessment data must be representative of the entire target population. The epidemiological and social approaches to assessment favor strategies that include individuals who look as much as possible like the target audience. However, if primary data are being collected, developing and employing strategies to achieve representativeness of the sample can be very difficult and costly. This issue is especially critical with hard-to-reach populations, such as intravenous drug users, emotionally abused spouses, or women who have experienced a perinatal loss. Less expensive but also less scientifically rigorous sampling strategies are certainly possible. The key decision to be made is how important it is to describe the population with a high degree of accuracy, based on data from less than the entire target population or target audience.

Overarching concerns are the cultural appropriateness of the data collection methods and the cultural competence of the data collectors and interpreters. Culture, language, and ethnicity all influence the responses of individuals to survey questions. The match between measurement approaches and the conceptualization of the neighborhood, for example, has subsequent implications for program intervention development (Nicotera, 2007).

Another key issue is the need to have community-level indicators—that is, data about the community, rather than data about individuals, that are then aggregated by community. For example, daily intake of fat is an individual-level indicator, and an average of percentage of daily intake of fat based on sampling of all residents in the community is still an individual-level indicator. The percentage of grocery store shelf space allocated to low-fat foods, by contrast, is a community-level indicator. The percentage of workers at a work site who smoke is an individual-level indicator, but the number of antismoking posters or announcements at the work site is a work site-level measure.

Very few ready sources of community-level measures or indicators of aggregates such as work sites or communities are available. Rather, it takes creativity, working with the community members, and careful consideration to develop community-level indicators and then to collect data reliably. But this extra work is worth the effort: For many of the health problems targeted by health promotion or disease prevention programs, what exists in the community will be extremely important as a component of assessing the antecedents and causal factors to the health problems.

In summary, there are five “principles” of collecting data for a community health assessment. One, collect data from more than one source; in other words, use multiple methods and multiple sources, and be multicultural. Two, involve members of the community in the design, collection, and interpretation of the community health data: Be inclusive and be empowering. Three, give full disclosure and then get informed consent from individuals from whom data are being collected: Be forthright, be honest, and be safe. Four, go beyond the available and collect data from unlikely but enlightening sources: Be creative, be inventive, and be open. Five, be as scientifically rigorous as time and other resources allow: Be scholarly, be interdisciplinary, and be systematic.

► Across the Pyramid

At the direct services level of the public health pyramid, health problems and conditions are viewed as individual problems that are best addressed by individual practitioners. Thus, at this level, assessments are of a focused type. A needs assessment is likely to concentrate on describing the magnitude of a specific medical problem. In addition, such an assessment would describe a subsystem of the community—namely, the diagnostic and treatment capabilities of the direct services providers within that community.

At the enabling services level of the public health pyramid, health problems and conditions are viewed as individual problems that are the direct results of nonindividual factors and that require

community-based or social services interventions. Assessments at this level hence would focus on describing the social context of those individuals with the health problem or condition, as well as the community subsystem in terms of local infrastructure capabilities and human services agencies.

At the population services level of the public health pyramid, health problems are viewed across a population. As a consequence, assessments at this level are likely to be epidemiological in approach, with attention being paid to describing the magnitude of various health problems or conditions. At the same time, social sciences approaches to assessment, using population data on social indicators, may provide valuable information about contributing and antecedent factors to the health problems and conditions.

At the infrastructure level of the public health pyramid, the concerns relate to the capabilities of the organization or the health delivery system to address the health problems or conditions at the direct, enabling, and population services levels of the pyramid. In the more ideal sense, the community health assessment is most appropriate for this level because it encompasses understanding the health problems and conditions within the social context of the target population as well as identifying the assets that are available to address those health problems and conditions. In addition, the organizational assessment fits at this pyramid level because it focuses on identifying the resources, capabilities, and mission currently available. The findings of the organizational assessment, when considered in conjunction with the findings of the community health assessment, ought to establish a solid foundation for garnering resources and planning health programs at the corresponding optimal level of the pyramid.

► Discussion Questions

1. Select one of the perspectives on assessment. In what way does that perspective change, alter, or influence each step in the process of performing a needs assessment?

2. Why is each type of assessment relevant to health program planning?
3. Consider the strengths of and the differences among the MAPP, PACE-EH, and CHIP models. Under which circumstances would you choose which approach? Access the websites for these programs (see the Internet Resources list), and compare the planning models in terms of how they demonstrate the planning approaches identified by Forester (1993).
4. Do an online search about community assessments. One suggestion is to begin with state health departments or the Community Tool Box (described below under Internet Resources). Which perspective on assessment is reflected in the context of the Internet site? Which of the steps described in this chapter received more or less emphasis by the authors of the Internet site? What implications does that shift in emphasis have for the problem statement?
5. Discuss the relevance of each perspective for developing programs at each level of the public health pyramid. What effects might choosing one perspective have on the level of the program subsequently developed based on its assessment results?
6. A health program planning committee wants to address various health problems by developing programs for each health problem at all four levels of the public health pyramid. What would be some implications of this perspective on setting priorities and on the nature of the subsequent community health assessment?

► Internet Resources

Community Tool Box

The Community Tool Box, found at <http://ctb.ku.edu/en>, has gained wide recognition for its information on community building, community

assessment, and program evaluation (evaluation framework, funders, developing evaluation plan, and more).

Group Dynamics and Community Building

This website offers a different perspective on thinking about communities, with a wealth of links to resources on community and team building. Find it at www.community4me.com.

National Center for Health Statistics (NCHS)

The NCHS website (www.cdc.gov/nchs/about.htm) has information on accessing and using existing national data sets.

Institute for Policy Research: ABCD

For information on the ABCD approach, check out the website www.northwestern.edu/ipr/abcd.html, maintained by the Institute for Policy Research at Northwestern University.

PRECEDE-PROCEED Model

The PRECEDE-PROCEED model can be viewed at the personal webpage of Lawrence Green, its author: www.lgreen.net/precede.htm.

MAPP

NACCHO's link to MAPP is found at <http://www.naccho.org/programs/public-health-infrastructure/mapp>.

National Association of County and City Health Officials: PACE-EH

NACCHO's webpage has links to PACE-EH resources and support at <http://www.naccho.org/programs/environmental-health/assessment-tools/pace-eh/resources>

Other PACE-EH Sources

You can look at the toolbox developed by New Mexico at this site: www.naccho.org/topics

/demonstration/PACE/upload/NMCommunity EnvHealthAss.pdf.

At the county level, Multnomah, Oregon has created a website describing its efforts. You can find it at www.pace-eh.org/index.shtml; find its publication at www.pace-eh.org/documents/pilot_nne.pdf.

Another example of PACE-EH being used by a county government is available at <http://www.delawarehealth.org/content.cfm?article=pace-eh>.

Disaster Preparedness

Examples of ethics in disaster preparedness planning in the United States can be found at http://ethics.missouri.edu/docs/Health_Ethics_Considerations.pdf for Missouri, and at <http://pandemic.wisconsin.gov/docview.asp?docid=14447> for Wisconsin.

References

- Alexander, E. R. (1992). *Approaches to planning: Introducing current planning theories, concepts, and issues* (2nd ed.). Philadelphia, PA: Gordon and Breach.
- Altschuld, J. W., & Witkin, B. R. (2000). *From needs assessment to action: Translating needs into solution strategies*. Thousand Oaks, CA: Sage.
- American Association of Colleges of Nursing. (2014, July 15). *Nursing shortage: Fact sheet*. Retrieved September 21, 2016, from www.aacn.nche.edu/media-relations/fact-sheets/nursing-shortage.
- Baker, I. R., Dennison, B. A., Boyer, P. S., Sellers, K. F., Russo, T. J., & Sherwood, N. A. (2007). An asset-based community initiative to reduce television viewing in New York state. *Preventive medicine*, 44(5), 437–441.
- Bart, C. (2007). A comparative analysis of mission statement content in secular and faith-based hospitals. *Journal of Intellectual Capital*, 8(4), 682–694.
- Bell, C., & Newby, H. (1971). *Community studies: An introduction to the sociology of the local community*. London, England: George Allen & Unwin.
- BeLue, R., Taylor-Richardson, K. D., Lin, J. M., McClellan, L., & Hargreaves, M. K. (2006). Racial disparities in sense of community and health status: Implications in community-based participatory interventions targeting chronic disease in African-Americans. *Journal of Ambulatory Care Management*, 29, 112–124.
- Benevise, G. (1989). *Mastering the politics of planning: Crafting credible plans and policies that make a difference*. San Francisco, CA: Jossey-Bass.
- Blum, H. L. (1982). Social perspective on risk reduction. In M. M. Farber & A. M. Reinhart (Eds.), *Promoting health through risk reduction* (pp. 19–36). New York, NY: Macmillan.
- Bradshaw, J. (1972). The concept of social need. *New Society*, 30, 640–643.
- Caan, W., Cassidy, J., Coverdate, G., Ha, M.-A., Nicholson, W., & Rao, M. (2015). The value of using schools as community assets for health. *Public Health*, 129, 3–16. Retrieved from <http://dx.doi.org/10.1016/j.puhe.2014.10.006>
- Centers for Disease Control and Prevention: <https://www.cdc.gov/>.
- Centers for Disease Control and Prevention. (2011). *Community Environmental Health Assessment (CEHA)*. Retrieved September 21, 2016, from www.cdc.gov/nceh/ehs/ceha/
- Centers for Disease Control and Prevention. (2014). *Improving environmental public health services: Performance to meet community standards*. National Center for Environmental Health. Retrieved September 21, 2016, from <https://www.cdc.gov/nceh/ehs/envphps/docs/improving-eph-serv-perf-comm.pdf>.
- Coffman, D. L., & BeLue, R. (2009). Disparities in sense of community: True race differences or differential items functioning? *Journal of Community Psychology*, 37(5), 547–558.
- Cottrell, L. S. (1976). The competent community. In B. H. Kaplan, A. Leighton, & R. Wilson (Eds.), *Further explorations in social psychiatry* (pp. 195–209). New York, NY: Basic Books.
- Dall, T., West, T., Chakrabrth, R., & Iacobucci, W. (2016). The complexities of physician supply and demand: Projections from 2014 to 2025. Retrieved October 11, 2016, from https://www.aamc.org/download/458082/data/2016_complexities_of_supply_and_demand_projections.pdf
- Denham, A., Quinn, S. C., & Gamble, D. (1998). Community organizing for health promotion in the rural South: An exploration of community competence. *Family and Community Health*, 21, 1–21.
- Dickerson, J. B., Smith, M. L., Ahn, S., & Ory, M. G. (2011). Associations between health care factors and self-reported health status among individuals with diabetes: Results from a community assessment. *Journal of Community Health*, 36(2), 332–341.
- Duhl, L. S. (2000). A short history and some acknowledgements. *Public Health Reports*, 115, 116–117.
- Durch, J. S., Bailey, L. A., & Stoto, M. A. (Eds.). (1997). *Improving health in the community: A role for performance monitoring IOM*. The National Academies Press. Retrieved September 7, 2012, from www.nap.edu/readingroom/books/improving/.
- Egger, E. (1999). Old ways of planning, thinking won't work in today's volatile health care industry. *Health Care Strategic Management*, 17(9), 18–19.
- Evaneshko, V. (1999). Mental health needs assessment of Tucson's urban Native American population. *American Indian and Alaska Native Mental Health Research*, 8(3), 41–61.

- Forester, J. (1993). *Critical theory, public policy and planning practice: Toward a critical pragmatism*. New York, NY: State University of New York Press.
- Goeppinger, J., Lassiter, P. G., & Wilcox, B. (1982). Community health is community competence. *Nursing Outlook*, 30, 464–467.
- Goodman, R. M., Steckler, A., Hoover, S., & Schwartz, R. (1993). A critique of contemporary community health promotion approaches: Based on a qualitative review of six programs in Maine. *American Journal of Health Promotion*, 7, 208–220.
- Green, L. W., & Kreuter, M. W. (2005). *Health promotion planning: An educational and environmental approach* (4th ed.). New York, NY: McGraw-Hill.
- Guerin, P. B., Allotey, P., Elmi, F. H., & Baho, S. (2006). Advocacy as a means to an end: assisting refugee women to take control of their reproductive health needs. *Women & Health*, 43(4), 7–25.
- Guyadden, D., & Seasons, M. (2016). Plan evaluation: Challenges and directions for future research. *Planning Practice & Research*, 31(2), 215–228. doi:10.1080/02697459.2015.1081335
- Hill, L. G., Maucione, K., & Hood, B. K. (2007). A focused approach to assessing program fidelity. *Prevention Science*, 8(1), 25–34.
- Hoch, C. (1994). *What planners do: Power, politics and persuasion*. Chicago, IL: Planners Press.
- Hubbard, B. (2006). Working to Build Healthy Communities: Community Environmental Health Assessments Using PACE EH. *Journal of Environmental Health*, 69(3), 32–33.
- Kamineni, V. V., Turk, T., Wilson, N., Satyanarayana, S., & Chauhan, L. S. (2011). A rapid assessment and response approach to review and enhance advocacy: Communication and social mobilisation for tuberculosis control in Odisha state, India. *BMC Women's Health*, 11, 463.
- Kegler, M. C., Painter, J. E., Twiss, J. M., Aronson, R., & Norton, B. L. (2009). Evaluation findings on community participation in the California Healthy Cities and Communities Program. *Health Promotion International*, 24(4), 300–310.
- Kono, A., Tadaka, E., Kanaya, Y., Dai, Y., Itoi, W., & Imamatsu, Y. (2012). Development of a community commitment scale with cross-sectional survey validation for preventing social isolation in older Japanese people. *BMC Public Health*, 12(1), 903. doi:10.1186/1471-2458-12-903
- Lee, Y., Altschuld, J. W., & White, J. L. (2007). Problems in needs assessment data: Discrepancy analysis. *Evaluation and Program Planning*, 30, 258–266.
- Lin, V., Zhang, X., & Dixon, P. (2015). Occupational therapy workforce in the United States: Forecasting nationwide shortages. *PM&R*, 7, 946–954. doi:10.1016/j.pmrj.2015.02.012
- Lindau, S. T., Vickery, K. D., Choi, H., Makelarski, J., Matthews, A., & Davis, M. (2016). A community-powered, asset-based approach to intersectoral urban health system planning in Chicago. *American Journal of Public Health*, 106, 10, 1872–1878.
- Lindsay, A. C., Sussner, K. M., Greaney, M. L., & Peterson, K. E. (2008). Influence of social context on eating, physical activity, and sedentary behaviors of Latina mothers and their preschool-age children. *Health Education and Behavior*, 36(1), 81–96.
- Lochner, K., Kawachi, I., & Kennedy, B. P. (1999). Social capital: A guide to its measurement. *Health and Place*, 5, 259–270.
- McNall, M., & Foster-Fishman, P. G. (2007). Methods of rapid evaluation, assessment, and appraisal. *American Journal of Evaluation*, 28, 151–168.
- Morgan, A., & Ziglio, E. (2007). Revitalising the evidence base for public health: An assets model. *Promotion and Education*, 2(Suppl.), 17–22.
- National Association of County and City Health Officials website: www.naccho.org/.
- National Association of County and City Health Officials. (2011). MAPP network: *MAPP in the news*. Retrieved September 7, 2012, from <http://mappnetwork.naccho.org/page/mapp-in-the-news>.
- Nicotera, N. (2007). Measuring neighborhood: A conundrum for human services researchers and practitioners. *American Journal of Community Psychology*, 40, 26–51.
- Nilsen, P. (2006). The theory of community based health and safety programs: A critical examination. *Injury Prevention*, 12, 140–145.
- Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. (2008). Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *American Journal of Community Psychology*, 41, 127–150.
- Norris, S. L., McNally, T. K., Zhang, X., Burda, B., Chan, B., Chowdhury, F. M., et al. (2011). Published norms underestimate the health-related quality of life among persons with type 2 diabetes. *Journal of Clinical Epidemiology*, 64(4), 358–365.
- Oconto County Department of Health and Human Services. (2008). *Community health improvement plan*. Retrieved September 7, 2012, from www.co.oconto.wi.us/i_oconto/d/whole_plan.pdf.
- Oliveira, V., & Pinho, P. (2010). Evaluation in urban planning: Advances and perspectives. *Journal of Planning Literature*, 24(4), 343–361.
- Ostadtaghizadeh, A., Ardalan, A., Paton, D., Jabbari, H., & Khankeh, H. R. (2015, April 8). Community disaster resilience: A systematic review on assessment models and tools. *PLoS Currents Disasters*. (1st ed.). doi:10.1371/journal.pcurrents.dis.f224ef8efbdfcf1d508dd0e4d8210ed
- Palmer, R. C., Samson, R., Batra, A., Triantis, M., & Mullan, I. D. (2011). Breast cancer screening practices of safety net clinics: Results of a needs assessment study. *BMC Women's Health*, 11, 6.
- Plough, A., Fielding, J. E., Chandra, A., Williams, M., Eisenman, D., Wells, K. B., . . . Maganz, A. (2013). Building community disaster resilience: Perspectives from a large urban county health department. *American Journal of Public Health*, 103, 1190–1197. doi:10.2105/AJPH.2013.301268

- Porepa, M., Chan, M., Huber, J., Lam, C. G., Au, H., & Birken, C. S. (2016). Creating a student-led health magazine with an urban, multicultural, resource-restricted elementary school: Approach, process and impact. *Paediatrics & Child Health* (1205-7088), *21*(3), 119–122.
- Quick, K. S., & Feldman, M. S. (2011). Distinguishing participation and inclusion. *Journal of Planning Education and Research*, *31*(3), 272–290.
- Reinke, W. A., & Hall, T. L. (1988). Political aspects of planning. In W. A. Reinke (Ed.), *Health planning for effective management* (pp. 75–85). New York, NY: Oxford University Press.
- Roberts-Gray, C., Gingiss, P. M., & Boerm, M. (2007). Evaluating school capacity to implement new programs. *Evaluation and Program Planning*, *30*, 247–257.
- Rogers, N., Guerra, F., Suchdev, P. S., & Chapman, A. S. (2006). Rapid assessment of health needs and resettlement plans among Hurricane Katrina evacuees—San Antonio, Texas. *Morbidity and Mortality Weekly Report*, *55*, 242–244.
- Selamu, M., Asher, L., Hanlon, C., Medhin, G., Hailemariam, M., Patel, V., Thornicroft, G & Fekadu, A. (2015). Beyond the biomedical: Community resources for mental health care in rural Ethiopia. *PLoS ONE*, *10*(5): e0126666. doi:10.1371/journal.pone.0126666
- Stimson, G. V., Fitch, C. J., & Don Des Poznyak, V. (2006). Rapid assessment and response studies of injection drug use: Knowledge gain, capacity building, and intervention development in a multisite study. *American Journal of Public Health*, *96*, 288–295.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *18*(4157), 1124–1131.
- U.S. Department of Homeland Security. (2007, September). National preparedness guidelines. Retrieved September 7, 2012, from www.fema.gov/pdf/emergency/nrf/National_Preparedness_Guidelines.pdf
- Wisconsin Department of Health Services. (2010). *Community Health Improvement Process & Plans (CHIPPs)*. Retrieved September 7, 2012, from www.dhs.wisconsin.gov/chip/

CHAPTER 4

Characterizing and Defining the Health Problem

In this chapter, the focus on the community health assessment phase of the planning cycle continues with a review of data collection options. This is followed by a discussion of the basic epidemiological analyses often used in community health assessment. After program planners have identified health problems through the assessment process, those health problems need to be summarized into statements and diagrams that facilitate both prioritization and the subsequent detailed program planning. A statement of the health problem can be developed in many ways and take a variety of forms. This chapter introduces an approach to developing a health problem statement, which is then used throughout this text as a means of linking the elements of the planning and evaluation cycle. The last phase of the assessment process is to prioritize the problems, preferably through a systematic and intentional process.

► Collecting Data From Multiple Sources

Numerous types and sources of data are used in a community health assessment. Each has the

potential to contribute to an understanding of the parameters of the health problem or condition. However, each has limitations and caveats that need to be considered.

Public Data

Data are called public data if they are available without a fee, are accessible by anyone, and have no personal identifiers. Public data primarily include national surveys, vital statistics, and census social indicators, which are gathered through the national surveys administered by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC), such as the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS). Secondary analyses of these public data sets can be used to extend data available for the community health assessment. Such national data have been helpful in making decisions about specific health problems, targeting populations, identifying barriers, and influencing health policy. They can also be used to create synthetic estimates, as explained later in this chapter.

Public data also include more and more often what is available via various and numerous social media outlets. Posts at public Facebook pages, images from photograph-sharing apps, and tweets from Twitter accounts can provide insights into cultural preferences, existing health information, emerging problems, and community resources. For example, Ramanadhan, Mendez, Roa, and Viswanath (2013) found that a substantial portion of community-based organizations use Facebook and Twitter to promote their respective organizations and their activities. Social media are used increasingly as the basis for developing visuals of the social networks among community organizations (e.g., Jung & Valero, 2016), which provide insights into the community power structure. Google Maps, as public data, can be useful in identifying commercial, religious, and educational resources, as well as public transportation routes. Such data complement and enrich the community assessment, especially when associated with geographic data distribution of health problems. Maps of the geographic distribution of health problems, as might be found at the website of a county or state health department, are yet another type of informative public data. Caveats to using public data exist. One caveat is that many of these data, especially the large national survey data sets, were collected years ago. The age of the data may influence their usefulness in understanding or identifying a current health problem. Another caveat is that social media, as public data, may not reflect an overall trend or sentiment, only that of a few vocal individuals. Although social media data tend to be quite current, its representativeness needs to be carefully considered.

Primary Data

Primary data are those collected by the community assessment or evaluation team, specifically to illuminate a need of interest. A wide variety of methods can be used to collect primary data, including interviews, surveys, community forums, focus groups, and interviews with key informants and service providers.

Three key points must be kept in mind when collecting primary data for a community health assessment. First, data from participants in a program are rarely used as the sole source of data for a community health assessment. Although the program participants can provide valuable insights into the perceived needs of the intended audience, that information must be considered in light of the fact that the participants are already in the program. This fact alone makes them potentially dissimilar to those persons targeted by the program. Second, rigor is required to obtain valid, reliable, trustworthy data. In most cases, only minor modifications to rigorous designs are needed for conducting a community health assessment. Third, primary data can, of course, be collected from members of the target audience, but providers can also provide valuable insights into the needs of the target population. Although data from providers are useful in identifying specific service needs of the target population, this information must be viewed as revealing the normative needs only. That is, providers are notorious for holding views of what is needed that differ from the views held by their clients. Naturally occurring discrepancies between providers' normative assessment of a problem and the clients' perceived needs can pose a particular challenge for the health program planning. Making program development decisions based only on provider data is likely to result in programs that are not attractive to the intended audience.

Observational Data

Unobtrusive (Webb, Campbell, Schwartz, & Sechrest, 2000) or nonreactive (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981) measures are also sources of data and are particularly relevant to community characteristics. These types of data constitute what is collected during a windshield or walking survey, as explained in the Community Toolbox (<http://ctb.ku.edu/en>). For example, walking around a neighborhood and observing how many blocks contain abandoned buildings or storefront churches is an unobtrusive

measure. Counting the number of liquor bottles in a garbage can, counting the number of billboard advertisements for unhealthy behaviors, estimating the ratio of bars and pubs to banks, watching the interactions among residents in a local bakery, and collecting local community newspapers are all examples of data collection of the least invasive nature. Each of these examples provides clues to the character, strengths, and problems in the community as a whole. The use of unobtrusive measures is inexpensive and can provide interesting data about what health problems exist and what may be contributing to those problems.

Archival Data

Archival data, as existing but not current, include newspaper articles, reports, and other types of agency records. On a local level, clinics, agencies, and nonprofit organizations may have data appropriate for a community health assessment they may be conducting. Archival data can provide the historical or cultural context of a problem, as well as information about past demand or need for a problem or historical information on previous program participants. One limitation to archival material is that the data may not include key information that is sought or may not be complete. Another potential problem with archival data is that the extent to which the data were initially collected accurately is unknown. These factors influence the data's overall usefulness and trustworthiness.

Proprietary Data

Another possible source of data is proprietary data—specifically, data that are owned by an organization and that can be purchased for use. For example, the American Hospital Association, American Medical Association, and health insurance companies own databases about their members that can contain information needed for a comprehensive community health assessment. Generally, medical record data are proprietary. Like archival data and public

data, the information that can be gained from proprietary data is limited to what has already been collected.

Published Literature

The published scientific literature is a good source of information, particularly for determining relative and normative needs. In other words, information may be available that allows for comparative statements about the health status in reference to other groups, professional health standards, or national goals. This inexpensive, reliable source of information should not be overlooked as a source of data when doing a community needs assessment.

Data Beyond Street Lamp

It can be important to collect data from sources that are not readily available. This practice is called “going beyond the streetlamp,” which derives its name from a little story.

One night, a man lost his keys. He began to look for them, crawling around on his hands and knees beneath a street lamp. Before long, a stranger stopped and asked the man what he was doing on his hands and knees. He replied that he was looking for his keys. The stranger offered to help and asked where he had lost his keys. The man replied, “Over there,” pointing to a dark area down the block just outside the bar. So the stranger asked, “Then why are you looking over here?” To which the man replied, “Because there is more light over here.”

The point of this story is that the information you need may not be the same as the data to which you already have access: You need to go beyond the streetlamp. Some of the sources of data just described are available under the streetlamp, whereas others are not readily available and require primary data collection. What determines the extent to which data need to be

collected from beyond the streetlamp are factors such as time constraints, fiscal resources, level of expertise, and endorsement or expectations of those who will be using the community health assessment.

► Collecting Descriptive Data

To understand the health problem and formulate a definition of the health problem or condition, it is necessary to collect data. Baker and Reinke (1988) suggest that, from an epidemiological perspective, four categories of information need to be collected as a prelude to health planning: the magnitude of the problem, the precursors of the problem, population characteristics, and attitudes and behaviors. These four categories provide a useful framework for organizing a community health assessment, especially when they are expanded to include elements from the public health, social, and asset perspectives.

Magnitude of the Problem

One category of information needed is the magnitude of the problem. The magnitude can be described in terms of the extent of the disease or health condition, the acute or chronic nature of the problem, and the intensity of the problem.

The extent of the health problem is described in terms of incidence and prevalence. The *incidence* is the rate at which new cases occur. The *prevalence* is the extent to which cases currently exist in a population. Incidence and prevalence, although typically used in reference to disease conditions, can be used to think about behaviors as well. For example, the number of new smokers among a defined group of adolescents (incidence) and the percentage of that same adolescent population that is currently smoking (prevalence) provide information that can be used to determine whether smoking is a problem of sufficient magnitude to warrant attention in the program planning effort.

The magnitude of a problem is also conveyed through measures such as rates and proportions. In epidemiological terms, these measures are a matter of numerators and denominators. The *denominator* is generally the total number in the population or the total number in the population that is potentially at risk. The *numerator* is generally the number of individuals who have the health problem or condition or who are actually found to be at risk. Using these basic numbers, a wide variety of commonly defined rates and proportions can be developed related to health. Increasingly, the rates and proportions for various health problems are available online at the websites for local and state health departments and federal agencies, such as the National Center for Health Statistics, which is housed within the CDC.

The ability to obtain accurate rates and proportions depends in part on the quality of the tests used to identify cases. Ideal tests have both high *sensitivity* (the extent to which there are no false negatives) and high *specificity* (the extent to which there are no false positives). Sensitivity and specificity are often used in reference to medical tests, such as occult blood tests, mammography, or urine tests for cocaine use, but they are also important characteristics of psychological and behavioral measures, such as the CES-D scale, which measures the level of depression in an individual (Radloff, 1977), and the SF-12 health survey, which measures overall health and functioning (Farivar, Cunningham, & Hays, 2007). The sensitivity and specificity of medical tests and of psychological or behavioral measures determine the extent to which a condition is accurately identified, which in turn influences the estimated incidence or prevalence rates for a given condition or behavior. In this way, sensitivity and specificity affect the accuracy of an estimated magnitude of a health problem or condition within a population.

Dynamics Leading to the Problem

Another category of data is information about the precursors of the health problem or condition.

As the community health assessment progresses, the planning group uses the data collected to generate lists of factors, conditions, situations, and events that in some way contribute to the health problem coming into existence and being observable. We broadly call the factors social determinants of health. All these factors are precursors or antecedents to the health problem. In addition, those factors, conditions, situations, and events that mediate, potentiate, or suppress the expression of the health problem may be uncovered during an assessment. While much may be known about a health problem from the scientific literature, the community health assessment is done to elucidate specific precursors that are unique to a locality, whether a neighborhood or a state, or to a target population.

Such information is necessary in order later to tailor interventions to the specific precursors of the health problem.

From an epidemiological perspective, the precursors to a health problem are understood in terms of agent, host, and environment. Baker and Haddon (1974), in studying childhood injuries, developed a model of factors associated with a health problem—namely, the human, physical, environmental, and sociocultural factors. **TABLE 4-1** is based on this model, albeit with the addition of the healthcare system as another element in analyzing the health problem or condition. Each cell in the table contains a definition of what might go into that cell, along with a few examples. For any single health problem that is the focus of a needs assessment, data can be placed into the

TABLE 4-1 Haddon's Typology for Analyzing an Event, Modified for Use in Developing Health Promotion and Prevention Programs

	Agent Factors	Human Factors	Physical Environment	Sociocultural Environment	Health System Environment
Pre-event	Latency	Genetic makeup, motivation, knowledge	Proximity, transportation, availability of agent (e.g., alcohol or drugs)	Norms, policy and laws, cultural beliefs about causes, family dynamics	Accessibility, availability, acceptability
Event (behavior)	Virulence, addictiveness, difficulty of behavior	Susceptibility, vulnerability, hardiness, reaction	Force	Peer pressure	Iatrogenic factors, treatments
Post-event	Resistance to treatment	Motivation, resilience, time for recovery	Proximity, availability of agent (e.g., alcohol or drugs)	Meaning of event, attribution of causality, sick role	Resources and services, treatment options, emergency response

Modified from Haddon, W., Jr. (1972). A logical framework for categorizing highway safety phenomena and activity. *Journal of Trauma*, 12, 193–207. Cited in D. C. Grossman. (2000), The history of injury control and the epidemiology of child and adolescent injuries. *The Future of Children*, a publication of the David and Lucile Packard Foundation., 10(1), 23–52.

cells in Table 4-1, thereby giving an overview and a preliminary analysis of precursors to the health problem or condition. This format is especially useful for infectious diseases and injuries. The Haddon (1972) model reveals the complexity of data that might need to be analyzed to understand the health problem or condition fully.

Population Characteristics

Population characteristics data, the third category, relate mainly to the social model of needs assessments. Obtaining this information involves collecting data on characteristics, such as distribution of age categories, income levels, educational levels, and occupation distribution within a community. If the “who” has been narrowly defined in terms of location, the population characteristics uncovered through this effort can be very specific. For example, if the assessment focuses on prison inmates, then their characteristics—such as types of crime committed, length of time incarcerated, or race—can become part of the population characteristics data collected for the community (prison) health assessment.

Attitudes and Behaviors

The fourth category of information concerns the attitudes and behaviors of the population being assessed, with particular attention being paid to the attitudes and behaviors of the target audience. Data about attitudes and behaviors help complete or flesh out the description of the factors related to a health problem.

Some attitudes and behaviors may be antecedents to health problems or conditions. For example, culturally held beliefs about illnesses, illness prevention, and treatments, as well as beliefs concerning appropriate health behaviors and the sick role, may all be important to understanding the health problem. Other lifestyle behaviors contribute to the existence of health problems. For example, secondhand smoke contributes to childhood asthma, whereas regular aerobic exercise contributes to reduced numbers of health

problems. Still, other attitudes and behaviors have a more direct, causal relationship with health problems. Distrust in medical providers and a failure to obtain preventive health services lead directly to severe morbidity conditions in some populations. Accordingly, attitudes toward health promotion and disease prevention behaviors as well as attitudes toward healthcare services and providers must be considered in order to have a comprehensive data set for a community health assessment.

Years of Life and Quality of Life

A number of measures have been developed to account not only for deaths but also the quality of years lived with an illness and the number of those years. **TABLE 4-2** summarizes these measures, whose definitions are drawn from a variety of sources. These measures are widely used to assess the value of new treatments and medication, which also makes them applicable to planning health programs. These burden-of-disease measures can also be used as part of a community needs assessment, albeit only if the population being assessed is sufficiently large to have stable statistics. Thus, measures of burden of disease might be used for larger states within the United States or regions of the United States. In the United States, national surveys, surveillance registries, and hospital discharge data are major sources that can be used to calculate burden-of-disease measures (Murry, 2013). These measures also are used in the economic evaluations of programs, particularly in cost–benefit analyses. Data on the negative consequences of health problems, beyond the familiar mortality rates, can be quite influential during the problem prioritization processes.

Definitions of quality of life include the notions of a perceived overall state of well-being across various domains such as sociocultural relationships and physical functioning or in relation to goals and expectations. Often quality of life is measured as it pertains to a particular illness or disease process—usually a chronic disease such as asthma

TABLE 4-2 Quality-of-Life Acronyms and Definitions

Acronym	Spelled-out Form	Definition
QALYs	Quality-adjusted life-years	Number of years of life expected at a given level of health and well-being.
DALYs	Disability-adjusted life-years	Number of years of life lost from living with a level of morbidity or disability.
YLL	Years of life lost	Number of years a person is estimated to have remained alive if the disease experienced had not occurred.
YPLL	Years of potential life lost	A measure of the impact of disease or injury in a population that calculates years of life lost before a specific age (often age 65 or age 75). This approach assigns additional value to deaths that occur at earlier ages.
HYE	Healthy years equivalent	Number of years in perfect health that are considered equivalent to a particular health state or health profile.
YHL	Years of healthy life	Number of healthy years of life lived or achieved, adjusted for level of health status.

or arthritis. Although each person has a sense of what constitutes quality of life, its measurement is complex; hence the plethora of quality-of-life measures that are available (Preedy & Watson, 2009). An issue with quality-of-life measures, as Kaplan (1996) stressed, is that quality of life is multidimensional, so the measures must address the relative importance of the many dimensions of quality of life. The choice of which measure to use in health program planning depends on the resources available for the assessment phase, the sophistication of the planning team, and the role played by the rational approach in the planning process.

The length of life is as important as its quality. Because individuals may live with the same health condition for varying lengths of time, the

length of life as affected by that health condition is what becomes important. In other words, the assessment must consider the quality of the life as lived with the health condition. The measures known as quality-adjusted life-years (QALYs) and disability-adjusted life-years (DALYs) were developed specifically to give a numeric value to the quality of years of life (Table 4-2). These composite scores are used with populations and therefore have the advantage of being indifferent to individual preferences. Because the number of years for which quality can be adjusted is naturally shorter for older persons, however, QALYs and DALYs mathematically discriminate against the elderly. Nonetheless, the use of DALYs reveals the extent to which diseases affect the years of life (TABLE 4-3).

TABLE 4-3 Global Leading Causes of Disability-Adjusted Life-Years (DALYs) and Years of Life Lost (YLL)

Rank	DALY Causes	% of DALYs	DALYs per 100,000
1	Ischaemic heart disease	6.0	2,342
2	Lower respiratory infections	5.4	2,076
3	Stroke	5.2	1,998
4	Preterm birth complications	3.9	1,515
5	Diarrheal diseases	3.6	1,409
6	Chronic obstructive pulmonary disease	3.4	1,306
7	HIV/AIDS	3.4	1,299
8	Road injury	2.9	1,113
9	Unipolar depressive disorders	2.8	1,081
10	Birth asphyxia and birth trauma	2.7	1,054
	Total % for top 10 DALY causes	39.2%	
Rank	YLL Causes	% of YLL	YLL per 100,000 population
1	Ischaemic heart disease	7.8	2,212
2	Lower respiratory infections	7.2	2,042
3	Stroke	6.8	1,933
4	Preterm birth complications	5.2	1,472
5	Diarrheal diseases	4.6	1,290
6	HIV/AIDS	4.4	1,234
7	Birth asphyxia and birth trauma	3.4	965
8	Road injury	3.2	916

Rank	YLL Causes	% of YLL	YLL per 100,000 population
9	Chronic obstructive pulmonary disease	3.1	871
10	Malaria	2.5	718
Total % of YLL from top 10 causes		48.2%	

Sources: World Health Organization. (2014). Summary: YLL (000s) by cause, in WHO Regions (a), estimates for 2004. Retrieved September 11, 2012, from www.who.int/healthinfo/global_burden_disease/risk_factors/en/; World Health Organization. (2004). The Global Burden of Disease 2004 Update. Accessed September 11, 2012, from www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf.

A slightly different perspective is based on the number of years of life that are lost due to a health condition. *Years of life lost* (YLL) reveals the number of years lost at the end of life because of a health condition. The shortened life span could be due to acute or chronic health problems, chronic environmental conditions, or injuries. *Years of potential life lost* (YPLL) is a similar measure but indicates the number of years of life lost at the beginning of life, such as the shortening of life caused by neonatal sepsis or childhood drowning.

All of these quality-of-life and life-year measures are particularly useful in public health assessments. They assist program planners in deciding which health condition warrants health promotion or disease prevention programs, particularly when resources are severely limited. Unfortunately, these measures are difficult to calculate and exist for nations or very specific populations. Therefore, it is challenging to use these measures in local community assessments.

► Statistics for Describing Health Problems

Data collected from the community health assessment need to be analyzed and interpreted, particularly the primary data that were collected.

For example, if surveys were conducted or if data were abstracted from medical records, those data need to be analyzed. This section briefly reviews statistical approaches, with particular attention to epidemiological considerations relevant to community health assessment. More details about quantitative data analysis are provided elsewhere in this text as well as in statistical textbooks. The intention of this section is to relate what is learned in statistics and epidemiology to the community health assessment.

Given that most community health assessments involve some population-based data, it is worth reviewing basic epidemiological techniques here. More complete and in-depth presentations are available in traditional epidemiology textbooks, such as those by Friss and Sellers (2014) or Fos (2011). For a direct application of epidemiology to community health assessment, Dever's (1980) book is a classic. However, the more recent publication by Dever (2006) also cover basic epidemiological techniques but from the point of view of healthcare executives planning for population health. Health program planners would do well to have at least one of these texts on their bookshelves for quick reference. With the widespread availability of computer spreadsheet and database programs, the calculation of most statistics is less a matter of doing the math and more a matter of understanding which numbers to use and how to make sense of the numbers generated by the software.

Descriptive Statistics

Descriptive statistics—the fundamentals of statistics—are a family of statistics that portray the distribution of values for a single variable. These statistics provide an amazing wealth of information but are often underappreciated for their ability to communicate important information simply.

The simplest descriptive statistic is the frequency, or count, of occurrences. Based on the frequency, two other informative types of descriptive statistics can be calculated. One type measures central tendency, such as the mean (average), whereas the other type measures dispersion as variance and standard deviation. Standard deviation is related to the range of values in the data and thus indicates the dispersion of the data. Remember that 68.3% of data are contained within one standard deviation, 95.5% within two standard deviations, and 99.7% within three standard deviations.

Descriptive statistics are easy to calculate with a calculator or with spreadsheet software, such as Excel. They are often presented in the form of graphical displays of frequency such as bar graphs. A bar graph of frequencies provides a rough picture of the distribution, thereby revealing visually whether the data approximate the normal curve.

Odds Ratio and Relative Risk

Two statistical tests that help estimate the likelihood of having or getting a given health problem are the *odds ratio* (OR) and the *relative risk* (RR). The odds ratio is calculated as the odds of having the health problem if exposed divided by the odds of having the problem if not exposed. The relative risk is calculated as the cumulative incidence in the exposed population divided by the cumulative incidence in the unexposed population. In conducting a community health assessment, planners mostly obtain the odds ratio and relative risk from published studies because having data on exposure usually requires epidemiological research.

Relative risk ranges from 0.0 to infinity: the larger the relative risk, the greater the chance of developing the health problem with exposure. Similarly, the odds ratio ranges from 0.0 to infinity. Odds ratios from 0.0 to 1.0 indicate a protective effect, whereas odd ratios greater than 1.0 indicate an increased likelihood of having the health problem. The larger the odds ratio, the more likely one is to have the health problem. Although the odds ratio can range to infinity, in practice it rarely exceeds 10. Relative risk compares two cumulative incidences, thereby providing a direct comparison of the probabilities. This makes the relative risk measure preferable to the odds ratio (Handler, Rosenberg, Kennelly, & Monahan, 1998). The odds ratio does not use the population in the denominator, making it less accurate than the relative risk. However, when the health problem is rare, the odds ratio begins to approximate the relative risk.

Both the relative risk and the odds ratio are used widely in epidemiology and thus are likely available for use in community health assessments. Both convey information about the comparative influence of factors or exposure variables on health outcomes. Having this information available then allows planners to prioritize which causal or exposure factors to address in a health program.

Population Parameters

The *confidence interval* (CI) indicates the upper and lower range of values between which the value for the true population is likely to fall. It helps to understand the likelihood that the score or mean value for a health condition derived from a sample is similar to the value in the true population. Confidence intervals, like standard deviations, provide a level of assurance about whether the mean or score value for the variable reflects the value for the whole population. For example, if a score is within the CI, then the value falls within a range that is reflective of the larger population. However, if the value falls outside the CI, then that score can be viewed as being important because it is different from the value in the general population.

Confidence intervals play a valuable role during the community health assessment by focusing attention on values that are unusual and thus merit attention. They also provide a clue about relative need because values that fall outside the CI are “abnormal” relative to the population. Naturally, the reverse of this can also be true: A value falling within the CI can be a cause for concern.

Tests of Significance

A test of significance is done to assess whether the probability is high or low that the statistical result can be accepted as being true. The test of significance is the same across the types of statistics used, and the interpretation of significance is the same. The first step is to set the alpha level, which is the probability of rejecting a null hypothesis, when in fact it is true. Commonly, the alpha is set at 5%, and the null hypothesis is stated as there will be no difference between groups. Using a table of critical values for the type of statistical test used, one can then determine whether the statistical result was above or below the cutoff for the alpha.

A critical issue for program planners in analyzing community health assessment data might be to determine whether the difference between two communities or two groups is just a random variation or whether the difference is sufficiently large to suggest something else is contributing to their difference. If the two groups are compared, the p -value gives the probability of falsely claiming that the groups are different. For example, if a test of significance is reported as $p = .05$ for a comparison of two groups, then there is less than a 5% chance that the two groups were actually the same. Remember, however, that a statistical significance does not automatically mean there is a clinical significance.

Geographic Information Systems: Mapping

Historically, an element of the epidemiological assessment model has been the geographic

mapping of health problems or population characteristics. As mapping software has become more accessible, the usual map-based display of the distribution of health problems or conditions can be done more readily at any level, such as by state, county, census tract, ZIP code, or street address. Mapping at very specific levels of geography provides an extremely refined picture of what is where. Geographic mapping technology can accommodate a variety of types of data, including survey data; social media data; and address data, such as that of community assets. The integrated use of these data lead to a useful and insightful statistical analysis of the intersection of needs or problems and existing resources. For example, Cebrecos et al. (2016) investigated the association between food environments and physical activity environments to identify neighbourhoods with both healthy food outlets and walkable streets. Another option is to overlay rates of chronic health conditions with location of health services resources, as can be done at a CDC website (<http://www.cdc.gov/dhdsp/maps/atlas/index.htm>). Such a map helps visualize that the highest rates of chronic health conditions are found in geographic areas with the lowest density of health services. Mapping health problems and suspected factors in the causal path provides very engaging information and can be crucial in reaching a consensus or attracting the attention of key stakeholders.

Advanced spatial analytics can provide quantitative information, including statistical evidence of association between location and other variables, and statistical difference in rates relative to adjoining geographies. As the geographic analysis software becomes more mainstream, the statistical analysis and interpretations will become more familiar to health program planners and evaluators.

Small Numbers and Small Areas

Small numbers are a big problem, whether one is looking at epidemiological data or social data. In particular, those conducting the community assessment and health planners for rural areas face

the issue of how to portray rates. The problem is a simple one: If a geographic area has a small population (denominator), then a small variation in the occurrence of a health problem (numerator) will inevitably lead to a large change in the rate or proportion of that health problem. This instability of the rate affects the conclusions that can logically be drawn from the data. The same statistical problem also occurs when the analysis focuses on a small geographic area, such as a parish, county, or legislative district. Another concern with small numbers is the potential breach of confidentiality, given the possibility of identifying an individual when the number of cases in a row by column cell is small. Guidelines for maintaining confidentiality exist (e.g., VanEenwyk & Macdonald, 2012) and ought to be followed.

Small numbers can also be a problem if the data collected are of a social or qualitative nature, as might be the case in an asset assessment. If the number of respondents to a community survey or the number of participants in a community focus group is small, then the information those individuals provide has a higher likelihood of not being representative of the range of views and opinions in the community.

Once data are collected, those conducting the assessment rarely have an opportunity to go back and gather more data. Given this caveat, careful planning and execution of the data collection must be done to avoid having too few respondents.

Several statistical techniques exist for addressing the small-numbers problem, utilizing counts, rates, or proportions (Dever, 1997) or pooling years of data (Cawley, Schroeder, & Simon, 2006). Additional techniques continue to be developed as well (Yu, Meng, Mendez-Luck, Jhawi, & Wallace, 2007). One set of techniques focuses on comparing the small area (population) with a larger area (population) or a standard. Another set of techniques is based on comparing two small areas. Yet another approach is to use data from multiple time periods, which may cumulatively produce a sufficient sample size to make comparisons either across time periods or with another small area using a similar time period.

One study of immunization rates (Jia et al., 2006) is illustrative of the challenge inherent in dealing with small numbers. Jia and colleagues color-coded county data to reflect whether the county was more than two standard deviations above or below the state mean, between one and two standard deviations above or below the state mean, or less than one standard deviation above or below the state mean. Their choice of displays on a map of the United States demonstrates the creativity that may be needed to deal with small numbers and small-area data.

Epidemiology Rates

TABLE 4-4 is a summary of commonly used epidemiological rates, with their corresponding numerators and denominators. As the table entries suggest, the epidemiological model relies heavily on having accurate counts that can be used in the denominators and numerators. To the extent that data are available, key elements of the epidemiological model are incorporated into most community health assessments.

► Stating the Health Problem

Data collected for the community health assessment can be organized in a variety of ways, such as a community profile, wellness profile, behavioral profile, or service profile (Paronen & Oje, 1998), or a community diagnosis (Muecke, 1984). Regardless of which format is chosen, the community health assessment ought to lead to a statement of what was found, phrased so that stakeholders, constituents, community members, and multidisciplinary health professionals can understand each health problem identified.

Diagramming the Health Problem

In health program planning, understanding what causes a health problem and how those causes

TABLE 4-4 Numerators and Denominators for Selected Epidemiological Rates Commonly Used in Community Health Assessments

Rate	Numerator	Denominator	Per
Crude death rate	Total number of deaths in a given period	Total population	1,000
Cause-specific death rate	Number of deaths due to a specific cause in a given time period	Total population	100,000
Disease-prevalence rate	Number of currently existing cases with the disease, during a given time period	Total population during that time period	(Varies depending on the size of the numerator)
Birth rate	Number of live births in a given time period	Total population	1,000
Fetal death rate	Number of fetal deaths at 28 weeks' gestation or more that occur in a given period	Number of fetal deaths at 28 weeks' gestation or more plus number of live births that occur in a given time period	1,000
Neonatal death rate	Number of deaths of infants 28 days old or younger that occur in a given time period	Number of live births that occur in a given time period	1,000
Infant-mortality rate	Number of infants (age birth to 1 year) who died in a year	Number of births in a year	1,000

lead to the problem is portrayed in a diagram or conceptual model that organizes key factors along a general sequential time line, as well as a written statement. Articulating the problem both visually and verbally can help in gaining clarity about the causes and possible solutions to the problem. In addition, having both the visual and verbal description allows the planner to communicate better with both those who prefer pictures and those who prefer text.

Each health problem has its own unique set of precursors given the specific context. The model of the health problem is distinct from a logic

model or a root cause analysis. The key distinction is that a logic model focuses on organizing the delivery of the program, whereas the model of the health problem focuses on understanding what causes the health problem. A root cause analysis includes a range of possible causes of a problem, but it does not rely on data or sequence those possible causes in terms of cause and effect. Attention to the causes of the health problem is critical if program planners are to select the best point of intervention and the appropriate intervention for that point. In this way, the model of the health problem contributes not only to

the development of a logic model but also to tailoring the programmatic intervention and to designing the evaluation of the program's effect.

Elements of a Causal Theory

The model of the health problem brings together, in a visual display, the key factors that were identified from the community health assessment as being important to the health problem. The combination of these factors ought to explain or hypothesize about what causes the health problem. For this reason, it is called a causal theory. For some health problems, an existing causal theory may be applicable to the current circumstances. If it is not, a new causal theory will need to be created. The decision of what to include in the causal theory and what to exclude intentionally from the model has ramifications through the program planning and evaluation cycle: It guides the intervention choice, establishes the parameters for the evaluation of the program effect, and influences the statistical analyses of the evaluation data.

Throughout this text, the same approach to displaying a causal theory (shown in **FIGURE 4-1**) is used for the sake of illustration. Each program, reflecting the unique perspective of its set of stakeholders, may develop its own approach to displaying a causal theory visually. The template presented in this chapter has been carefully crafted to include four key elements to be considered in developing any causal theory: existing factors,

causes, mediating factors, and moderating factors. Not included in the causal theory are factors, conditions, variables, and elements that may exist but are not immediately relevant to the health problem or that are so complex that they are not contenders for programmatic intervention. In other words, the process of developing the causal theory also is a process of narrowing the problem and of prioritizing. As with all processes, the creation of the causal theory is an iterative, evolving, and ongoing process that cannot be completed in a single short session.

Precursors can be thought of as existing physical or contextual factors as well as the causes of the health condition. *Required antecedent factors* are those elements that must be present for the health problem to come into existence or are direct precursors of it. For example, required factors may include genetic predisposition, being in the right place at the right time, prior exposure and vulnerability, or legal or policy conditions. From an asset perspective, required factors might also include variables such as the political clout of the local representatives or the existence of economic empowerment zones. The predisposing factors of health services utilization models (Anderson, 1995; Anderson & Aday, 1978; Green & Kreuter, 1991) generally fall into the required antecedent factors category.

Causal factors are those elements that influence whether the health problem will manifest itself, given the presence of the required antecedents. Depending on the health problem, causal factors

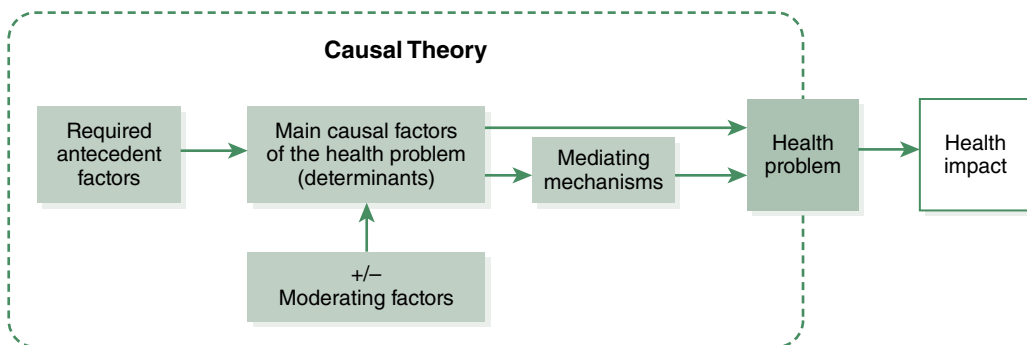


FIGURE 4-1 Generic Model of a Theory of Causes

might be exposure to the health hazard, susceptibility, or the virulence of the hazard. From an asset perspective, causal factors might include health knowledge, the existence of healthy food choices in local grocery stores, the existence of environmental pollutants, the existence of road safety features (intersection lights), or the accessibility and availability of local health and social services agencies. These causal factors could also be called the determinants of the problem because they are directly responsible for outcomes. In recent years, however, the term *determinants* has been more widely applied to encompass a host of social and ecological factors that lead to health problems. To avoid confusion and to be more specific, the term *causal factors* is used throughout this text to refer specifically to the factors identified from the community assessment as leading directly to the health problem.

Causal factors include, for example, the water that determines whether a seed will sprout, the hole in a boat that determines whether the boat will sink, and the presence of potassium in the blood that determines whether the heart muscles will contract. Nevertheless, as these concrete examples suggest, the situation involves both required antecedent factors and factors that lead one to say “yes, but. . .” Those “yes, but” factors can be sorted into two types: moderating factors and mediating factors.

Moderating factors are those elements that have the potential either to exaggerate or to lessen the presence of the health problem. Again, depending on the health problem, these factors might consist of laws and policies or social support. Such factors generally act to affect the causal factors. Complex and interacting relationships exist among required antecedents, causal factors, and moderating factors. In a model of the causal theory, moderating factors are shown as possibly influencing the identified causal factors. By their nature, moderating factors can either increase, potentiate, exaggerate, and stimulate or, alternatively, lessen, diminish, and suppress the presence or strength of the causal factors. Factors that would be classified as enabling and

reinforcing factors (Anderson, 1995; Anderson & Aday, 1978; Green & Kreuter, 1991) must be reevaluated for their role in the causal theory, as many are likely to function as moderating factors.

Mediating factors come between causes and outcomes. In fact, without the mediating factor, the causes will not result in the health outcome. In other words, without this process or mechanism, the causal factors cannot cause the health outcome. Depending on the health problem, there may not be any mediating factors. For example, if an individual has the genetic mutation that causes cystic fibrosis, the disease will appear—there is no mediating variable. However, if the health outcome is defined as longevity for persons with cystic fibrosis, mediating factors would include quality of health care and individual response to treatments. In contrast, if someone has a stroke, both morbidity and mortality depend on the response time and quality of the emergency medical care, which are mediating factors.

Examples

TABLE 4-5 shows how data collected during the community assessment can be presented in a tabular format. It contains examples of information for the five health problems identified in the community assessment of the imaginary Layetteville. Possible required antecedent factors, causal factors, moderating factors, and mediating factors that lead to five different health problems are given. The information that is included in a table such as Table 4-5 and in a causal theory diagram is based on the data collected during the community health assessment, including the scientific literature related to the health problem. These hypothetical details of the five health problems demonstrate how a causal theory diagram might look using the data from this table. These five health problems are carried throughout the text at the subsequent stages of the planning and evaluation cycle.

One example is the health problem of adult immunizations, particularly against influenza. Not surprisingly, we want the rate

TABLE 4-5 Existing Factors, Moderating Factors, Key Causal Factors, Mediating Factors, and Health Outcome and Impact for Five Health Problems in Layetteville and Bowe County

Required Antecedent Factors or Conditions	Moderating Factors or Conditions	Key Causal Factors	Mediating Factors	Health Outcome	Health Impact
Age, existing health conditions, pathogens in environment	Knowledge about adult immunizations, media attention, quality of medical care	Motivation to be vaccinated, fear of the communicable disease, perceived susceptibility	Vaccine supply and distribution, vaccine cost	Immunity level	Rate of preventable hospitalizations
Age, food availability, type of employment, genetics	Knowledge about folic acid, taking prenatal vitamins, genetic counseling	Inadequate intake of folic acid, quality of prenatal care, genetic counseling	Preconception nutritional status, biological processes	Presence of neural tube defect	Rate of congenital anomalies
Psychological development, physical development stage	Media messages, knowledge, family support, availability of birth control	Sexual activity, sexual self-efficacy, partner and peer pressure	Use of birth control methods	Diagnosis of pregnancy	Child abuse rate
Developmental stage, local history of violence, local lack of jobs, state gun laws	Parental supervision, school antiviolence program, community action	Lack of conflict resolution skills, school dropout rate, local gang activity, gun availability	Individual resilience, inadequate policing, quality of emergency care	Death from gunshot wounds	Adolescent death rate due to gunshot wounds
Genetic predisposition, age, race, safe place to exercise	Knowledge about diabetes prevention, family support for self-care	Specific health behaviors (e.g., exercise), quality of medical supervision	Physiological processes	Diagnosis of type 2 diabetes	Morbidity due to chronic illness

of adult immunizations to increase. The rate of immunizations is based on individuals actually receiving the vaccine. What causes an adult older than age 55 to seek immunization is motivation. However, even a motivated individual cannot be vaccinated if no vaccine is available or if the cost is too high. The level of motivation varies depending on the person’s knowledge about adult immunizations, the amount of media attention given to the importance of adult immunizations, and the quality of medical care in terms of having a provider who recommends getting immunized. Ultimately, the need for adult immunization and

the degree of motivation is based on the person’s age, current health condition, and the presence of the pathogens in the environment.

The factors listed in Table 4-5 related to deaths from adult immunization rates, adolescent death rates, and rates of congenital anomalies are shown as causal theory diagrams in FIGURES 4-2, 4-3, and 4-4, respectively. These examples use hypothetical data, but findings from the literature were used to substantiate those data: Hwang and Jaakkola (2003) found an association between exposure to chlorination and birth defects; Calhoun, Dodge, Journal, and Zahnd (2005) used police

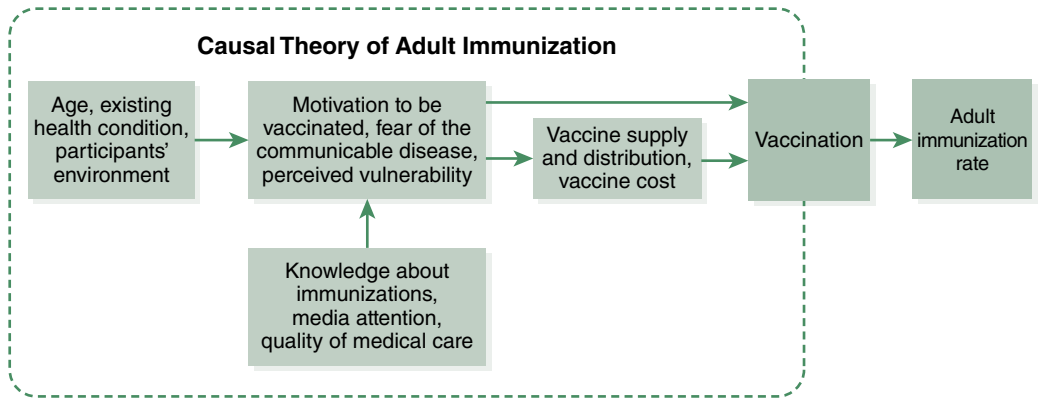


FIGURE 4-2 Diagram of Theory of Causes/Determinants of Receiving Immunizations, as Contributing to Adult Immunization Rates, Using the Layetteville Example

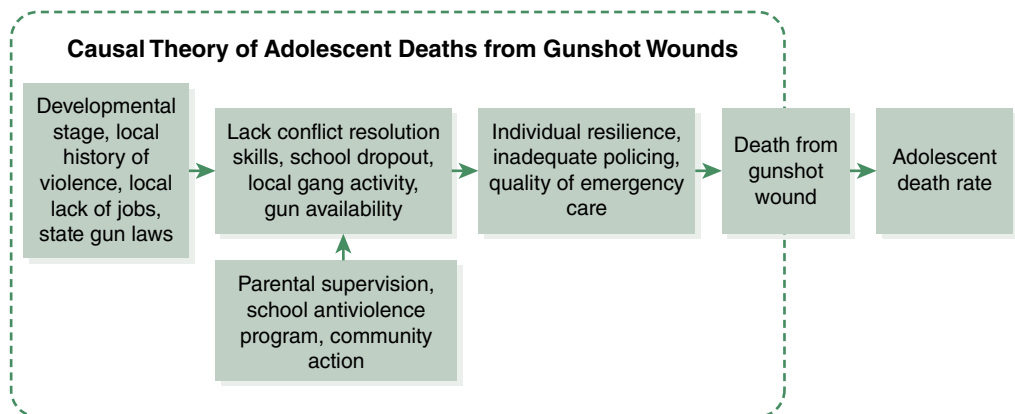


FIGURE 4-3 Diagram of Theory of Causes/Determinants for Deaths from Gunshot Wounds, as Contributing to Adolescent Death Rates, Using the Layetteville Example

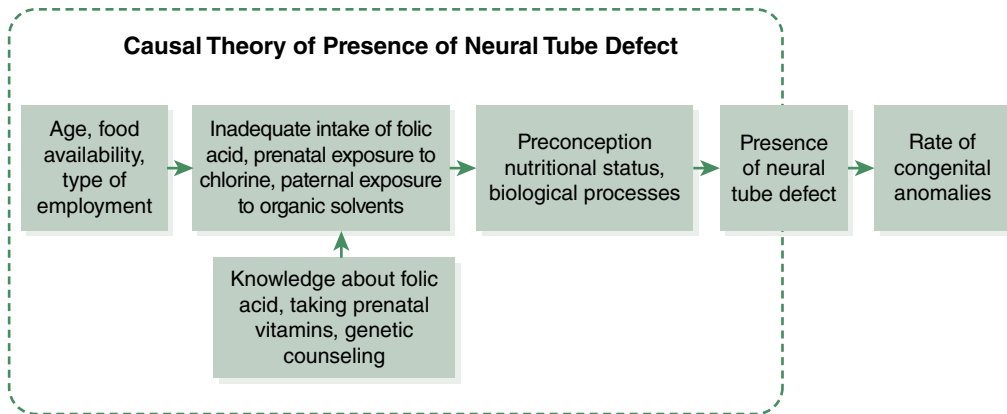


FIGURE 4-4 Diagram of Theory of Causes/Determinants for Neural Tube Defects, as Contributing to Rates of Congenital Anomalies, Using the Bowe County Example

records and gun sales records as part of their assessment. The important consideration here is to gain confidence in pulling the diverse community assessment data together in a coherent, systematic, and scientifically defensible manner.

Writing a Causal Theory of the Health Problem

Community diagnosis, as a diagnosis-type formula, was suggested by Muecke (1984) as one technique for synthesizing needs assessment data into a statement that can be understood by various health disciplines. Since then, the term *community diagnosis* has been used to encompass the data collection as well as the planning. As a formula, it follows the general diagnosis format that specifies who is at risk for what health problem and to what extent those risks are related (Stanhope & Lancaster, 2010). It has also been the basis for graduate courses (Quinn, 1999) and for preparedness (Matsuda & Okada, 2006; Okada et al., 2006). Because these definitions might more aptly describe the community health assessment process, the focus here is on developing a coherent statement that is the equivalent of the diagram of the causal theory of the health problem. The value in writing the causal theory statement is that it complements and extends the value of the diagram as a reference

point for planning the intervention, writing objectives, and planning the evaluation. It also serves as a reference point if the planning and evaluation process begins to drift away from the original health problem and factors identified in the community health assessment.

The causal theory statement adds two elements to the causal theory diagram. The full statement includes both the “who” in reference to the health problem and the relevant health indicators. The elements of the causative theory are used to develop the causal theory statement. The basic format is as follows.

Basic Template

Risk of [health problem] among [population/community] is indicated by [health outcome indicators] and results from [causative factors].

Note that the population or community is the “who” of the community health assessment.

Using the example gun violence as described in the Layetteville community health assessment, the basic causal theory would read as in the following example.

Example

Risk of [death from gunshot wounds] among [adolescents of Layetteville] is indicated in the

[*high rate of admissions for gunshot injuries at the local hospitals and police reports*] and results from [*lack of conflict resolution skills, being a high school dropout, local gang activity, and gun availability*].

The basic template can be modified to incorporate findings about existing factors and mediating factors. The resulting template has more of a public health tone. However, the template does not yet include a way to incorporate assets or strengths of individuals, families, or the community. If we consider that assets and resources have the potential to modify the causative factors' ability to result in the health problem, then including assets identified through the community needs assessment statement is key. Thus, the final full causal theory statement template is as follows.

Causal Theory Statement Template

[*Health problem*] among [*population/community*], indicated in [*health outcome indicators*], is caused by [*causative factors*], but is mediated by [*mediating factors*] given that [*moderating factors*] moderate the causes and that [*required antecedent factors*] exist prior to the causes.

Continuing with the Layetteville community example, the community diagnosis template statement would read like Causal Theory Statement A.

Causal Theory Statement A

Risk of [*death from gunshot wounds*] among [*adolescents of Layetteville*], indicated in the [*high rate of admissions for gunshot injuries at the local hospitals and police reports*], is caused by [*lack of conflict resolution skills, being a high school dropout, local gang activity, and gun availability*] but is mediated by [*individual resilience, adequacy of policing, and quality of emergency medical care*], given that [*community action, parental supervision, and school anti-violence programs*] moderate the causes and that [*adolescent developmental stage, local history of*

violence, lack of job opportunities, and state gun laws] exist prior to the causes.

The community health assessment is likely to identify or uncover numerous health problems or conditions that potentially need to be addressed. For each health problem, a community health statement can be developed. The factors identified through the community assessment may apply differently to the different health problems. A separate community diagnosis statement can be written for each highly prioritized health problem. For example, the birth defects health problem ranked high in priority, so a community diagnosis statement was developed, as shown in Causal Theory Statement B.

Causal Theory Statement B

Risk of [*birth defects*] among [*residents of Bowe County*], indicated by the [*rate of neural tube defects and congenital anomalies*], is caused by [*low folic acid intake, parental exposure to organic solvents, prenatal exposure to chlorine, and the Zika virus*], but is mediated by [*preconception nutritional status and biological processes*], given that [*genetic counseling, use of prenatal vitamins, knowledge about folic acid, and mosquito abatement*] moderate the causes and that [*the mother's age, type of employment, and availability of food high in folic acid*] exist prior to the causes.

A causal theory statement ought to convey information about the health problem so that it stands as a well-articulated base from which to engage in the prioritization and subsequent program designing processes. The statements can be used in a priority setting, as will be seen in the next section of this chapter. Statements about health problems can be compared with regard to the extent to which the required antecedents, causal factors, and moderating and mediating factors are amenable to change as well as the level of seriousness or importance of the health problem. Although a health problem might be considered a high priority initially, data from the community health assessment might in fact potentially lead to a reprioritization of the problem. In short, prioritization and

TABLE 4-6 Relationship of Problem Definition to Program Design and Evaluation

Diagnosis	Problem →	Program →	Evaluation
Risk of:	Health problem or condition	Program goal	Outcome variables
Among:	At-risk population or group, target audience	Recipients	Intervention group
As demonstrated in:	Health indicators	Program objectives	Outcome and impact variables
Resulting from causal factors:	Specific processes, conditions, and factors	Interventions or treatments for the target population	Outcome evaluation
But is mediated by:	Factors that must be present for the health problem to occur	Possible intervention	Possible control variables
Given moderation of the causes by:	Factors that increase or decrease the potency of the causative factors	Possible intervention	Possible control variables
And required existing factors of:	Sociodemographic characteristics and social ecological factors	Program eligibility criteria	Control variables or comparison groups

assessment are often iterative processes rather than a straightforward linear process. The nature of these processes hints at the extent to which community assessors and health planners need to be flexible and act as guides throughout the planning–assessment process.

In summary, the elements contained in the causal theory statement are related to both program design and program evaluation, as shown in **TABLE 4-6**. The causal theory diagram and the corresponding causal theory statement become the basis for developing the program theory with the corresponding logic model. One point is critical to understand at this juncture: The program interventions that will be developed

can be targeted at either reducing *or* enhancing key factors.

► Prioritizing Health Problems

The final problems to be addressed by a program are those selected from among the many health concerns identified through the needs assessment. A highly rational approach to prioritization is presented here because the decision process generally begins as a rational approach. Individuals with inclinations toward other planning

approaches may alter this process accordingly. Nonetheless, health professionals involved in a community health assessment need to have the skills to guide the decision process in a way that does not ignore the data and that results in a plan with the highest overall potential to improve the health status of the community and target community.

Nominal Group Technique

The nominal group technique is not strictly a health planning or prioritizing method but is more typically used in small-group processes and in research. Because it is widely used and can easily be applied to stakeholders who have little experience, it is included in this section as a prioritization approach. The nominal group technique has been used for a wide variety of prioritization needs, ranging from ranking research priorities (Elliott et al., 2016) to environmental changes or barriers for physical activity (Lees et al., 2007; Jahns, McDonald, Wadsworth, Morin, & Liu, 2014), to relative ranking of health interventions under resource constraints (Makundi, Kapiriri, & Norheim, 2007), to developing health programs (Annang, Hannon, Fletcher, Sykes, & Cornish, 2011; Dreer et al., 2013).

This technique involves a round-robin series of voting and narrowing lists based on the results of the voting. In essence, the process begins with the complete, usually long list of health problems identified from the community health assessment. Each member of the planning group is given three votes to be used to select which problems to address. The problems with the most votes are kept, and the problems with the fewest votes are eliminated. Next, the voting is repeated with each member of the planning group having only one vote. The health problem with the most votes becomes the problem to be addressed. The results may or may not be logical, but they often stimulate dialogue and discussion about why the highest priority problem emerged.

For the nominal group process to be successful in selecting and prioritizing health problems, participants must agree before voting that the

results will be honored and used as the basis for moving forward in the program planning. It is also important to give the group adequate background information so that group members can make informed decisions when they cast their votes.

Basic Priority Rating System

A more systematic approach was developed by Hanlon (1973), whose model now has a history of being used or adapted in efforts to formulate a community health plan (New York Department of Health, 2006; Sogoric et al., 2005). An overview of this process provides some initial insight into the depth and breadth of the data that are needed for making decisions about program directions. Hanlon's basic priority rating system (BPRS) is now one of several tools used for establishing priorities and is one of the few tools to highlight an objective approach. Over the years, the adaptability of the Hanlon method can be seen in the small variations made that accommodate community participation in the planning process and major local concerns (Neiger, Thackeray, & Fagen, 2011). The extent to which the BPRS has been adapted in practice underscores the usefulness of knowing the original system (Platonova, Studnicki, Fisher, & Bridger, 2010).

Hanlon's approach to planning public health programs has been codified into the deceptively simple formula known as the BPRS. This method entails prioritizing health problems based on the magnitude of the problem, the severity or importance of the health problem, and the potential effectiveness of interventions. A key part of the process involves assigning values to each of these three factors. The formula is:

$$\text{Basic priority rating} = (A + 2B) \times C$$

where A is the score for the magnitude of the problem, B is the score for the seriousness of the health problem, and C is the score for the potential effectiveness of the intervention.

Unfortunately, the scores assigned to the problem magnitude, seriousness, and intervention effectiveness (TABLE 4-7) can be biased by

TABLE 4-7 Criteria for Rating Problems According to the BPRS

BPRS Factor	A	B				C
	Size	Urgency	Severity	Economic Consequences	Willingness or Involvement of Others	Intervention Effectiveness
Rating scale	1 (small) to 10 (endemic)	1 (not at all) to 10 (extremely urgent)	1 (low) to 10 (high)	1 (low) to 10 (high)	1 (low) to 10 (high)	1 (low) to 10 (high)
Factors to consider in the rating	Stability of incidence or prevalence over time	Rate of spread	Extent to which QALYs and DALYs are affected; virulence of health problem	Healthcare costs; extent to which YLL and YPLL are affected	Political support for addressing the problems; popular awareness of the health problem	Recalcitrance to change; entrenchment of contributing factors

the personal preferences of those involved in the planning process. By going through a group process to arrive at a score for each factor, however, members of the planning group are forced to make explicit the assumptions underlying their assignment of values. This understanding in turn helps establish consensus and consistency within the group.

The first factor to determine is the magnitude of the health problem (A). Magnitude is reflected in expressed need, such as the demand for and utilization of services. The magnitude can also be the burden of disease in terms of the incidence rate (Pooripussarakul, Rewpai-boon, Bishai, Muangchana, & Tanivess, 2016). The normative needs—namely, what health professionals view as being a deviation from a baseline or normally acceptable level—also captures the magnitude. Normative need is reflected in epidemiological measures, such as mortality and morbidity rates, incidence,

prevalence, and relative risk. One difficulty with using mortality rates as the sole criterion for determining the size of a health problem is that mortality data are medical, making them less helpful in planning that focuses on behavioral or social health problems. In addition, disability, pain, and quality of life are just as important considerations as death, as we have seen with regard to QALYs and DALYs. Thus, the size of a health problem and the factors leading to its manifestation ought to be viewed from various angles and incorporate a diversity of measures or indicators.

Not all health problems are equally serious (B), where seriousness encompasses the degree of urgency for addressing the problem, the degree of severity of the health problem, the degree of economic losses possible from the health problem, and the degree to which others can be motivated to become involved. Each of these four elements of seriousness can be rated on a scale of 1 (at the

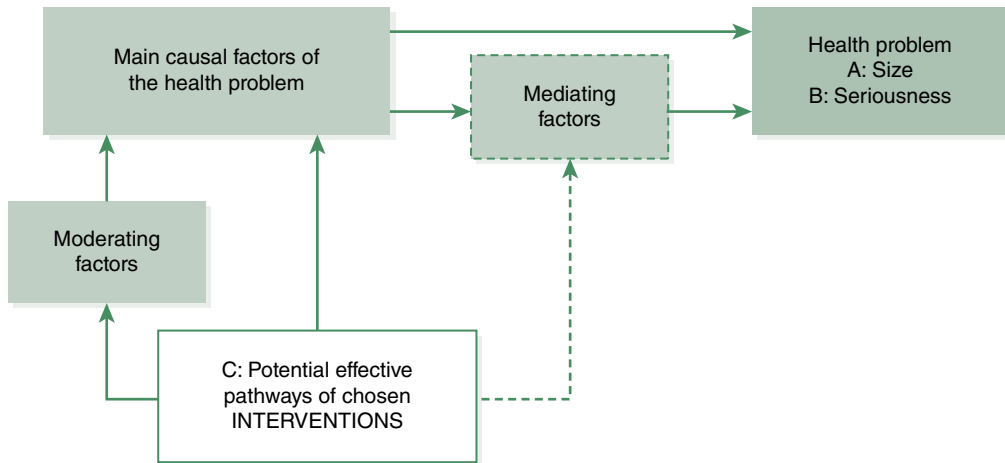


FIGURE 4-5 Theory of Causes/Determinants with Elements of the BPRS Score: Size, Seriousness, and Interventions

lowest end) to 10 (at the highest end). Again, the specific data derived from the community assessment are used to score each element. The severity of a health problem or condition is also related to its virulence.

Seriousness is best determined by examining information from experts, the scientific literature, and input from key stakeholders on the long-term consequences of the health problem. The degree of economic loss focuses on individual loss due to disability and death, but it also might include the societal costs of providing care and the loss of revenue from disabled individuals. Utility measures that capture individuals' preferences for different states of health also play a role, implicitly or explicitly, in determining seriousness.

Intervention effectiveness (C) is the third element in the BPRS. Scoring the effectiveness of the interventions that might be used to address a health problem also utilizes a scale of 1 to 10. Interventions for which considerable and favorable evidence exists would be rated highest, where *favorable* means having a clinically and practically significant effect on the health problem. The choice of an intervention deserves considerable attention, in terms of whether and how it has the potential to affect causal or other

factors (**FIGURE 4-5**). The effective and efficient (lower-cost) interventions can be viewed as contributing to the extent to which it is possible to change the health problem. Naturally, at the point that the prioritizing process occurs, the planning group may not have complete data on intervention effectiveness. In this situation, some data gathering may be necessary, with subsequent rescoreing.

Propriety, Economics, Acceptability, Resources, and Legality (PEARL) Component

Once the calculation of the priorities has been done, the high-priority problems can next be discussed in terms of their propriety, economics, acceptability, resources, and legality (PEARL) characteristics (Vilnius & Dandoy, 1990). *Propriety* refers to whether addressing the health problem is the responsibility of those represented by the planning group. The *economic aspect* relates to the economic feasibility of addressing the problem. *Acceptability* is assessed in terms of the culture's and population's preference for the potential intervention. *Resources* refers to the availability of all types of resources. Naturally,

resources available for addressing a health problem is a concern that affects prioritization. The *legality* element reflects whether there are legal constraints or mandates in addressing the health problem. Each health problem being considered for program planning is evaluated on the five dimensions and scored as either yes (1) or no (0).

Applying the PEARL component to prioritizing among health problems may not be possible until sufficient data about the health problem and characteristics of the community are known. In other words, it may not be possible to apply the PEARL scoring until a substantial amount of information has been collected about the sociopolitical context of the health problem. The need to revisit priorities when new information is uncovered is likely to be frustrating for everyone involved in the priority setting, but it is a reality of program planning.

Prioritizing Based on Importance and Changeability

A simpler approach to understanding the priority of a health problem is to consider only whether an intervention can actually make a change in the health problem and whether the health problem is important or worth addressing. The importance of a health problem can be assessed

through the use of utility measures or through reliance on the *Healthy People 2020* objectives, which indicate whether the health condition is sufficiently important to warrant national attention. The changeability of a health problem or condition is the degree to which any intervention has the potential to alter its course.

When multiple health problems or conditions are being considered as targets for an intervention or health program, each of the health problems can be rated with regard to its degree of importance and changeability. Using these two dimensions, program planners can use the high and low changeability and high and low importance parameters to form four quadrants of a 2×2 matrix into which health problems can be sorted (TABLE 4-8). Health problems classified as having both high changeability and high importance ought to be addressed first. In contrast, health problems in the low changeable, low importance quadrant are either at the bottom of the list or off it entirely.

The above discussion of methods for prioritizing health problems assumes that the planning group and stakeholders will make rational decisions. Against this backdrop of rational planning, other factors will likely come into play, such as personal experiences with a specific health conditions, political preferences, or time

TABLE 4-8 Program Prioritization Based on the Importance and Changeability of the Health Problem

	Highly Important Health Problem	Less Important Health Problem
Highly Changeable Health Problem, More Effective Intervention	High priority for developing a program	Low priority, unless resources are available for developing a program
Less Changeable Health Problem, Less Effective Intervention	High priority, if an innovative program can be developed	No program development is warranted

needed to complete the prioritization process. Another important factor is the existence and popularity of rankings, particularly of the county health status (<http://www.countyhealthrankings.org/>). Published rankings, while intuitive in their interpretation, can be misleading for planning purposes. Rankings provide a relative or comparative status, but they do not take into account the magnitude or severity of the item being ranked nor the changeability of the problem, particularly in the local context. Attention to making a thoughtful, transparent, and explicit selection of health problems helps to engage stakeholders in implementing solutions, and minimizes the possibility of idiosyncratic and inappropriate selections.

► Across the Pyramid

In this chapter, the community health assessment process, problem statement development, and prioritization process have been described as though they constituted a linear set of activities. This is not the reality. These activities are iterative processes that must be adapted to the local situation. Reflecting on whether the data gathered during the assessment process are representative of the levels of the public health pyramid can be helpful in identifying gaps in the assessment and prioritization.

TABLE 4-9 provides a few examples of how prioritization data can be sorted by level of the pyramid. Note that, across all levels of the pyramid, the determination of the effectiveness of proposed interventions ought to come from the scientific literature or existing, rigorously conducted program evaluations. **TABLE 4-10** shows the elements of the causal theory and illustrates how health problem statements incorporate data from across the pyramid.

At the direct services level of the public health pyramid, the data for prioritization come mostly from individual patients through surveys or from provider, clinic, and hospital

archival data. Accordingly, the information that is incorporated into the health problem statement tends to be personal; episodic; and specific to a small, highly homogenous target audience. At the individual level, the health problem statement resembles a very comprehensive medical diagnosis that incorporates sociocultural determinants.

At the enabling services level of the pyramid, the data used for prioritization need to be applicable to the aggregate of concern. Thus, the assessment data are from community sources, such as local news media and groups advocating for the aggregate. Other sources of data may be gathered by unobtrusive means, such as walking through a park to determine its level of safety. A key archival source of data at the enabling level is the list of existing services, which would be informative with regard to the urgency for creating a program that does not exist. The various data collected can be sorted into the elements of the health problems, perhaps as a first step toward determining which data ought to be woven into the final health problem statement.

At the population-based services level of the public health pyramid, data for the prioritization process are more typically drawn from epidemiological data sources, and trend analysis may be necessary to determine the urgency of the health problem. As for the elements of the health problem statement, the factors at the population level would apply to the entire population and therefore are not specific to individuals or aggregates. In addition, the factors affecting the health problem are likely to be environmental (either social or physical).

Last, at the infrastructure level of the pyramid, sources of data for prioritization come from the organization, the relevant legal jurisdiction, and workforce records. At this level of the pyramid (and at the other pyramid levels), relevant characteristics of the infrastructure can be sorted into the elements of the health problem statement. Doing so serves as a nice double check on barriers and facilitators to the proposed program.

TABLE 4-9 Examples of Sources of Data for Prioritizing Health Problems at Each Level of the Public Health Pyramid*

	Level			
	Individual and Direct Services	Community and Enabling Services	Population-Based Services	Infrastructure
Health problem	Nonimmunized adult; infant with neural tube defect; diagnosis of pregnancy; diagnosis of diabetes; death from gunshot wound	Insufficient parenting support services; lack of diabetes management classes	Cultural acceptance of obesity as normal; lack of media messages about safe sex for adolescents and diabetes prevention	Availability of guns; inequities in vaccine distribution; lack of school-based clinics; insufficient inspection of workplaces for teratogens
Data sources for size (A)	Vital records data; clinic or service provider data; hospital discharge data; police records; survey	Census data; waitlists; human services sources of data	Epidemiological data; acute care and outpatient discharge data	Personnel records
Data sources for seriousness (B)	Medical literature on course of the health problem; literature on associated DALYs, YLL, and YHL	Advocacy group pressure; local media on health problem	Statistical trends; Medicaid and Medicare cost data	Regulatory requirements; lawsuits; effects on capacity requirements
Data sources for intervention effectiveness (C)	Scientific literature; professional associations' practice guidelines	Scientific literature; existing health program evaluations	Scientific literature; existing health program evaluations	Scientific literature; existing health program evaluations

* These examples correspond to the five example health problems in Layetteville.

TABLE 4-10 Examples of Required Existing, Causal, and Moderating Factors Across the Pyramid

Factors	Level			
	Individual and Direct Services	Community and Enabling Services	Population-Based Services	Infrastructure
Required existing factors	Genetic predisposition, health beliefs, attitudes, values	Local economic empowerment zones, safe places to exercise and play	Deep cultural practices and beliefs	Legal or policy considerations, organizational mission, workforce capacity
Causal factors	Lifestyle and health practices, individual physiology, quality of medical supervision, exposure to toxins	Road safety features, local gang activity	Exposure to environmental toxins	Resources allocated, capacity
Moderating factors	Lifestyle practices, family, norms, patterns of health services utilization	Culture, accessibility and availability of local health and social services agencies	Income level, educational level	Health workforce competence, needs assessment and planning, information systems
Mediating factors	Quality of emergency medical care	Police presence in community	Vaccine supply and distribution	

► Discussion Questions and Activities

- Which statistical tests would be used to determine statistical significance, and which statistical tests would be used to determine the variance from a population mean? Give a brief description of their key differences.
- Imagine that you are part of a community health assessment and planning group. Your group believes it is important to know the rate of type 2 diabetes in the three census tracts being assessed. However, the only data available are the county statistics. Describe the process by which you would create a synthetic estimate of the rate of type 2 diabetes for whites, blacks, and Hispanics in the three census tracts.
- Using the hypothetical data in Table 4-5, create a causal theory diagram for either morbidity due to chronic illness or for child abuse rates. Which decisions did you make in developing the diagram?

4. Create priority scores for the health problems in Table 4-5 using the BPRS. If you had scored these problems as though you were the state health officer, how might the scores be different?

► Internet Resources

University of North Carolina (UNC)

At UNC, students' community assessment papers are made public at this website: <http://hsl.lib.unc.edu/specialcollections/communitydiagnosispapers>. A perusal of the papers may help generate some ideas of what is involved and what a needs assessment can yield.

County Health Needs Assessments

Most counties in the United States, especially the larger ones, have a website devoted to sharing their most recent community health need assessments and reports. These informative assessments provide insights into not only how a community's needs change but also how the approach to the assessment changes over time. Google "county health needs assessment report" and browse the links.

Statistical Tests

Two websites list freeware that is available to perform the most commonly used statistical tests. Both websites give brief descriptions and links to the downloads: www.healthcarefreeware.com/calc.htm and <http://statpages.org/>.

Community Tool Box

The Community Tool Box (<http://ctb.ku.edu/en>) provides resources and guidance on conducting community assessments with a focus on building health communities. The "Tools" section includes an entire area dedicated to community assessment.

THRIVE: Tool for Health and Resilience In Vulnerable Environments

Here is an example of creatively using an Internet version of the online priority application. In this example, raters can see their own priorities. Look at the "Collaboraton Multiplier": <https://www.preventioninstitute.org/tools>.

Workbook for Prioritization

The National Association of County and City Health Officials (NACCHO) has a wealth of resources that are available to the public, including these materials related to establishing health priorities: www.naccho.org/topics/infrastructure/accreditation/upload/Prioritization-Summaries-and-Examples.pdf.

References

- Anderson, R. (1995). Revisiting the behavioral model and access to medical care: Does it matter? *Journal of Health and Social Behavior*, 36(1), 1–10.
- Anderson, R., & Aday, L. A. (1978). Access to medical care in the U.S.: Realized and potential. *Medical Care*, 16, 533–546.
- Annang, L., Hannon, L., Fletcher, F. E., Sykes, W., & Cornish, D. (2011). Using nominal technique to inform a sexual health program for black youth. *American Journal of Health Behavior*, 35(6), 664–673.
- Baker, S. P., & Haddon, W., Jr. (1974). Reducing injuries and their results: A scientific approach. *Milbank Memorial Fund Quarterly*, 52, 377–389.
- Baker, T. D., & Reinke, W. A. (1988). Epidemiologic base for health planning. In W. A. Reinke (Ed.), *Health planning for effective management* (pp. 117–130). New York, NY: Oxford University Press.
- Calhoun, D., Dodge, A. C., Journel, C. S., & Zahnd, E. (2005). The supply and demand for guns to juveniles: Oakland's gun tracing project. *Journal of Urban Health*, 82, 552–559.
- Cawley, J., Schroeder, M., & Simon, K. I. (2006). How did welfare reform affect the health insurance coverage of women and children? *Health Services Research*, 41, 486–506.
- Cebrecos, A., Diez, J., Gullon, P., Bilal, U., Franco, M., & Escobar, F. (2016). Characterizing physical activity and food urban environments: A GIS-based multicomponent proposal. *International Journal of Health Geographies*, 15, 35. doi:10.1186/s12942-016-0065-5

- Dever, G. E. (1997). *Improving outcomes in public health practice: Strategy and methods*. Gaithersburg MD: Aspen Systems.
- Dever, G. E. (1980). *Community health assessment*. Germantown, MD: Aspen Systems.
- Dever, G. E. (2006). *Managerial epidemiology: Practice, methods and concepts*. Sudbury, MA: Jones and Bartlett Publishers.
- Dreer, L. E., Girkin, C. A., Campbell, L., Wood, A., Gao, L., & Owsley, C. (2013). Glaucoma medication adherence among African-Americans: Program development. *Optom Visual Science, 90*(8), 883–897.
- Elliott, M. J., Staus, S. E., Pannu, N., Ahmed, S. B., Laupacis, A., . . . & Hemmelgarn, B. R. (2016). A randomized controlled trial comparing in-person and Wiki-inspired nominal group techniques for engaging stakeholders in chronic kidney disease research prioritization. *BMC Medical Informatics and Decision Making, 16*, 113. doi:10.1186/s12911-016-0351-y
- Farivar, S. S., Cunningham, W. E., & Hays, R. D. (2007). Correlated physical and mental health summary scores for the SF-36 and SF-12 health survey, V.1. *Health Quality and Life Outcomes, 5*, 54–62.
- Fos, P. J. (2011). *Epidemiology foundations: The science of public health*. San Francisco, CA: Jossey-Bass.
- Friis, R., H. & Sellers, T. A. (2014). *Epidemiology for public health practice* (5th ed). Burlington, MA: Jones & Bartlett Learning
- Green, L., & Kreuter, M. (1991). *Health promotion planning* (2nd ed.). Mountain View, CA: Mayfield.
- Haddon, W., Jr. (1972). A logical framework for categorizing highway safety phenomena and activity. *Journal of Trauma, 12*, 193–207. Cited in Grossman, D. C. (2000). The history of injury control and the epidemiology of child and adolescent injuries. *Future of Children, 10*(1), 23–52.
- Handler, A., Rosenberg, D., Kennelly, J., & Monahan, C. (1998). *Analytic methods in maternal and child health*. Vienna, VA: National Maternal and Child Health Clearinghouse.
- Hanlon, J. J. (1973). Is there a future for local health departments? *Health Services Report, 88*, 898–901.
- Hwang, B. F., & Jaakkola, J. J. (2003). Water chlorination and birth defects: A systematic review and meta-analysis. *Archives of Environmental Health, 58*, 83–91.
- Jahns, L., McDonald, L. R., Wadsworth, A., Morin, C. & Liu, Y. (2014). Barriers and facilitators to being physically active on a rural U.S. Northern Plains American Indian reservation. *International Journal of Environmental Research and Public Health, 11*, 12053–12063. doi:10.3390/ijerph111112053
- Jia, H., Link, M., Holt, J., Mokdad, A. H., Li, L., & Levy, P. S. (2006). Monitoring county-level vaccination coverage during the 2004–2005 influenza season. *American Journal of Preventive Medicine, 31*, 275–280.
- Jung, K., & Valero, J. N. (2016). Assessing the evolutionary structure of homeless network: Social media use, keywords, and influential stakeholders. *Technological Forecasting & Social Change, 110*, 51–60. http://dx.doi.org/10.1016/j.techfore.2015.07.015
- Kaplan, R. M. (1996). Utility assessment for estimating quality-adjusted life years. In F. A. Sloan (Ed.), *Valuing health care: Costs, benefits, and effectiveness of pharmaceuticals and other medical technologies* (pp. 32–60). Cambridge, UK: Cambridge University Press.
- Lees, E., Taylor, W. C., Hepworth, J. T., Feliz, K., Cassells, A., & Tobin, J. N. (2007). Environmental changes to increase physical activity: Perceptions of older urban ethnic minority women. *Journal of Aging and Physical Activity, 15*, 425–438.
- Makundi, E., Kapiriri, L., & Norheim, O. F. (2007). Combining evidence and values in priority setting: Testing the balance sheet method in a low-income country. *BMC Health Services Research, 7*, 152.
- Matsuda, Y., & Okada, N. (2006). Community diagnosis for sustainable disaster preparedness. *Journal of Natural Disaster Science, 28*, 25–33.
- Muecke, M. (1984). Community health diagnosis in nursing. *Public Health Nursing, 1*, 23–35.
- Murry, C. J. L. (2013). The state of US health, 1990–2010: Burden of diseases, injuries, and risk factors. *Journal of the American Medical Association (JAMA), 310*, 591–608. doi:10.1001/jama.2013.13805
- Neiger, B. L., Thackeray, R., & Fagen, M. C. (2011). Basic Priority Rating Model 2.0: Current applications for priority setting in health promotion practice. *Health Promotion Practice, 12*(2), 166–171.
- New York Department of Health. (2006). *Building on community health assessment: Workshop agenda*. Retrieved September 9, 2012, from www.health.state.ny.us/statistics/chac/agenda2_bcha.htm
- Okada, N., Yokomatsu, M., Suzuki, Y., Hagihara, Y., Tatano, H., & Michinori, H. (2006). Urban diagnosis as methodology of integrated disaster risk management. *Annals of Disease Prevention Research, 49*, 1–7.
- Paronen, O., & Oje, P. (1998). How to understand a community: Community assessment for the promotion of health-related physical activity. *Patient Education and Counseling, 33*(suppl), S25–S28.
- Platonova, E. A., Studnicki, J., Fisher, J. W., & Bridger, C. (2010). Local health department priority setting: An exploratory study. *Journal of Public Health Management & Practice, 16*(2), 140–147.
- Pooripussarakul, S., Rewpaiboon, A., Bishai, D., Muangchana, C., & Tanivess, S. (2016). What criteria do decision makers in Thailand use to set priorities for vaccine introduction? *BMC Public Health, 16*, 684. doi:10.1186/s12889-016-3382-5
- Preedy, V. R., & Watson, R. R. (Eds.) (2009). *Quality of life measures*. New York, NY: Springer.

- Quinn, S. T. (1999). Teaching community diagnosis: Integrating community experience with meeting graduate standards for health education. *Health Education Research, 14*, 685–696.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–400.
- Ramanadhan, S., Mendez, S. R., Roa, M., & Viswanath, K. (2013). Social media use by community-based organizations conducting health promotion: A content analysis. *BMC Public Health, 13*, 1120. Accessed October 13, 2016, from <http://www.biomedcentral.com/1471-2458/13/1129>
- Sogoric, S., Rukavina, T. V., Brborovic, O., Vlahugic, A., Zganec, N., & Oreskovic, S. (2005). Counties selecting public health priorities: A “bottom up” approach (Croatian experience). *Collegium Antropologicum, 29*, 111–119.
- Stanhope, M., & Lancaster, J. (2010). *Foundations of nursing in the community: Community-oriented practice*. St. Louis, MO: Mosby/Elsevier.
- VanEenwyk, J., & Macdonald, S. C. (2012). *Guidelines for working with small numbers*. Washington State Department of Health. Accessed November 29, 2016, from www.doh.wa.gov/portals/11/documents/5500/smallnumbers.pdf
- Vilnius D., & Dandoy S. (1990). A Priority Rating System for Public Health Programs. *Public Health Reports, 105*(5):463–470.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (2000). *Unobtrusive methods* (rev. ed.). Thousand Oaks, CA: Sage.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Non-reactive measures in the social sciences* (2nd ed.). Boston, MA: Houghton Mifflin.
- Yu, H., Meng, Y., Mendez-Luck, C. A., Jhawr, M., & Wallace, S. P. (2007). Small-area estimation of health insurance coverage for California legislative districts. *American Journal of Public Health, 97*, 731–737.

SECTION III

Health Program Development And Planning

CHAPTER 5

Program Theory and Interventions Revealed

After developing statements about health problems that have been ranked as a high priority, the next steps in health program planning involve a more intellectual and creative effort to articulate an explanation of what caused the problem. This is a critical step toward identifying which intervention or group of interventions will be most effective in addressing the health problem. Wild guesses, past experience, and personal preferences might be used as the basis for decision making, but a more rational approach is to identify existing scientific knowledge and theories that can be used to develop a program theory.

A theory is a description of how something works. It is a set of statements or hypotheses about what will happen and therefore contains statements about the relationships among the variables. We use working theories in everyday life, usually in the form of working hypotheses, such as “If I ask the children to clean their rooms, they are not likely to do it.” We also use theories based in science. For example, based on theories of thermodynamics and heat conduction, we can predict how long the turkey needs to roast.

With regard to planning a health program, a primary consideration is to specify what is to be

explained or predicted with a theory. The health problem is what needs to be explained from a programmatic perspective. To explain how to change or affect the health problem, a theory must contain relevant variables, or factors, and must indicate the direction of the interactions among those variables related to the health problem. Identifying the relevant antecedent, contributing, and determinant factors of the health problem gives planners the foundation for developing a working theory of how the programmatic interventions will lead to the desired health outcome. A difficult part of this task is to identify where a health programmatic intervention can have an effect on those factors. As more details and more factors are included in the explanation of the health problem and beliefs about how the programmatic interventions will work, the theory becomes increasingly more complex.

The theory development phase of program planning requires thinking rather than doing, so it often receives less attention than is needed to develop an effective health program fully. However, using a systematic approach to develop a program theory and to engage stakeholders in the development of the theory has big and

long-term payoffs that outweigh any delay or costs associated with developing the theory.

► Program Theory

A sound basis for developing the health program and for guiding the program evaluation is the use of a program theory. Rossi, Freeman, and Lipsey (2004) acknowledged that the need for a program theory has long been recognized by evaluators in the social sciences, and many have advocated for program theory as useful in public health program development (Potvin, Gendron, Bilodeau, & Chabot, 2005). *Program theory* is a conceptual plan, with some details about what the program is and how it is expected to work. The comprehensive overview of how the program is to work has various names, including logic model, causal model, outcome line, program model, and action theory. These names all refer to a conceptual plan of how the program will work. Whether one is developing a new health program or designing an evaluation for an existing health program, understanding and articulating the program theory is essential.

The use of the words *impact* and *outcome* is inconsistent across the literature, in practice, and in government. It is prudent to look beyond the words themselves and ask for definitions. In this text, *outcome* refers to the immediate effects resulting from an intervention, whereas *impact* refers to the long-term or cumulative effects attributable in part to the programmatic interventions. The term *effect* generically refers to changes or consequences of an intervention, regardless of whether the changes are immediate, proximal outcomes, or longer term distal impacts.

There are two main components of program theory, as shown in the top half of **FIGURE 5-1**. The theory about resources and actions is called the *process theory*, and the theory about interventions and outcomes is called the *effect theory*. The concept of program theory is used throughout this text rather than the more widely used term *logic model*. The key difference is that a full program theory, compared to a logic model, contains a far more explicit explanation of the relationship of the factors related to the health problem with the interventions. These relationships are the effect theory. Similarly, the process theory offers a more explicit and detailed description of the resources

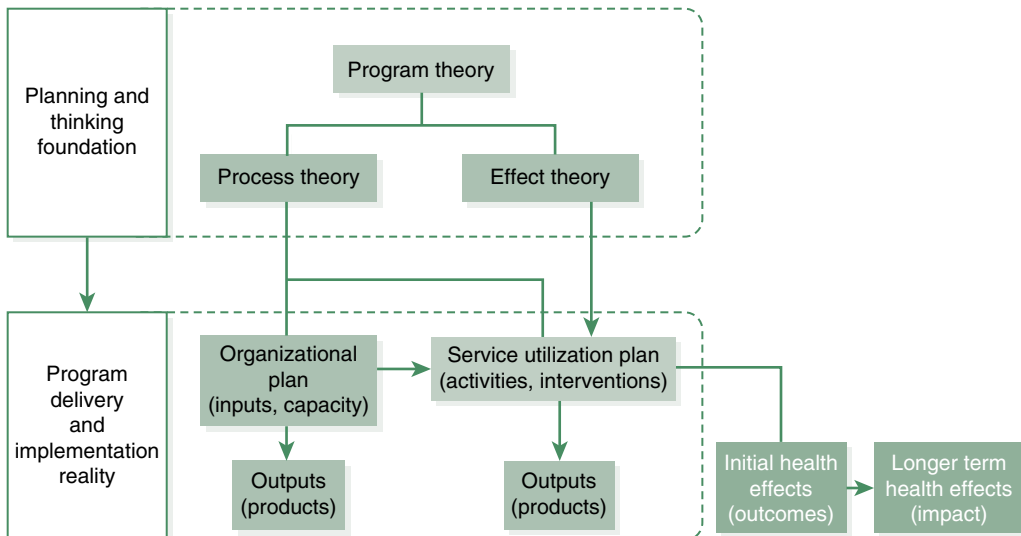


FIGURE 5-1 Model of Program Theory

Modified from Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.

used than is normally found in a logic model. The major similarity is that both a logic model and the program theory provide road maps to creating a successful program. The development of a program theory and its components leads to a stronger program and a more convincing argument for the program's existence.

Process Theory

The process theory includes three components: the organizational plan, the service utilization plan, and specifications of their outputs (Rossi et al., 2004). Process theory includes inputs, which are part of the organizational plan; activities, which are part of the service utilization plan; and outputs, which are by-products of the organizational and service utilization plans.

The *organizational plan*, according to Rossi et al. (2004), encompasses the nature of the resources needed to implement and sustain the program. As such, it includes specifications about personnel; the organization of resources to be used in the program; and elements of capacity, such as infrastructure, information technology, fiscal resources, and personnel. It covers all the behind-the-scenes work needed to provide a program. The organizational plan implicitly contains if-then statements.

For example, if program staff members are adequately supported with regard to supplies and managerial support, then these staff members will deliver the interventions as planned. These if-then statements are useful not only for checking the logic behind requesting specific resources but also for guiding the portion of the evaluation plan that focuses on the processes behind the delivery of the health program.

The *service utilization plan*, according to Rossi et al. (2004), specifies how to reach the intended audience and deliver the programmatic interventions and services to that audience. It constitutes the nuts and bolts of providing the program and of implementing the program plan. The service utilization plan includes specifics about social marketing of the program, accessibility and availability of the program, screening

procedures, and other logistics of providing the program. Development of the service plan ought to reflect cultural sensitivity and appropriateness of the services and intervention given the target audience.

Within the context of planning a program, the organizational plan needs to be in place before the program can begin. Developing both the organizational plan and the service utilization plan draws on the results of the organizational and community health assessments, particularly with regard to incorporating existing resources into the plans and addressing structural issues that can affect the delivery of the program. The organizational plan is influenced by the service utilization plan to the extent that the planned intervention must be adequately supported by the resources outlined in the organizational plan. As a consequence, the development of the organizational and service utilization plans is an iterative process, with considerable back-and-forth adjustments as each element is more fully explicated. Likewise, the service utilization plan evolves as the effect theory is revised, which then leads to adjustments in the organizational plan. Thus, the process theory elements are continually adjusted throughout this phase of planning for the program. Although the time it takes to make adjustments and revisions may be frustrating, it is much easier to make the adjustments at this stage of planning than it is to do so after the program has begun.

Effect Theory

Effect theory consists of the explanations of how the programmatic interventions will affect the causal factors and moderating or mediating factors of the health problem and describes the relationship between the programmatic interventions and the desired immediate and long-term outcomes for program participants. Three sets of relationships, or theories, constitute the effect theory (FIGURE 5-2): the theory of causal/determinant factors, the theory of intervention mechanisms, and the theory of outcome to impact. Depending on the health problem, it

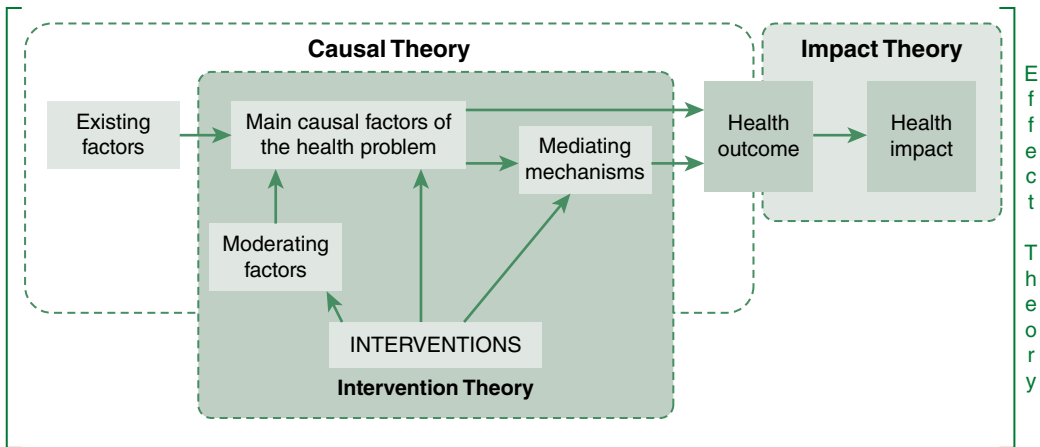


FIGURE 5-2 The Effect Theory Showing the Causal Theory Using Community Diagnosis Elements

can be useful to develop each of these theories. Often these theories are implicitly stated and understood by health professionals and program staff members. By explicitly stating and discussing these theories, however, program planners can refine programmatic interventions, thereby increasing the likelihood of program success.

This set of three theories and the associated informally stated hypotheses constitute the effect theory portion of the program theory. The term *effect theory* makes it clear that this part of the program theory deals with both outcomes and impacts. Generating each of the theories that constitute the effect theory may seem complicated. Program experts agree on the complexity of constructing an effect theory as well as its central role in program evaluation (Rossi et al., 2004).

► Interventions

Interventions are actions that are done intentionally to have a direct effect on persons or populations with the health problem. In other words, interventions are the verbs that tell what is being done to make a change in program recipients. Using this definition allows for the inclusion of a broad range of actions, such as medical treatments, pharmacological treatments,

and education, as well as psychological strategies and policy formulation. Such a broad definition also allows for the inclusion of strategies not typically considered treatments, such as providing transportation (an enabling service) or community development (an infrastructure-level intervention). Clearly identifying and labeling the interventions as such make developing the intervention and outcome theories easier, facilitate developing outcome objectives, and help distinguish outcome and impact objectives from process objectives.

Sometimes interventions are included as or couched in terms of “activities.” Doing so makes the intervention actions indistinguishable from the myriad of other activities done as part of the organizational or service utilization plans; the latter activities are supportive of the interventions but are not actions that will make a difference on the health problem. Interventions are the heart of all health programs. A clear understanding and statement of the role of interventions is made in the intervention theory.

Finding and Identifying Interventions

Selecting and then articulating the chosen interventions are cornerstone activities of health

program planning. It is important when planning a health program to draw on existing knowledge in multiple disciplines. A literature review, for example, can generate ideas and information with regard to existing theories that have been used to explain what leads to the health problem, as well as explanations of why some interventions have been effective and others have not.

The use of existing theories can expedite the development of the effect theory and lend it credibility. Heaney and van Ryn (1996) provided a nice example of this phenomenon. In their case, the health problem was worksite stress. These authors wanted to develop a health program to reduce worksite stress but were concerned that existing programs had been designed for a target audience of middle-class employees within the cultural majority. Recognizing this fact, Heaney and van Ryn sought to improve the effectiveness of worksite stress-reduction programs for employees of low status or of a cultural minority. Their premise was that the potential exists for different subgroups to vary in both their participation in and benefit derived from a program.

Heaney and van Ryn (1996) began by reviewing the literature on stress and coping. From this literature, they constructed a theoretical model of stress and coping, and identified the major variables along with the direction of the interaction among those variables. They also reviewed the literature on the content of worksite stress-reduction programs and the sociological literature on status, class, culture, and stress. From their literature reviews, they were able to identify program interventions that might potentially alter specific variables in the stress and coping model. This information became part of their effect theory for the worksite stress-reduction program for low-status minority workers.

Unfortunately, for many health problems, widely accepted theories are not available to guide the development of an effect theory or the selection of interventions. However, health program planners and the planning team have options for determining how to proceed.

Types of Interventions

A simple starting point for thinking about types of interventions is to consider the levels of prevention. In the most common typology in public health, prevention activities are classified into three levels: primary, secondary, and tertiary. *Primary prevention* includes those activities that are done to prevent a disease or illness from beginning. Adequate exercise, good nutrition, immunizations, and wearing seat belts are examples of primary prevention. *Secondary prevention* involves screening for undiagnosed problems so that a disease can be treated before it manifests itself. Blood pressure screenings at health fairs, fecal immunochemical tests for colon cancer, testing for anemia, and cholesterol tests are all secondary prevention activities. *Tertiary prevention* involves activities to limit the extent of an existing disease. For example, it includes taking blood pressure medications, receiving physical rehabilitation after an injury, and taking stress management classes for individuals with cardiac problems. The three levels of prevention provide a starting point, but they are not sufficiently detailed to provide guidance in the development of programmatic interventions.

Another approach to thinking about types of interventions is to consult one of the various classification schemes of interventions that have been developed across the health disciplines. In medicine, the Current Procedural Terminology (CPT®) codes (American Medical Association, 2016) enumerate the various procedures that physicians perform. Excluding diagnostic procedures, all other procedures can be thought of as interventions in the sense that they are intended to affect the health of an individual. In nursing, the equally detailed Nursing Intervention Classification (Bulechek, Butcher, Dochterman, & Wagner, 2013) or the Omaha System (Monsen & Newsom, 2011) may be used to categorize interventions. Given the highly clinical nature of these intervention classifications, they would be helpful only in the development

of health programs at the individual level. One typology of interventions, the Public Health Intervention Wheel, shows a core set of interventions as applied at the individual, community, population, and systems levels (Olson-Keller, Strochschein, Lia-Hoagberg & Schaffer, 2004). Nevertheless, their use with electronic medical or health records is a big advantage in subsequent program monitoring.

A more global intervention typology is needed to identify interventions across the public health pyramid levels. One such typology, developed by Grobe and Hughes (1993), had seven categories of interventions; an eighth category was added by Issel (1997) when studying case management of pregnant women. This typology, by providing an encompassing perspective, can aid in identifying which activities are programmatic interventions. In a meta-analysis of 131 studies of community-based interventions, most frequently used interventions were education (80% of the studies), advice (54%), social support (44%), and skill development training (39%) (O'Mara-Eves et al., 2015).

Each of the eight types of interventions exists at the direct services, enabling services, and population levels of the public health pyramid (TABLE 5-1). The typology also accommodates both secondary and tertiary prevention because these are activities of health professionals undertaken with the intent of having an effect on the health of the program participant. Primary prevention is not included in the typology because providers cannot “do primary prevention” for or to the participant. Rather, primary prevention is a rubric for a variety of interventions: Individuals receive education about primary prevention, are encouraged to engage in primary prevention behaviors, and might be monitored for the extent to which they practice primary prevention behaviors. This is one example of how such a typology of interventions forces program planners to be specific about the actions (as reflected in verbs in the written plan) that are undertaken to affect the health condition or situation of the target audience.

Specifying Intervention Administration and Dosage

Many health program interventions differ from medical interventions because they are thought of in more general terms, such as “hand out informational flyers” or “provide emotional support.” Nonetheless, health program interventions also need to be thought of in terms of dosage, route of administration, and site of administration.

Dosage refers to the amount and strength of the intervention, with a minimum dosage presumed to have an effect. We normally think of dosage in terms of a medication regimen or an exercise program. However, each intervention strategy that is included in a health program needs to be developed and tailored with regard to the dosage of the intervention exposure for the program participants. For example, Morone, Greco, and Weiner (2008) provided a stress-reduction mindfulness program and described the dosage in terms of practicing the meditation three times per week, for a minimum of 50 minutes each time, over an 8-week period. Specifying the dosage is important for achieving the optimal program for the intended audience; it also provides the information needed to develop the process theory adequately and appropriately. Once the dosage is specified, that information is incorporated into the service utilization plan and is used to modify the organizational plan to ensure that adequate resources have been allocated.

Dosage consists of five elements: frequency, duration, strength, route of administration, and administration credibility. The first four of these are fairly straightforward. *Frequency* is how often the intervention is received, such as hourly, daily, weekly, monthly. *Duration* specifies over what time period the intervention is delivered, such as one session, 8 weeks of classes, or 6 months of exposure. O'Mara-Eves and colleagues (2015) found that the duration of the intervention associated with an effect varied by health outcome studies. This suggests the need to tailor the duration to the specific desired outcome. *Strength of the*

TABLE 5-1 Examples of Interventions by Type and Level of the Public Health Pyramid

Intervention Type	Direct Services Level	Enabling Services Level	Population Level	Infrastructure Level
Treating	Medical or dental procedures, medications, physical manipulations, tertiary prevention	Respite care, exercise classes or groups	Water treatment and fluoridation, mass immunizations	Codification of new treatment modalities
Assessing	Determination of needs and preferences by asking individuals, secondary prevention	Determination of needs and preferences by needs assessment	Use of epidemiological data to identify trends and rates of illnesses and conditions	
Coordinating	Care coordination, client advocacy, referral, linking to services	Case coordination, local provider networks and collaborations	Systems integration, records and data sharing, disaster response, planning	Interorganizational and intersectoral collaboration
Monitoring	Reassessment, follow-up	Local trends and news reports	Trends analysis, epidemiological surveillance	
Educating	Skills building, information giving	GED programs, job training programs	Media campaigns	Workforce training
Counseling	Psychotherapy, emotional support, marital counseling, cognitive behavioral therapy	Group counseling, family counseling, grief counseling for groups	New alerts and advice	Codification of new trauma diagnoses
Coaching and nudging	Role modeling, motivational interviewing, empowerment, encouragement, stress management	Community development	Policy formation	Urban redesign, built environment changes
Giving tangibles	Giving vouchers for food or clothing	Medical supplies loan programs	Income supplements, insurance supplements	Establishing food banks, tax incentives

intervention refers to the powerfulness or intensity. For example, a smoking-cessation mass-media campaign is less intense to receive than smoking cessation counseling by an individual's primary care physician. In an exercise program to improve cardiovascular fitness, stretching has less strength than Zumba even if both are done at the same frequency and duration. A policy intended to prevent gun violence is stronger if there are more and more severe consequences, such as jail time, than if it results in citations and no jail time. *Route of administration* is the mechanism by which the intervention is delivered or the medium used to deliver the intervention, whether interpersonal communication, public mass media, educational brochures, or injection. Thus, dosage of health program interventions might include the number of hours of education, days of respite, micrograms of fluoride, number of home visits, or number of weeks of counseling, all along with the amount of the intervention provided (e.g., six education topics versus one topic) was provided and at what level of intensity (e.g., leaving a brochure versus providing detailed feedback on performing a task).

Administration credibility refers to the perceived degree to which the person or agency providing the health program is knowledgeable and believable. In other words, it involves whether the intervention is provided by a health professional of a particular discipline, a lay health worker, or a paraprofessional. For some health problems, the cultural values attached to a physician may be a key factor in the effectiveness of the intervention, whereas for other health problems and programs, a community member will have more credibility. Thus, among facets of dosage, administration credibility is particularly relevant for health programs.

For many health problems, research reported in the literature or official documents can provide information on which doses are needed to be effective. For example, Martin, Church, Thompson, Earnest, and Blair (2009) found that the amount of exercise per week among postmenopausal women was related to the amount of change in physical and mental quality of life. The dosage of exercise included

the number of kilocalories expended per kilogram of body weight (strength), and the number of sessions per week (frequency) over 6 months (duration) administered by trained professionals. In another example, Guttentag and colleagues (2014) developed an early parenting intervention at a high and a low intensity. Based on the findings from a comparison of infant and maternal outcomes in the two groups, they hypothesized the need to extend the high-intensity intervention in order to maintain the effects gained. Such findings highlight the need to be very specific in the development of not only the intervention but also the target audience and the format elements of the service utilization plan.

Interventions and Program Components

One key challenge in selecting an intervention strategy is deciding whether a single intervention is warranted or whether a package of interventions would be more effective in addressing the health problem. A *program component* comprises an intervention or set of interventions, with the corresponding organizational plan. Thus, if a health program includes multiple interventions, each addressing one of several causes of the health problem or one of several moderating or mediating factors for the health problem, and these interventions are grouped in some way that makes sense for either effectiveness or efficiency reasons, then the program has multiple program components.

Using program components is appropriate if, to address the health problem, changes must occur across levels, such as at both the family and the community levels. Levels are nested within other levels, and each can be the focus of the program. It is extremely difficult to develop a single intervention that can affect all or most of the causes and moderating or mediating factors for a health problem at multiple levels. Instead, program components are typically needed. For example, if individuals as well as the community as a whole in which those individuals live are

targets for the intervention, then interventions tailored to both individuals and communities will be needed. If both individual behavior and actions of the gun industry are targeted to address the problem of gunshot deaths, then different interventions (program components) are needed.

Another reason to include multiple program components is to address micro and macro health problems. As early as 1982, Blum suggested that some health problems or risks require individual behavioral changes, whereas others require group behavioral change. From a public health perspective, an individual behavioral change needed to protect against a health risk is called *active protection*; in contrast, protection that does not require individuals to make a behavioral change but is instituted through policy, laws, or some other means that does not involve the individual is called *passive protection*. Passive protection often occurs at a macro level because it encompasses more than a small group of individuals. However, macro-level changes can also involve active protection, such as the immunization of all infants and vulnerable adults. Immunization involves individual healthcare-seeking behavior but is intended to have a population effect. In contrast, fluoridation of the water supply and reduction of factory pollutant emissions as health programs are both intended to provide passive protection of a population. The distinctions between micro and macro programs, as well as between active and passive protection, may be important in developing the interventions and the effect theory. If the health program is intended to be community-based or community-focused, then it will likely include components at the micro level as well as at the macro level.

Of course, it is important to consider the package of interventions that the recipients actually receive. For example, Harris (2007) used dance and movement therapy as an intervention with African adolescents who were former child soldiers and survivors of torture. The group cohesion that developed during this program was important to the success (i.e., the effectiveness) of the intervention. Similarly, Lipman et al. (2007) identified group cohesion as being critical for

program outcome. These examples highlight the synergistic effects of interventions that can occur when they are provided in a group context, as well as the delivery of a psychological intervention that may or may not have been planned. Understanding such interactions and identifying the presence of implicit interventions is critical to later evaluations of what made the difference in health outcomes.

Some interventions are packaged with mnemonics to assist practitioners with remembering the set of interventions. For example, the “five A’s” consist of assess, advise, agree, assist, and arrange. Fisher et al. (2005) suggested that these interventions may be helpful in programs for diabetes self-management, in addition to drawing attention to the resources and support needed for successful self-management. Alternatively, for programs addressing diabetes and other chronic illnesses, standards have been developed by national associations that specify recommended interventions. Use of national standards is encouraged given that national standards tend to be evidence based, updated regularly, and used as the community standard of practice.

Because each program component will have a slightly different effect, acknowledging the individual components is important in subsequent evaluation plans. The intervention and outcome theories will vary slightly for each program component and for each of the different units of intervention of the program.

Characteristics of Good Interventions

The final choice of an intervention or a package of interventions can be evaluated against a set of criteria for useful interventions. Having a list of criteria for good interventions is not new (Blum, 1982) but is helpful.

Evidence-Based

Choosing an intervention that has been studied for its effectiveness and found to achieve the intended effect enhances the program’s likelihood

of success. The increased awareness of the need to have an evidence-based practice has resulted in an increase in the number of meta-analyses and literature syntheses that provide a summary of the effectiveness of interventions for a specific health condition or problem. Some reviews provide information on which interventions are more effective for a specific health problem (Cooke, 2010; Lundgren & Amin, 2014), other reviews provide information on the dosage characteristics of effective programs (Virués-Ortega, 2010; Travagin, Margola & Revenson, 2015), yet other reviews assess the effectiveness of a known intervention on alternative outcomes (De-Regil, Peña-Rosas, Fernández-Gaxiola, & Ryaco-Solon, 2015).

In choosing an intervention based on scientific evidence for its effectiveness, program planners sometimes face the question, What constitutes “evidence”? The array of possibilities ranges from meta-analyses of existing studies to a single randomized clinical trial, qualitative reports, or practice guidelines. The other challenge when selecting an evidence-based intervention is dealing with equivocal findings. Many meta-analysis studies of interventions conclude few of the interventions had been inadequately described, making comparisons unreliable. Thus, many authors are reluctant to recommend one intervention over others. This ambiguity over the relative effectiveness of interventions may apply to many health problems.

Tailored to the Intended Population

A good intervention is tailored to the characteristics of the intended population. Tailoring the intervention encompasses adapting the program for cultural sensitivity, linguistic appropriateness, group similarity, cultural beliefs, and ethnic values. It can occur either through a modification of the intervention to fit the intended audience or through screening members of the intended audience for eligibility based on an important characteristic. Either approach achieves the goal of having an intervention that can be readily accepted by the program recipients.

Even widely accepted interventions may need tailoring. For example, Mapuana, Antonio, and Chung-do (2015), in a review of positive youth development programs for Native Americans, found that culturally grounded programs had higher completion rates, leading to higher dosage and thus more favorable outcomes. At the individual level, Resnicow and colleagues (2009) found that tailoring breast cancer prevention messages to the ethnic identity of African Americans was associated with a greater increase in their fruit and vegetable intake compared to the control group. However, tailoring interventions—and especially public health prevention messages—can be very difficult, as Perchmann and Reibing (2006) discovered when comparing seven different antismoking messages.

Causes Health Gains

A third criterion is that health gains must result from the intervention. That is, the problem must be able to be changed with the available knowledge of how to change it. This criterion acknowledges that some interventions may have unintended consequences or side effects. Awareness of potential harmful effects of public health interventions has led to the development of a typology of harms and a call for articulating possible mechanisms of harm when developing interventions (Bonell, Jamal, Melendez-Torres, & Cummins, 2014). Other programs are simply ineffective, such as the Drug Abuse Resistance Education (DARE) program, which has been widely adopted but is ineffective (Des Jarlais et al., 2006; Pan & Bai, 2010).

This criterion also speaks to an advantage of fully articulating the effect theory. A common tendency among health professionals and program planners is to jump to a favorite solution, albeit one that may not necessarily be a good match for addressing the health problem. One technique that helps avoid this tendency is to specify the mechanisms and processes that would result in the health gains. In some scenarios, interventions could be useful and effective with regard to one type of outcome, but they may not lead to the

outcome or impact of interest. For example, health education about family planning methods may be effective in reducing the birth rate in a target audience, but it may not be effective in reducing rates of sexually transmitted infections. Again, having done the work of developing the effect theory helps program planners be certain that the intervention will lead specifically to the desired health gains.

In addition, the program planners need to have the requisite expertise for designing the intervention and activities so that those activities will actually affect the health problem. As was discussed earlier in terms of prioritizing the health problems, the changeability of a health problem is considered to be one aspect of its importance. In terms of interventions, a more technologically feasible intervention ought to result in a more changeable health problem.

Manipulability

The fourth criterion is that the intervention must be manipulable (Rossi et al., 2004). *Manipulability* refers to the ability of the program planners and program staff members to adjust the intervention to the specific needs of the participants. A major element of manipulability is dosage, as discussed earlier in this chapter. If the dosage of the intervention can be tailored to the intended audience, then the intervention meets the manipulability criterion. Effective and efficient interventions are customized to some extent to account for the variations among potential participants.

Related to manipulability is the ability to achieve synergy by taking into account other programmatic interventions. For example, Thomas, Sayers, Godon, and Reilly (2009) described a set of community-level interventions to increase physical activity. The program was designed to mutually reinforce the effects of the social marketing and local policy changes. The approach of intentionally developing a program intervention to maximize the effects of program components being delivered to a community is essential as communities become the focus of

complex health promotion programs intended to address social and environmental determinants of health.

Another aspect of manipulability is the notion that the intervention ought to be designed to address moderators as well as determinants of the health problem. This means that the intervention needs to have sufficient strength to overcome factors indirectly affecting the health problem. For example, a school-based violence prevention program has an effect not only on reducing violence in general but also notably among youth who had experienced maltreatment (Crooks, Scott, Ellis, & Wolfe, 2011). The program “buffered” the youth from the consequences of maltreatment. Using the psychological theory of buffering guided the inclusion of specific elements in the program. In other words, existing theories can be helpful in manipulating the intervention so that it is sufficiently strong.

Technologically and Logistically Feasible

The feasibility of an intervention needs to be considered from the point of view of whether it is technologically realistic and logistically doable within the context in which the intervention will be provided. These aspects of an intervention could be determined through a pilot study in which the intervention is provided on a small scale and on a trial basis. For example, Abroms, Hershcovitz, Boal, & Levine (2015) conducted a feasibility study of a text-messaging program to promote smoking cessation in a new population, Israelis. Ensuring involvement of the stakeholders—and particularly those likely to be providing the intervention—in the planning can provide insights into the feasibility of providing the intervention within an everyday context.

Another aspect of feasibility considers the technology to be used as part of the intervention. In some settings or situations, the availability or acceptability of technology is minimal, limiting the nature of interventions. For example, use of mammography for early detection of breast cancer

would not be possible in undeveloped nations, but it also might not be feasible in some remote and impoverished regions in the United States.

Reasonable Cost

The sixth criterion is that the cost of the intervention must be reasonable rather than prohibitive. The cost of the intervention depends on many factors, such as the extent to which the health behavior or problem is resistant to change, the duration of the program, and the number of program components.

Politically Feasible

The seventh criterion of a good intervention is that it be politically feasible. Not all interventions are equally acceptable to the intended audience, to funding agencies, or to other stakeholders. During the assessment phase, program planners should have determined the preferences and willingness of various stakeholders to endorse different types of interventions. Interventions need to be culturally appropriate and sensitive as a first step toward being politically feasible. Various strategies, such as conducting focus groups and pretesting an intervention, can be used to design culturally sensitive and competent health program interventions for use with ethnically or racially distinct target populations.

A corollary to the political feasibility criterion is that meeting this criterion helps the program planner, as well as the program, to survive. Proposing interventions that are not politically feasible can result in the planner being used as a scapegoat and blamed for a “bad” intervention. Worse yet, politically sensitive programs run the risk of not being funded, which will reflect poorly on the qualifications of the program planner.

Addresses Societal Priorities

The last criterion is that the intervention must address societal priorities; in other words, the problem must be important in the larger picture. Sufficient agreement first needs to exist with regard to the importance of the health problem. This consensus should have been established

during the priority-setting and assessment phases. Lack of the desired healthy state or a high prevalence of the problem may contribute to granting high priority. By contrast, many effective interventions can be used to address trivial problems of low priority.

Health program planners and evaluators might potentially play a role in raising the priority of the issue so that the health problem takes a more prominent place. To some extent, societal priority is set by celebrity spokespersons for specific health problems or by the nightly news covering the current health research. These societal pressures may conflict with the local assessment data. Nevertheless, the intervention must be aligned with the societal priorities assigned to health problems if it is to receive public credibility and backing. Also, the new behavior or health state must be important to the intended audience members or they will not make attempts to change. Although the importance of the health problem to the intended audience may have been included as an element in the community needs assessment, this issue can resurface during program theory development in terms of societal versus public health priorities.

► Path to Program Outcomes and Impacts

Careful consideration of which interventions will be used in the health program must go hand-in-hand with specifying which effects are anticipated from the program. Several factors can distract program planners from having a clear vision of the relevant effect. For example, a plethora of possible outcomes from programmatic interventions may exist. There may also be many ways to think about changes resulting from programs (Patton, 1997, p. 160). Yet another distraction is that, with extensive stakeholder involvement, it is quite possible to become sidetracked and end up with an extensive list of what “our program could do.” For these reasons, having the community diagnosis, as

written at the conclusion of the community needs assessment, is important because it helps those involved in the planning process stay focused on both the health problem and those health outcomes and impacts that are directly related to the health program.

Further complicating the choice of key health outcomes and impacts is the reality that change is not always the purpose of health programs; some programs are, in fact, intended to stabilize, prevent, or maintain a health state. In addition, health is multidimensional, encompassing bio-psycho-social dimensions, and social determinants of health might be the focus of health program. Nonetheless, a specific end point is desired and that essentially constitutes the outcome or impact.

Components of the Effect Theory

After having considered the type of intervention and the criteria for choosing an intervention, the next step is to articulate more fully the effect theory by enumerating the causal, intervention, and impact theories that constitute the effect theory. This iterative process requires going back and forth between the needs assessment, priorities, and intervention choice.

Both inductive and deductive approaches can be used to generate an effect theory. In other words, theory development can proceed through a deductive process that uses reason and existing knowledge, or it can occur through an inductive process that uses experience and intuition. Either approach will lead to an effect theory. In practice, a combination of both inductive and deductive approaches is typically used and yields the optimal results. Generating an effect theory requires both creativity and intellectual rigor.

Elements of the effect theory draw on the community diagnosis developed for each of the high-priority health problems as well as the literature. Recall the template for the community diagnosis: Risk of [*health problem*] among [*population/community*], indicated in [*health indicators or measures*], is caused by [*causative factors*], but is mediated by [*mediating factors*] given that [*moderating factors*] moderate the causes

and that [*existing factors*] exist prior to the causes. Outlining and articulating the components of the effect theory help ensure that the program will achieve the desired effect. The effect theory, as the overall explanation of how the program interventions lead to the desired effect, integrates the three narrowly focused subcomponent theories.

Theory of the Causal/Determinant Forces

One component of the effect theory is the *causal theory*, which is an explanation of the process that currently underlies the health problem. It includes statements or hypotheses that describe which causal factors, including key relevant social ecological determinants, are directly responsible for the health problem. The causal theory ought to include the factors found present through the community needs assessment and draw on the scientific literature to justify the causal theory.

For example, we can use the community diagnosis related to deaths from gunshot wounds to develop a causal theory. The causal theory states that deaths from gunshot wounds stem from causal factors of local gang activity, lack of conflict resolution skills, dropping out of school, and gun availability. Individual resilience, adequacy of policing, and quality of emergency medical care are mediating factors that determine whether the causal factors actually result in a death. In addition, the adolescent's developmental stage, local history of violence, lack of job opportunities, and state laws, as preexisting forces, influence whether the causal factors exist. Community action, parental supervision, and school antiviolence programs all have the potential to moderate—either decreasing or increasing—the potency of the causal factors.

Similarly, the community diagnosis for birth defects is the basis for a causal theory of birth defects in Bowe County. The causal theory states that birth defects among residents of Bowe County are caused by low folic acid intake, parental exposure to organic solvents, prenatal exposure to chlorine, and the Zika virus. However, preconception nutritional status and biological processes (mediating factors) influence whether

the causal factors actually result in a birth defect. In addition, the mother's age, type of employment, and availability of food high in folic acid, as contextual preexisting factors, determine whether the causal factors exist. Genetic counseling, taking prenatal vitamins, knowledge about folic acid, and mosquito abatement all have the potential to moderate the influence of the causal factors by either increasing or decreasing their potency.

Theory of Intervention Mechanisms

The *intervention theory* explains how interventions affect which of the causal/determinant factors, or possibly the moderating or mediating factors. It contains hypotheses about the relationships of the programmatic interventions to the factors in the causal theory that the interventions are intended to affect. More important, it must address how the intervention alters the causal factors or breaks the chain between causal factors and health outcome. The intervention theory includes statements describing the relationships connecting interventions and outcomes. The intervention might also affect some of the moderating or mediating factors. Thus, the intervention theory articulates all the connections between the programmatic intervention and the intended effects on the health problem. Having the intervention theory explicitly stated and understood by the program staff members contributes to the success of the program.

The intervention theory describes how the program “works its magic.” Developing an intervention theory is useful to refine the number, types, and quality of interventions that are carried out as part of the health program. Interventions that are not likely to alter or change the key factors in the causal theory can, in turn, be eliminated, which results in a more effective and efficient program.

In the birth defects health problem, planners might identify several possible points at which to intervene to ensure that the causal factors do not lead to neural tube defects. For example, the program might focus on the moderating factor regarding knowledge about the importance of folic acid. Accordingly, one part of the intervention theory would state that nutritional

education [intervention] changes the behavior of the woman with regard to eating dark green vegetables. Another point at which to intervene on the causal factors might be by encouraging the use of prenatal vitamins [intervention] to remove the causal factor of inadequate folic acid intake. Also, screening for occupational exposures followed by an early ultrasound [intervention] could identify fetuses with abnormalities. Receiving nutritional education, taking supplements, and making changes in prenatal care can, together, alter the biological processes that result in a neural tube defect.

As this example shows, not all moderating, causal, or mediating factors need to be, or can be, addressed within a single health program. An equally plausible intervention theory might state that education about occupational exposures [intervention] leads to decreased exposures and subsequently fewer infants with neural tube defects. Alternatively, an intervention theory might focus on the environment context related to eliminating the mosquito carrying the Zika virus. The decision regarding which intervention theory to use as the basis of a program is influenced by the preferences of stakeholders, the mission of the organization, and the science regarding which factors are more readily changeable and effective.

Theory of Outcome to Impact

The final element of effect theory is the *impact theory*, which explains how the outcomes lead to impacts. Usually, a health program has a very limited number of health outcomes that it seeks to affect. Impact theory helps substantiate the sometimes seemingly wild and wishful claims of program planners about the effects of their program by specifying the relationship between the immediate outcome of the program and the long-term, ultimate changes to the health problem. It is possible to have multiple impact theories for one long-range impact, especially if multiple intervention theories are used within a single program. Given the complex nature of many health problems and conditions, this is a likely scenario. Continuing with the birth defects example, the impact theory states that fewer

infants born with neural tube defects leads to a decrease in the rate of birth defects of all types.

Funding agencies commonly specify program impacts—for example, a decrease in infant mortality or an increase in early detection of preventable disease. These impacts might be stated as program goals that the funded programs are to achieve. In such cases, program planners must essentially work backward to generate the impact theory and the intervention theory. In addition, impact theories show the links and explain the relationships between the program objectives regarding outcomes and program goals specifying desired impacts.

Matching Levels: Audience, Cause, Intervention, and Effects

In developing the effect theory, care must be taken to ensure that the level of the pyramid where the intervention functions matches both the intended audience and the level of the public health pyramid at which the outcomes and impacts are expected. Intended audiences may consist of individuals, families, aggregates, or populations, with effects occurring at each of those levels. Program

interventions need to be tailored to reach that specific intended audience, essentially matching the audience level at which the intervention is aimed to the level at which the intended audience exists, and to the level at which the outcome is desired. For example, if the intervention is designed to affect family eating patterns, then the health outcome sought ought to be family nutritional health rather than reducing anemia in children or increasing the daily consumption of milk in a neighborhood. The importance of being clear about the level or unit for the intervention has long been recognized as pivotal for the evaluation phase; the unit of intervention becomes the unit of analysis in the evaluation phase (Jackson, Altman, Howard-Pitney, & Farquhar, 1989). In the family eating pattern example, the family, as a unit, is the level at which the intervention is focused and would be the unit of analysis in the evaluation of effects.

In summary, the effect theory integrates the theories of causal/determinant forces, intervention mechanisms, and outcome with impact. The combination of these three components of the effect theory explains the complexity of addressing a health problem. **FIGURE 5-3** brings together all the components of the effect theory in the birth defects example.

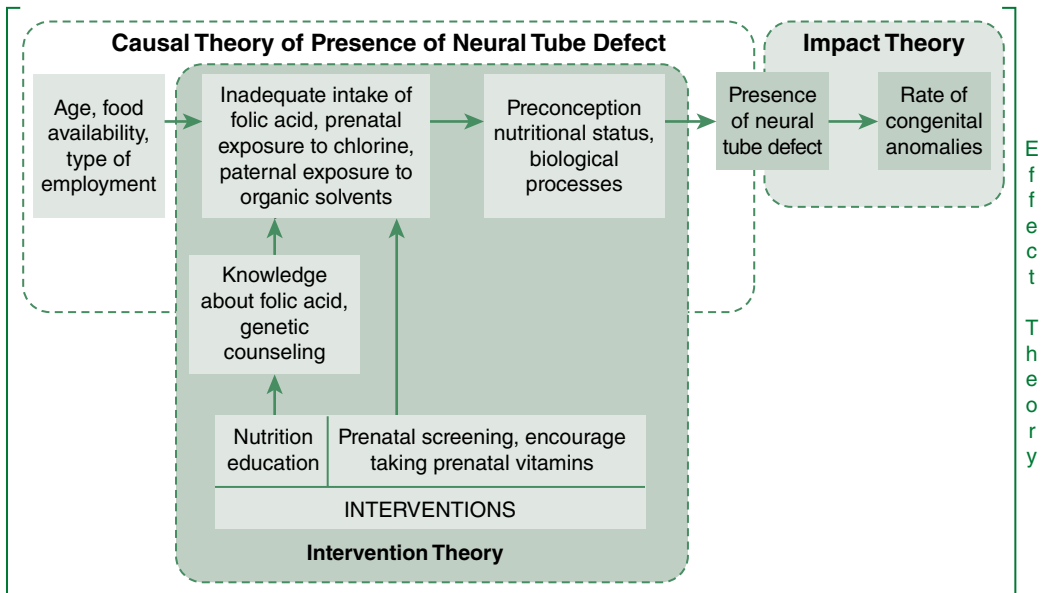


FIGURE 5-3 Effect Theory Example: Effect Theory for Reducing the Rate of Congenital Anomalies

► Generating the Effect Theory

Involve Key Stakeholders

Generating a program theory is not a solitary task; it requires brain power, diverse ideas, and sustained energy. Involving key stakeholders not only makes good ideas evident but also encourages stakeholders to become invested in the health program and to address the health problem. This type of involvement is a critical step toward having a politically feasible intervention.

Potential program participants and providers typically have their own working explanation, or theory, of how a program will affect participants. One type of theory they may advocate is an espoused theory. Argyris and Schon (1974) were among the first to understand the importance of espoused theories. They found that employees had explanations for why things happen in their organizations; these stated explanations are the espoused theories. People know what they are supposed to do or say, regardless of whether they actually do or say it. The espoused theory consists of this stated and repeated explanation. Argyris and Schon (1974) also found that espoused theories were not always congruent with the behaviors they observed. What people do to achieve their ends is termed their theory-in-use, sometimes called a theory-in-action. For example, Nichols and Badger (2008) found that hospital staff members espoused one practice for infection control but actually had a different practice or theory-in-use. The theory-in-use is crucial in program evaluation because it consists of the interventions that actually make up the health program and affect participants. For example, in a diabetes management program, if the staff members become friends with the patients and provide encouragement in a supportive manner, but they rarely focus on teaching patients, then their theory-in-use is coaching or social support rather than education.

As seen in the diabetes management example, espoused theories and theories-in-use may

not be congruent. The theories-in-use denote how the program is implemented and are the source of the effects on participants. One way to avoid incongruity between the espoused theory and the theory-in-use is to include the theory-in-use explicitly in the effect theory. Being aware of the differences among espoused theories, theories-in-use, and effect theories (TABLE 5-2) can help planners generate an effect theory that incorporates useful elements of both the espoused theories and the theories-in-use. If the program has been in existence for some time, an alternative is to decide either to incorporate the theory-in-use into the program theory or to exclude the theory-in-use explicitly from the program. Modifying the program theory based on the practical experience gained through the theory-in-use may be efficient and prudent if the theory-in-use has had the desired effect on program participants.

Draw Upon the Scientific Literature

Program planners should review articles published across the health disciplines for information that can help them generate the theories by providing information on the relationships among the antecedents and causal, moderating, and mediating factors. Abstracts available through online databases are another good source of ideas that can be incorporated into the effect theory. The published literature is also helpful in developing the process theory, particularly with regard to the service utilization elements.

Existing theories from multiple disciplines can be used to develop the effect theory. For example, if the health program is intended to have a physiological effect or to address a certain pathology, then theories from genomics, biochemistry, pharmacology, or physiology might be useful. In contrast, theories from psychology or social work about psychopathology, stress, coping, or family functioning might be used to explain a family problem. For programs

TABLE 5-2 Comparison of Effect Theory, Espoused Theory, and Theory-in-Use

	Effect Theory	Espoused Theory	Theory-in-Use
What it is	Explanation of how program interventions affect participants	What staff members say about how the program affects participants	What staff members actually do that affects participants
Where it resides	Manuals and procedures; program descriptions	Minds of program staff members; program manuals and descriptions	Actions of program staff members; on-the-job training
How it is identified	Review of scientific literature, program materials	Listen to staff members describe the program; read program materials	Watch what staff members do in providing the program and intended interventions
Importance	Guides program evaluation; basis for claiming outcomes	Becomes what staff members, clients, and stakeholders believe and expect of the program	Is the actual cause of program outcomes

intended for population-level changes, theories from epidemiology, decision sciences marketing, social networks, or public health may be helpful. Integrating or borrowing theories from across disciplines may be necessary, especially as a means to address multiple determinants of health problems.

Many existing theories can help health program planners develop causal theories for health problems and situations. The examples listed in **TABLE 5-3** are grouped by the domain of health outcomes anticipated by the program as a reminder that ultimately the program intervention theory must be matched with both the health problem and the desired outcomes of the program. In addition, existing theories can be used in developing the process theory; examples of such theories are shown in **TABLE 5-4**.

The theories used by program planners are generally specific to the level of the public health pyramid. In this section, the examples are largely at the individual level. For problems at the other levels of the pyramid, theories can be found in the literature. For example, Gay (2004) relied on the theory of disease transmission as a framework for understanding what is required to develop a program to eliminate measles, an infectious disease. Although measles is an individual illness, the elimination of any infectious disease—whether measles, tuberculosis, or HIV/AIDS—requires thinking in terms of populations as well as individual susceptibility. To change population behaviors related to alcohol use problems, Wallin (2007) reported that the successful program was based on the diffusion of innovation theory.

TABLE 5-3 Examples of Types of Theories Relevant to Developing Theory of Causative/Determinant Factors or Theory of Intervention Mechanisms by Four Health Domains

Physical Health	Psychosocial Health	Knowledge and Abilities	Lifestyle Behaviors
Pathophysiology Immunology Endocrinology Pharmacology Wound healing Biochemistry Metabolism	Psychopathology Social cognition Stress and coping Family functioning Addiction Violence Resilience Discrimination	Learning Communication Cognition Attention Memory Diffusion of innovation Acculturation	Peer pressure Decision making Self-efficacy Self-worth Risk taking Social stratification Motivation Social networks Economic nudging Urban design Health services utilization

TABLE 5-4 Examples of Types of Theories Relevant to Developing the Organizational Plan and Services Utilization Plan Components of the Process Theory

Organizational Plan	Service Utilization Plan
Social network Communication Leadership Accounting Quality improvement	Social marketing Marketing Cueing Economic incentivizing Pricing

Diagram the Causal Chain of Events

Drawing or creating a visual representation of the various theories is important given the complex nature of the causes of health problems and the equally complex systems of services required to address health problems. Diagrams that depict the effect theory, the process theory, and the overall program theory can be created with pencil and paper or by using graphics software. A picture showing how each intervention changes a characteristic of the participants provides an expedient

means of engaging program staff members and getting feedback from other professionals in the field. As the scientific literature is reviewed and assimilated, additional relevant variables and their interrelationships can be incorporated into the map of the causal chain of events. Including every possible variable is neither realistic nor desirable, of course; instead, program planners should include only those variables that relate to the essence of the program and that, according to the community health assessment and available scientific literature, are mostly likely to influence the success of the proposed interventions.

In some instances, a health program is started in response to a mandate or a health policy initiative and therefore may not have an explicit program theory. If a program has been in existence or is ongoing, the development of a program theory is still possible, and its creation instead can contribute to program improvements. In such cases, the espoused theory of program staff members is a good starting point for the development of a program theory. Observations of program staff members would then help identify the theory-in-use. Together with findings from the literature, these elements could be formalized into an effect theory. It is quite possible that new areas for program monitoring and evaluation would emerge from such an exercise with program staff. In addition, program staff members may come to see the value of their work and become more committed to the program and the participants. Involving program staff members in reconciling their espoused theories and theories-in-use can lead to new program approaches and the identification of areas of inefficiencies.

For some health programs, timing is critical such that some intervention components must be accomplished before other intervention components are implemented. If either the intervention or the outcomes must proceed in stages, these increments need to be reflected in the effect theory of the causal chain of events leading to the health outcome.

Check Against Assumptions

The program theory—and the effect theory in particular—needs to be checked against alternative assumptions about theories. Patton (1997) referred to these points as validity assumptions. One assumption is that the theory is really about the phenomenon of interest. In other words, program planners assume that the program theory truly deals with the health problem or condition that is the focus of the health program. Through the multiple interactions and discussions with stakeholders, this assumption can inadvertently be violated.

Another assumption relates to parsimony. Improving the health of individuals, families, and communities is a complex task, so most health programs address only one aspect of a complex puzzle of factors affecting health. Including too much in a program theory can lead to confusion, diffuse interventions, and frustration, not to mention exorbitant expenditures. Parsimony is a crucial characteristic of a good theory, including an effect theory. Relying on the priorities set earlier in the planning process by focusing on the most important factors about the target audience helps achieve parsimony.

► Functions of Program Theory

Having an articulated theory of how the health program will lead to improved health, and specifically how the interventions will affect participants, serves several purposes (Bickman, 1987) that range from providing guidance and enabling explanation to forming a basis for communication.

Provide Guidance

A program theory that can be stated in one or two sentences provides a description of what is being implemented. To say that a program is helping asthmatic children is less compelling or descriptive than saying that a program teaches children how to be aware of their bodies and thereby avoid situations that may trigger an asthma attack. The latter is a description of how the program works to reduce asthma attacks and provides direct guidance on what to include in the program.

In a world of complex and interactive health problems, identifying the specific health problem and the appropriate intended audience for a program can be difficult. Blum's (1982) caution against failure to analyze problems adequately is avoided by developing the program theory,

which specifies the problem and the intended audience. If the program theory is inordinately difficult to develop, it may indicate that the health problem has not been sufficiently narrowed, the intended audience is not specific enough, or too many program elements have been included. Having an intended audience that is too broad can lead to a program theory that is too complex to be of value in designing and implementing the program.

The program theory guides what to measure in both the process and the effect evaluations of the program. In terms of the process evaluation, it specifies what needs to be measured with regard to the delivery of the intervention. In terms of the effect evaluation, the effect theory specifies the desired effects and therefore what needs to be measured. When a health program has several possible outcomes, the effect theory clarifies which outcome is most directly a result of the intervention. This information makes the evaluation of outcomes more efficient and enables program planners and evaluators to design an evaluation that will find those program effects that are arguably the result of the program.

Just as theory is used to guide the development of the health program, so theory can be used to guide the development of the evaluation. For example, Newes-Adeyi, Helitzer, Caulfield, and Bronner (2000) used ecological theory to guide their formative evaluation of New York State's Women, Infants, and Children (WIC) nutritional program. Their use of ecological theory strengthened the evaluation in terms of its design and ability to explain how the program worked. Their report also serves as a reminder that the same underlying social or psychological theory that guides the effect theory can be applied to the effect evaluation as well.

When a new health program is first provided, its evaluation helps refine the subsequent delivery of the program. A program theory helps identify needed inputs and determine what needs to be evaluated and where improvements or changes in the delivery of the interventions are appropriate.

Enable Explanations

The program theory helps identify which interventions are likely to have the greatest effect on program participants and clarify how the interventions cause the desired effect in program participants. In this way, the theory enables planners and evaluators to explain more easily how the program should and does work.

One task of program planners is to anticipate the unintended. Careful attention to the development of the program theory can help uncover unintended consequences that may result from the program. The development of an effect theory, in turn, helps generate plausible explanations for those unintended consequences. Engaging in this kind of exercise in speculation helps program planners avoid another source of unsuccessful programs: failure to examine and compare relevant possible interventions (Blum, 1982).

A program theory also enables the evaluators to distinguish between process theory failure and effect theory failure (**FIGURE 5-4**). If the evaluation results show no effect on program participants, then the evaluator must explain what failed. A successful program sets into motion the interventions (causal processes) that lead to the desired outcome. However, if a program is not effective, the evaluator needs to identify the roots of that failure. A lack of program success can result from the program not being provided—a process theory failure. A lack of program success also can result from an ineffective intervention—an effect theory failure. This distinction between process and effect theory failures, based on the notions of program and theory failure put forth by Weiss (1972), helps evaluators sort out what went wrong or right with the program and explain the evaluation findings to stakeholders.

Form a Basis for Communication

Health programs compete for resources. A program theory helps convince organizational or legislative policy makers that the program is

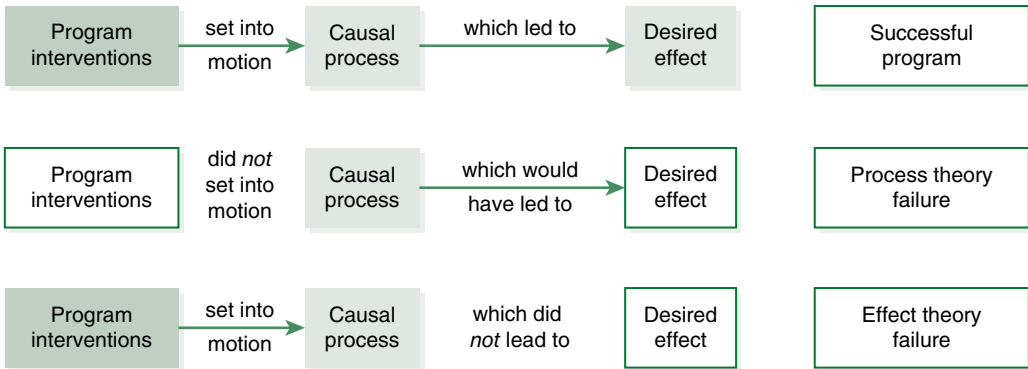


FIGURE 5-4 Two Roots of Program Failure
 Weiss, C. H. (1998). *Evaluation*, (2nd ed.). Reprinted by permission of Pearson Education, Inc., New York, NY.

worthy and deserving of support. The causal chain of events outlined in the effect theory serves to frame discussions on a more rational basis, leading to a more rational decision-making process about the health program. The effect theory also helps policy makers understand the extent to which the program interventions are ideologically compatible with their stance and are based on science rather than biases and opinions. In other words, the effect theory provides a basis for clear communication of the program intent and content.

Starting and maintaining a program requires that key stakeholders agree on supporting the program. Gaining consensus from stakeholders—whether program staff members, administrators, or legislators—is an important step in ensuring the success and acceptance of the health program. If stakeholders understand the program theory, it becomes easier to gain consensus on the usefulness of the program. Having gone through the exercise of developing the causal, intervention, and impact theories, the program planners are in the position of being better able to anticipate questions and provide alternative rationales for the health program. As mentioned earlier, stakeholders can be included in the development of the program theory as a way to gain consensus on the program interventions. For controversial programs, such as those dealing with sexuality education or family planning for adolescents,

consensus on the program theory could be critical to the program’s survival.

Make a Scientific Contribution

In a sense, every health program is an experiment that tests the program theory. In other words, every evaluation has the potential to contribute to our understanding of human nature and health. Evaluations based on the program theory can be used to modify existing theories relevant to the target population and types of interventions used.

► Across the Pyramid

At the direct services level of the public health pyramid, the health problems are related to specific individuals, so the relevant theories will focus on individual behavior and intra-individual responses to treatment or pathology. In other words, the focus is on the micro level. As a result, the interventions delivered are one-on-one, with providers directly delivering the interventions to their clients. (Examples of direct services interventions appear in Table 5-1.) If the program will have subcomponents, those components would involve different types of interventions that are delivered directly to individuals.

At the enabling services level, the health problems are related to aggregates of individuals, so the relevant theories will focus on the interactions

of individuals with family or community characteristics. Because enabling services are still provided to individuals, the focus continues to be at the micro level. Hence, interventions are delivered on a one-on-one basis, as well as to groups with similar characteristics. Different intervention types can be applied at the enabling services level (Table 5-1).

At the population level, the health problems are related to entire populations, so the relevant theories will focus on group responses that lead to the health problem, cultural theories that explain behaviors and beliefs related to the health problem, and social theories about interactions among groups. Liddle and Hogue (2000), for example, described an intervention for high-risk adolescents. One key feature of their intervention model was that the theoretical foundation included risk and protection theory, developmental psychopathology theory, and ecological theory. This blend of theories is consistent with the intent of the program. In terms of the public health pyramid, however, the use of ecological theory reflects the theoretical awareness of the program planners that the population level influences both the enabling level (i.e., the family) and the individual level. At the population level, the interventions are designed and intended to have a universal focus. Such interventions are more likely to be delivered through the mass media or to involve policy formation. Having program components at the population level may create synergies that enhance the intervention, as well as possibly increasing feasibility and decreasing long-term program costs.

At the infrastructure level, the problems are related not to individuals but rather to processes and structures that enable the delivery of health programs, so relevant theories might focus on organizational behavior, management and leadership style, personnel motivation, political action, and communication. The interventions can be delivered one-on-one with personnel, as well as with groups of workers or entire organizations. Because workforce capacity building is a key focus at the infrastructure level, it may be appropriate to use individual-level theories. For example, Kirk, Tonkin, and Burke (2008) used the theory

of planned behavior as the basis for enhancing genetics literacy among health professionals.

► Discussion Questions and Activities

1. Select a health program with which you are familiar.
 - a. Briefly state the hypotheses that constitute the effect theory of the program.
 - b. What are the intervention components and the specific interventions?
 - c. Develop an effect theory of the program theory used by the program.
 - d. Do a brief literature search to determine whether the scientific evidence supports the interventions used.
2. What are the relationships among the possible functions of effect theory and the selection of optimal interventions?
3. Which of the theories that make up the effect theory are likely to be affected by the cultural, ethnic, or racial differences of target populations? In what ways might you make those theories culturally appropriate or sensitive?
4. Identify possible primary, secondary, and tertiary prevention interventions for each level of the public health pyramid.
5. Figure 5-3 shows a possible effect theory, with the interventions, to address the health problem of congenital anomalies. Using this as a template or an example, try developing an effect theory diagram for one specific health problem of interest to you.

► Internet Resources

University of Iowa, College of Nursing

This website (www.nursing.uiowa.edu/center-for-nursing-classification-and-clinical-effectiveness)

provides an overview of standard nursing interventions (NIC) and outcomes (NOC). The detailed list can be helpful to show the level of specificity for interventions, which may be needed for some programs.

Wheel of Public Health Interventions

Each of 16 different interventions can be explored at this interactive website. The intervention examples are more widely applicable than is the public health nursing focus. <http://www.people.vcu.edu/~elmiles/interventions/>

Understanding Change and Theories Critical in Developing Program Theory

The following websites focus on understanding and generating change. For example, the chapter found in the “Theory at a Glance” published in 2005 (https://cancercontrol.cancer.gov/brp/research/theories_project/index.html) is part of a short, online text and nicely summarizes the theories often used in public health. If you want to broaden your repertoire of change theories, then the information at the Communication Initiative Network (www.comminet.com) would be helpful.

Community Guide to Preventive Services

This website lists interventions for various health topics and the degree of scientific evidence for the use of the intervention: www.thecommunityguide.org. These systematic literature reviews focus on interventions for populations.

References

- Abroms, L., Hershcovitz, R., Boal, A., & Levine, H. (2015). Feasibility and acceptability of a text messaging program for smoking cessation in Israel. *Journal of Health Communication, 20*, 903–990. doi:10.1080/10810730.2015.1018585
- Agryis, C., & Schon, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco, CA: Jossey-Bass.
- American Medical Association. (2016). *CPT 2016: Professional edition*. Chicago, IL: American Medical Association.
- Mapuana C. K., Antonio, M. A., & Chung-Do, J. J. DrPH. (2015). Systematic review of interventions focusing on indigenous adolescent mental health and substance use. *American Indian and Alaska Native Mental Health Research (Online)*, 22(3), 36–56. Retrieved from <https://librarylink.uncc.edu/login?url=http://search.proquest.com/docview/1787152385?accountid=14605>
- Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation, 43*, 5–18.
- Blum, H. L. (1982). Social perspective on risk reduction. In M. M. Farber & A. M. Reinhart (Eds.), *Promoting health through risk reduction* (pp. 19–36). New York, NY: Macmillan.
- Bonell, C., Jamal, F., Melendez-Torres, G. J., & Cummins, S. (2014). ‘Dark logic’: Theorizing the harmful consequences of public health interventions. *Journal of Epidemiology and Community Health, 69*, 95–98. doi:10.1136/jech-2014-204671
- Bulechek, G., Butcher, H., Dochterman, J., & Wagner, C. (Eds.). (2013). *Nursing interventions classification (NIC)* (6th ed.). St. Louis, MO: Elsevier.
- Cooke, E. V. (2010). The effects of increased dose of exercise-based therapies to enhance motor recovery after stroke: A systematic review and meta-analysis. *BMC Medicine, 8*(60). doi:10.1186/1741-7015-8-60
- Crooks, C. V., Scott, K., Ellis, W., & Wolfe, D. A. (2011). Impact of a universal school-based violence prevention program on violent delinquency: Distinctive benefits for youth with maltreatment histories. *Child Abuse and Neglect, 35*(6), 393–400.
- De-Regil, L. M., Peña-Rosas, J. P., Fernández-Gaxiola, A. C., & Rayco-Solon, P. (2015). Effects and safety of periconceptional oral folate supplementation for preventing birth defects. *Cochrane Database of Systematic Reviews*, 12. Article Number CD007950. doi:10.1002/14651858.CD007950.pub3
- Des Jarlais, D., Sloboda, Z., Friedman, S. R., Tempalski, B., McKnight, C., & Braine, N. (2006). Diffusion of the D.A.R.E. and syringe exchange programs. *American Journal of Public Health, 96*, 1354–1358.
- Fisher, E. B., Browston, C. A., O’Toole, M. L., Shetty, G., Anwuri, V. V., & Glasgow, R. E. (2005). Ecological approaches to self-management: The case of diabetes. *American Journal of Public Health, 95*, 1523–1535.
- Gay, N. J. (2004). The theory of measles elimination: Implications for the design of elimination strategies. *Journal of Infectious Diseases, 189*(1S), S27–S35.
- Grobe, S. J., & Hughes, L. C. (1993). The conceptual validity of a taxonomy of nursing interventions. *Journal of Advanced Nursing, 18*, 1942–1961.
- Guttentag, C. L., Landry, S. H., Williams, J. M., Noria, C. W., Borkowski, J. G., . . . Warren, S. F. (2014). “My Baby & Me”: Effects of an early, comprehensive parenting intervention on at-risk mothers and their children. *Developmental Psychology, 50*, 1482–1496. doi:10.1037/a0035682

- Harris, D. A. (2007). Dance/movement therapy approaches to fostering resilience and recovery among African adolescent torture survivors. *Torture, 17*, 134–155.
- Heaney, C. A., & van Ryn, M. (1996). The implications of status, class, and cultural diversity for health education practice: The case of worksite stress reduction programs. *Health Education Research, 11*(1), 57–70.
- Issel, L. M. (1997). Measuring comprehensive case management interventions: Development of a tool. *Nursing Case Management, 2*, 3–12.
- Jackson, C., Altman, D. G., Howard-Pitney, B., & Farquhar, J. W. (1989). Evaluating community-level health promotion and disease prevention interventions. In M. T. Braverman (Ed.), *Evaluating health promotion programs* (pp. 19–32). San Francisco, CA: Jossey-Bass.
- Kirk, M., Tonkin, E., & Burke, S. (2008). Engaging nurses in genetics: The strategic approach in the NHS national genetics education and development center. *Journal of Genetic Counseling, 17*, 180–188.
- Liddle, H. A., & Hogue, A. (2000). A family-based, developmental-ecological preventive intervention for high-risk adolescents. *Journal of Marital and Family Therapy, 26*, 265–279.
- Lipman, E. L., Waymouth, M., Gammon, T., Carter, P., Secord, M., Leung, O., Hicks, F. (2007). Influence of group cohesion on maternal wellbeing among participants in a support/education group for single mothers. *American Journal of Orthopsychiatry, 77*, 543–549.
- Lundgren, R., & Amin, A. (2014). Addressing intimate partner violence and sexual violence among adolescents: Emerging evidence of effectiveness. *Journal of Adolescent Health, 56*, S42–S50. doi:10.1016/j.jadohealth.2014.08.012
- Martin, C. K., Church, T. S., Thompson, A. M., Earnest, C. P., & Blair, S. N. (2009). Exercise dose and quality of life: A randomized controlled trial. *Archives of Internal Medicine, 169*(3), 269–278.
- Monsen, K. A., & Newsom, E. T. (2011). Feasibility of using the Omaha System to represent public health nurses manager interventions. *Public Health Nursing, 28*(5), 421–428. doi:10.1111/j.1525-1446.2010.00926
- Morone, N. E., Greco, C. M., & Weiner, D. K. (2008). Mindfulness meditation for the treatment of chronic low back pain in older adults: A randomized controlled pilot study. *Pain, 134*, 310–319.
- Newes-Adeyi, G., Helitzer, D. L., Caulfield, L. E., & Bronner, Y. (2000). Theory and practice: Applying the ecological model of formative research for a WIC training program in New York State. *Health Education Research, 15*(3), 283–291.
- Nichols, A., & Badger, B. (2008). An investigation of the division between espoused and actual practices in infection control and of the knowledge sources that may underpin this division. *British Journal of Infection Control, 9*(4), 11–15.
- Olson-Keller, L., Strochschein, S., Lia-Hoagberg, B., & Schaffer, M. A. (2004). Population-based public health interventions: Practice-based and evidence-supported, Part I. *Public Health Nursing, 21*, 453–468.
- O'Mara-Eves, A., Brunton, G., Oliver, S., Kavanagh, J., Jamal, F., & Thomas, J. (2015). The effectiveness of community-engagement in public health interventions for disadvantaged groups: A meta-analysis. *BMC Public Health, 15*, 129. doi:10.1186/s12889-015-1352-y
- Pan, W., & Bai, H. (2010). A multivariate approach to a meta-analytic review of the effectiveness of the D.A.R.E. Program. *International Journal of Environmental Research and Public Health, 6*(1), 267–277.
- Patton, M. Q. (1997). *Utilization focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.
- Perchmann, C., & Reibing, E. T. (2006). Antismoking advertisements for youths: An independent evaluation of health, counter-industry, and industry approaches. *American Journal of Public Health, 96*, 906–913.
- Potvin, L., Gendron, S., Bilodeau, A., & Chabot, P. (2005). Integrating social theory into public health practice. *American Journal of Public Health, 95*, 591–595.
- Resnicow, K., Davis, R., Zhang, N., Strecher, V., Tolsma, D., Calvi, J., William, C. (2009). Tailoring a fruit and vegetable intervention on ethnic identity: Results of a randomized study. *Health Psychology, 28*(4), 394–403.
- Rossi, P., Freeman, H., & Lipsey, M. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Thomas, I. M., Sayers, S. P., Godon, J. L., & Reilly, S. R. (2009). Bike, walk, and wheel: A way of life in Columbia, Missouri. *American Journal of Preventative Medicine, 37*(6), 322–328.
- Travagin, G., Margola, D., & Revenson, T. A. (2015). How effective are expressive writing interventions for adolescents? A meta-analytic review. *Clinical Psychology Review, 36*, 42–55. Retrieved from <http://dx.doi.org/10.1016/j.cpr.2015.01.003>
- Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose-response meta-analysis of multiple outcomes. *Clinical Psychology Review, 30*(4), 387–399.
- Wallin, E. (2007). Dissemination of prevention: Community action targeting alcohol use-related problems at licensed premises. *Substance Use and Misuse, 42*, 2085–2097.
- Weiss, C. (1972). *Evaluation*. San Francisco, CA: Jossey-Bass.

CHAPTER 6

Program Objectives and Setting Targets

In this chapter, the focus is on setting the parameters by which the program is judged as successful—in other words, developing goals and objectives for the program. Setting goals and objectives gives clarity and specificity to the implementation plan and guides subsequent evaluation effects. After a logic model has been developed, setting goals and objectives are the next step in program planning.

► Program Goals and Objectives

Goals and *objectives* are terms that are widely used in program planning and evaluation. *Goals*, in a strict sense, are broad, encompassing statements about the impact to be achieved, whereas *objectives* are specific statements about outcomes to be achieved and are stated in measurable terms. Funding bodies do not use the terms *objectives* and *goals* consistently, so program planners and evaluators must understand the difference. The distinction between objectives and goals forms the basis for the conceptual distinctions between short-term outcomes and long-term impacts of the program.

Goals and their corresponding objectives reflect the logic model and the program theory. Involving the stakeholders and program staff members in the development of the program goals and objectives can be useful in gaining their support, stimulating good ideas, and reaching a consensus on what will constitute the program. However, the process of reaching a consensus, particularly on objectives, can be a bit of a struggle if stakeholders have vested interests in achieving particular health outcomes for their constituents. In addition, program planners often must cope with tight schedules for preparing a program proposal, making timely involvement of stakeholders a challenge. The efforts devoted to arriving at a set of clearly articulated goals and objectives do pay dividends, however; they lay the foundation from which to develop the evaluation and establish standards against which to assess the success of the program.

Goals

Goals are always statements about the health impact or status of the intended audience, and generally apply to a longer time horizon, such as 5 years. Typically, goals do not incorporate a quantifiable

measure but instead refer in broad terms to the most important anticipated effect of the program. A program will have at least one goal, and a well-focused program with several components may have more than one. In general, however, the number of goals is quite low. The use of creative activities, stories, and clear communication can make writing goals a positive experience.

Good goals are congruent with and contribute to the strategic plan, whether it consists of a national (e.g., *Healthy People 2020*), state, or local health plan. For health programs being developed by local health agencies or local community-based organizations, this larger context of health programs can be crucial in achieving synergies within the program as well as between existing programs with complementary foci. Having program goals and objectives that also are compatible with the strategic or long-term plan of the healthcare organization can affect the priority given to the program and hence the fiscal support provided for, and organizational approval of, the program.

On a cautionary note, Friedman, Rothman, and Withers (2006) provocatively remind evaluators that using goals for evaluation creates a paradox. Goals implicitly or explicitly embody values. The paradox arises when goals are used to guide an evaluation because the evaluation then also embodies those values. This is not necessarily a bad thing, but it serves as a reminder that who is involved in developing and approving the goals is important.

Foci of Objectives

Objectives are precise statements about the work to be done or an immediate effect on program participants, and have a short time frame. Development of objectives begins with having conceptual clarity regarding whether the objective is related to the program's process theory or effect theory. Objectives focus specifically on the process theory or the effect and thus are called either process or effect objectives. Objectives address specific program processes or effects, so they provide more direct guidance for the program implementation. For this reason, program planners and evaluators

need to understand the particulars of the format and development of objectives.

An easy format that helps in remembering the parts of a good objective is this: "by when, who will achieve what, by how much." For example, one objective may be: "By 2012, the Layetteville Innovation—Adolescent Preventing Pregnancy (i-APP) Program will reduce the pregnancy rate among program participants by 20% compared to girls not participating." The "by how much" portion of the objective, or the target value, is the quantifiable measure that distinguishes an objective from a goal. The target value is the essence of the objective; without it, no objective exists.

The statement, "The percentage of pregnant adolescent girls among girls enrolled in the Layetteville i-APP Program during 2010 will be reduced" is a goal, not an objective. A goal may have several objectives that delineate more precisely what achieving the goal entails. Thus, adding a target value would yield the following for reaching that goal: "The percentage of adolescent girls who become pregnant among girls enrolled in the Layetteville i-APP Program during 2010 will be 8%." The "8%" quantifies the reduction and is measurable.

The time frame used in the objectives needs to be short term and well within the life span of the program. For direct services and enabling services, objectives generally are set with a 1- or 2-year time horizon. This time line contrasts with that for objectives for population services, which are more likely to have 5- or 10-year horizons that are aligned with the national *Healthy People* targets set for a particular decade.

Process Objectives: TAAPS

The process theory component of the program theory—and specifically the organizational plan and the service utilization plan—provides a framework for stating process objectives (**FIGURE 6-1**). The *process objectives* encapsulate the essentials of the process theory that describes how the program is delivered, focusing on the activities of the program staff members or the program participant. Process objectives ought to have the following elements: Time frame,

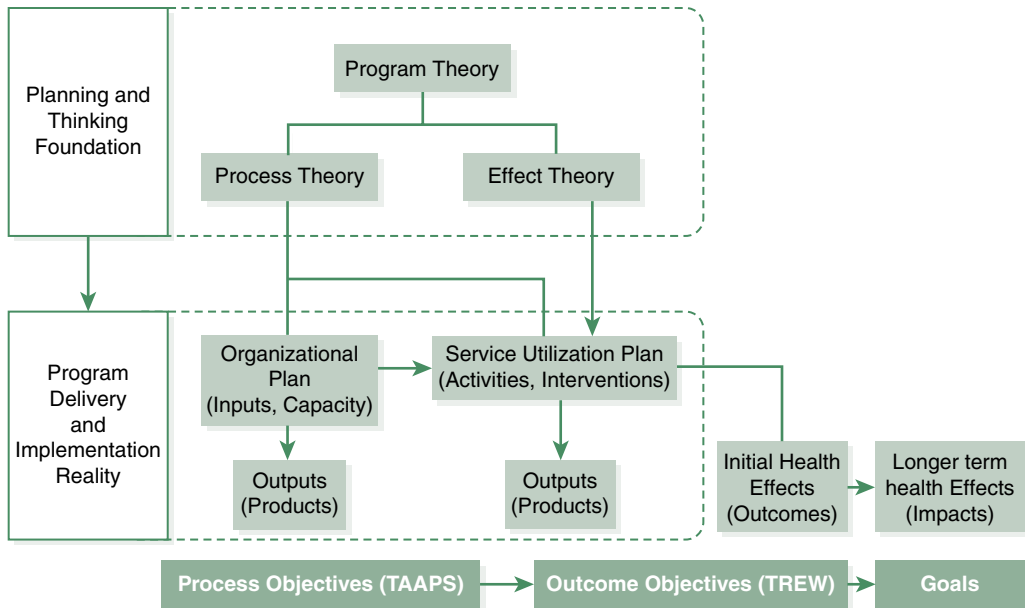


FIGURE 6-1 Using Elements of Program Theory as the Basis for Writing Program Objectives

Amount of what **Activities** done by which **Participants/program Staff** (TAAPS). TAAPS objectives can be written in the general format for writing objectives; process objectives would then state, “by when, which staff members will do what, to what extent” (**TABLE 6-1**).

TAAPS objectives focus on actions of participants or on the activities of the program staff members that don’t directly cause the effect. The organizational plan and service utilization plan provide insights into what ought to be included in each process objective, particularly for the “do what” portion. The “to what extent” portion will be determined based on past experience with the capabilities of the staff and on the amount of work to be done within the time frame. Objectives can be identified for the capacity of the infrastructure, as is commonly done in terms of personnel qualifications. Capacity objectives are best considered as objectives about the organizational plan.

Thus, a process objective might be, “By month 6, 100% of program staff will have participated in 90% of the training sessions on how to use health education modules being used in Layetteville’s i-APP Program.”

Effect Objectives: TREW

Effect objectives focus on benefits that program participants will experience as a result of receiving the program interventions. Following the formula for writing objectives, effect objectives would state, “by when, how many of which program participants will experience what type of health benefit or state and to what extent.”

Effect objectives have the following elements: in what **Time** frame, what portion of **Recipients** experience what **Extent** of **Which** type of change (TREW). *TREW objectives* can be written in the general format for writing objectives; effect objectives would then state, “After how much intervention, how many recipients will experience what extent of which type of change.” The extent refers to how much or what degree of change, which is anticipated by having received a sufficient dosage of the intervention.

The effect theory—most notably, the theories of cause/determinants, of the intervention, and of outcomes—provides the basis for stating intervention, outcome, and impact objectives, remembering that impacts are more appropriately

TABLE 6-1 Aspects of Process Objectives as Related to Components of the Process Theory, Showing the TAAPS Elements

	Organizational Plan	Service Utilization Plan	Process Theory Outputs
Objective	By when (T), how much (A) of obtaining or organizing which types of resource in what ways (A) will be done by which program staff members (PS).	By when (T), how many (A) of what types of interactions will be done (A) by which participants/program staff members (PS).	By when (T), how many (A) of what types of outputs or products will be created or finalized (A) by whom (PS).
Objective examples	By [date] (T), full funding (A) for three new computers with electronic clinical record software will be secured (A) by the program manager (PS). By [date] (T), 4 hours of training about community gang violence (A) will be provided (A) by the medical director (PS).	By [date] (T), three evidence-based interventions for improving self-management of diabetes (A) will be identified (A) by the health educator (PS).	By [date] (T), 100 brochures (A) will be distributed to women receiving genetic counseling at the clinic (A) by outreach liaisons employed by the program (PS).

T for Time; A for Amount; A for Activities; PS for Participant/Program Staff.

called goals (**FIGURE 6-2**). In most program literature, all three types of objectives are referred to as outcome objectives. The purpose of distinguishing among the three types is to ensure that, during the planning process, connections between the planned interventions and health changes are made explicit. Being explicit at this phase of the planning facilitates subsequent development of the evaluation, particularly with regard to which changes, benefits, or health outcomes should be measured. Because funding agencies generally require objectives dealing with effects, intervention objectives can be included with outcome and impact objectives.

The format for writing good objectives can be used to write objectives in terms of increasing or reducing the level of a certain outcome compared to some benchmark level. For example, “After attending 90% of the sessions, 100% of

the Layingville i-APP Program participants will have a 20% lower pregnancy rate compared to girls not participating,” uses nonparticipants as the comparison—that is, the benchmark. Another approach to writing an outcome objective is to have the “extent” parameter reflect a preferred level of achievement or a target value that the program seeks to achieve. An objective written in this way might state, “After attending 90% of the sessions, 100% of Layingville i-APP Program participants will have a 3% pregnancy rate.” Regardless of how the “extent” parameter is stated, the objectives need to reference a quantity of dosage that can be translated into a time frame, the program participants to be affected, a health outcome related to the program interventions, and a quantifiable target value for that health outcome. Clearly stated objectives that include these components serve

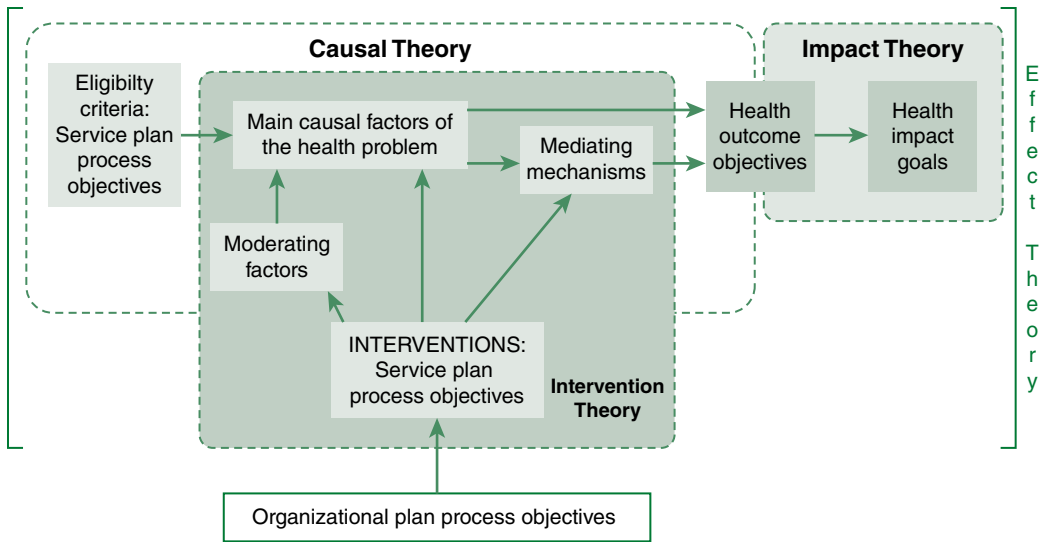


FIGURE 6-2 Diagram Showing Relationship of Effect Theory Elements to Process and Outcome Objectives

as clear guideposts in designing the evaluation of program effect.

Objectives and Indicators

One aspect of developing objectives is to consider indicators. Like so many other terms in program planning and evaluation, *indicator* has many uses and interpretations. It can refer to the “what” portion of the objective, to the variables used to measure that “what,” or to performance benchmarks used to determine the failure or success of a program. There is no easy way to distinguish among these uses or to state prescriptively that one is better than the others. It is important to be aware, however, that the results of the evaluation regarding “extent” can be influenced not only by the program’s true effect but also by the sensitivity of the measure (indicator) selected. For example, if an outcome objective concerns improvement in cognitive functioning of children with special healthcare needs, indicators of cognitive functioning might consist of a score on a standardized scale such as the Bayley Scales of Infant Development or the Denver Developmental Screening Test, or a parental report of cognitive functioning. Because the standardized scales are more sensitive and

specific than the parental report, they can detect a smaller true change.

By choosing—or at least considering—indicators when developing objectives, planners and evaluators can set reasonable target numbers for the extent of or how much change is expected, given the indicator chosen. For full-coverage programs at the population level, it may be more appropriate to think of indicators in terms of benchmarks. For example, the national standard for a healthy birthweight could be used as an indicator, as in the objectives for infants born to women in the Women, Infants, and Children (WIC) nutrition program.

Most health programs address one or more domains of health or well-being. Typically, these domains encompass physical, mental, cognitive, behavioral, knowledge, social, and financial issues. For each of these domains of health, specific variables (indicators) are used to measure the program effect on that domain. **TABLE 6-2** lists some commonly used variables for each health or well-being domain. The developers of a health program would need to select those indicators that reflect the specific health domain targeted by that program. Reliance on the effect theory and identification of the antecedent, causal, moderating, or mediating factors of the

TABLE 6-2 Domains of Individual or Family Health Outcomes with Examples of Corresponding Indicators and Standardized Measures

Outcome Domain	Examples of Indicators (Variables) to Measure Objectives
Physical health	<p><i>General measures:</i> cardiovascular fitness, weight, dental diagnosis, medical diagnosis of acute illness, medical diagnosis of chronic illness.</p> <p><i>Standardized measures:</i> normal range on laboratory tests, ICD-10, DSM-5.</p>
Mental health	<p><i>General measures:</i> motivation, values, attitudes, emotional bonding, medical diagnosis of mental illness, medical diagnosis of addiction, stress.</p> <p><i>Standardized measures:</i> CES-D, Edinburgh Postnatal Depression Scale, Daily Hassles Scale.</p>
Cognitive processes	<p><i>General measures:</i> decision making, judgments, problem-solving ability, cognitive development, cognitive impairment.</p> <p><i>Standardized measures:</i> Bayley Scales of Infant Development, IQ tests, Short Mental Status Exam.</p>
Behavior	<p><i>General measures:</i> smoking, exercise, acts of aggression, seat-belt use, food purchasing behavior, specific parenting behavior, risk-taking or risk-seeking behaviors.</p> <p><i>Standardized measures:</i> Self-Care Inventory, Child Behavior Checklist, Adolescent Sexual Activity Index, items from the Behavioral Risk Factors Surveillance Survey.</p>
Knowledge	<p><i>General measures:</i> skill, ability, performance, education of others, recall of facts, synthesis of facts.</p> <p><i>Standardized measures:</i> Smith Alcohol Knowledge Test, Self-Rated Abilities for Health Practices.</p>
Social health	<p><i>General measures:</i> marital status, social network, recreation activities, volunteerism.</p> <p><i>Standardized measures:</i> Norbeck Social Support Questionnaire, Dyadic Adjustment Scale, Life Events Scale.</p>
Resources	<p><i>General measures:</i> income, insurance source, housing situation, employment status, education level.</p> <p><i>Standardized measures:</i> Hollingshead Index, dissimilarity index.</p>

health problem being targeted may also suggest the optimal indicators of program effect.

A variety of criteria can be applied when selecting indicators. The first and foremost

criterion to consider is any indicators that are required or mandated by the funding agency. For example, the Maternal and Child Health Bureau (MCHB) of the Health Resources and

Services Administration (HRSA) requires that all grantees of Title V funds use its set of 18 performance measures, 6 outcome measures, and 8 capacity-related objectives. MCHB Title V indicators include measures such as the rate at which children are hospitalized for asthma, the percentage of women with a live birth who have had an adequate number of prenatal visits according to the Adequacy of Prenatal Care Utilization index (Kotelchuck, 1997), and the percentage of live births where the newborns weigh less than 2,500 grams (Maternal and Child Health Bureau of the Health Resources and Services Administration, n.d.). Although some of the MCHB indicators could be considered related to process, grantees must use these indicators as outcomes.

Another criterion for selecting indicators includes the requirement that the data for the indicator, if it is a variable, must be feasible to

collect. Indicators such as variables also ought to be scientifically defensible—hence the use of standardized or existing questionnaires and tools. Indicators of any type also ought to be relevant to users, such as the program managers and program stakeholders. Indicators (measures) need to be relatively easy to analyze. It is pointless to rely on an indicator (measure) that is so difficult to analyze that it is not used in program management or improvement.

Indicators may also be selected by returning to the community health diagnosis statements developed about health problems. In those statements, the health status indicator can be directly applied to the outcome objectives. For example, for each the five health problems identified as being a high priority in Bowe County, indicators or variables are used in the objectives (TABLE 6-3). TABLE 6-4 provides examples of intervention, outcome, and impact objectives for the goal of

TABLE 6-3 Bowe County Health Problems with Indicators, Health Outcomes, and Health Goals

Health Problem	Indicator of Health Problem	Health Outcome	Health Goal or Impact
Vaccination	Rates of underimmunization, by age group	Vaccine-preventable illness	Decrease vaccine-preventable hospitalizations
Presence of neural tube defect	Rates of neural tube defects and congenital anomalies	Absence of neural tube defects	Reduce current rate of congenital anomalies
Diagnosis of pregnancy during adolescence	Pregnancy rate, by age group	Diagnosis of pregnancy	Reduce child abuse related to unwanted pregnancy
Hospital admissions for gunshot wounds	Rate of admissions for gunshot injuries at local hospitals, number of police reports	Adolescent death rate due to gunshot wounds	Reduce adolescent death rate due to gunshot wounds
Diagnosis of type 2 diabetes	Incidence rate of diabetes, prevalence rate of diabetes	Rates of amputation and vision loss due to diabetes	Reduce morbidity due to chronic illness

TABLE 6-4 Effect Objectives Related to the Theory of Causal/Determinant Factors, Theory of the Intervention Mechanisms, and Theory of Outcome to Impact, Using Congenital Anomalies as an Example, Showing the TREW Elements

	Effect Objectives Related to the . . .		
	Theory of Causal/ Determinant Factors	Theory of the Intervention	Theory of Outcome to Impact
Format of objective	After how much intervention (T), what portion of recipients (R) will experience what extent of change (E) on which causes of the health problem (W)	After how much intervention (T), what portion of recipients (R) will experience what extent of change (E) on which immediate or mediating factors that lead to the health problem (W)	After how much intervention (T), what portion of potential recipients (R) will experience what extent of change (E) on which long-term or global health status indicators (W)
Example	After attending the 3 sessions (T) on environmental hazards that are teratogenic, 100% of the women (R) will have avoided (E) exposure to environmental hazards (W) during the year before becoming pregnant	After taking prenatal vitamins the year before pregnancy (T), 100% of women in the program (R) will have normal (E) blood levels of iron and folic acid (W)	After three complete cycles of the program (T), newborns in Bowe County (R) will have no (E) preventable congenital anomalies (W)

T for Timeframe; R for Recipients; E for Extent of change; W for Which type of change.

reducing birth defects, and **TABLE 6-5** provides examples of intervention, outcome, and impact objectives for adolescent pregnancy. In each example, the objectives are derived from—and thus correspond to—the intervention, causal, and impact theories.

Good Goals and Objectives

Obviously, good goals and objectives are both meaningful and useful (Patton, 1997). The objectives need to be distinctly related to either process or effect. Likewise, both process and effect objectives need to be tailored to the specific health program being planned. Thus, program planners are encouraged to adapt—rather than plagiarize—objectives from similar programs.

Each objective ought to convey only one idea so that each statement can be related to only one measure. This precludes using the word *and* in a goal or objective; each part ought to be separate goal or objective. Ideally, the objectives will be understandable to any stakeholder who might read them. Most important, objectives facilitate implementation by keeping the processes and activities in mind and facilitate direct evaluation by specifying the desired effect to be found.

Goals and objectives are often written using language that indicates a direction, such as improve or reduce. This use of a direction in the objective can be confusing and misleading, and therefore ought to be avoided. A goal of “improving birth outcomes” encompasses both reducing congenital anomalies and increasing

TABLE 6-5 Effect Objectives Related to the Theory of Causal/Determinant Factors, Theory of the Intervention Mechanisms, and Theory of Outcome to Impact, Using Adolescent Pregnancy as an Example, Showing the TREW Elements

	Effect Objectives Related to the . . .		
	Theory of Causal/ Determinant Factors	Theory of the Intervention	Theory of Outcome to Impact
Format of objective	After how much intervention (T), what portion of recipients (R) will experience what extent of change on which causes (E) of the health problem (W)	After how much intervention (T), what portion of recipients (R) will experience what extent of change (E) on which immediate or mediating factors that lead to the health problem (W)	After how much intervention (T), what portion of potential recipients (R) will experience what extent of change (E) on which long-term or global health status indicators (W)
Example	After attending two i-APP sessions (T), 100% of adolescents (R) will be able to describe three ways (E) to avoid pregnancy (W)	After 18 months of the i-APP (T), among sexually active adolescents (R) in Bowe County, less than 1% (E) will give birth (W)	After 2 years of full program implementation (T), among residents of Bowe County (R) the incidence of infant abuse (W) will be 8.0 per 1,000 (E)

T for Timeframe; R for Recipients; E for Extent of change; W for Which type of change.

birthweight, which may not accurately reflect the breadth of the program. Objectives and goals ought to be written to reflect the final rate or state of health, not the change needed to get there. A way to double-check that the program is directly responsible for the elements addressed in the objective is to ask the question, Without the program, would whatever is stated in the objective occur? Similarly, the program goals and objectives need to be reviewed for alignment with the needs, problems, and assets identified through the community health assessment.

Although this chapter presents the development of the goals and objectives as being derived from the logic model and the program theory, in actuality, discussions that develop about objectives may prompt the program planners and the

program stakeholders to revise the logic model or the program theory. Similarly, the process of selecting indicators for objectives may cause the objectives to be revised. These iterations ought to be viewed as a positive sign that ways to strengthen and streamline the health program are being identified and attempts to do so are occurring during the planning phase rather than after the program has been implemented.

Sooner or later someone will mention the acronym SMART in reference to objectives. SMART stands for five qualities of a good objective: specific, measurable, achievable, realistic, and time. *Specific* refers to the “what” that is to be achieved. *Measurable* refers to the metric or measure being used to determine whether the objective was met. *Achievable* is a reality check

ensuring that the target being set can actually be achieved or attained. *Realistic* asks whether, given the resources and conditions, it is plausible that the objective will be achieved. *Time* refers to the time frame for achieving the objective. In developing objectives, it can be helpful to reflect on the SMART qualities to ensure that the objective is good.

► Using Data to Set Target Values

All types of objectives, whether related to process theory or effect theory, have the “by how much” portion for each “what.” A critical step in developing a meaningful objective is choosing a numeric value as the target for the “by how much” portion of the objective. Despite the importance of quantifying the amount, surprisingly little literature exists about providing explanations of how to calculate the amount or about examples of calculating the amount. That gap in the literature reflects either the assumption that the knowledge is intuitive or that the calculation is totally idiosyncratic. As the following demonstrates, neither assumption holds.

For process objectives, the procedure for establishing target values for the “by how much” portion generally means using data from the organizational and marketing assessments. National standards or objectives are not always available for use as a guide. Nevertheless, professional standards can often be used, particularly for organizational plan objectives. For example, legal and professional standards have been established for minimum qualifications for personnel. These standards can be used as a starting point for setting targets—say, for the percentage of program staff members with a given certification.

For effect objectives, the target-setting process relies on the effect theory. The assumption is that as long as the objectives are consistent with the program theory and the level of programmatic effort, targets can be achieved. What is achievable, of course, depends on many factors both internal and external to the program. Having

reasonable target values directly influences the extent to which a program is perceived as successful, particularly with regard to outcomes and impacts. Consequently, the measurement of success must be scientifically credible. During the development of the objectives and their corresponding targets, planners and evaluators will want to agree on a strategy that accounts for both the program theory and any extraneous factors, and they will also want stakeholders to become involved during this crucial initial stage.

Developing a rational target-setting strategy is preferred over using mere “guesstimation.” Using a rational approach to setting targets leads to targets that are meaningful from a programmatic perspective, that are achievable to the extent that they represent an outgrowth of the program theory, and that are based on empirical data. The target-setting approaches outlined here begin with the process of establishing guidelines based on a decision framework and then move to choosing one or more relatively simple statistical procedures to yield target values consistent with the decisions. These approaches are best suited for setting target values for effect objectives. The framework described in the next section was developed by Rosenberg (2012) as an outgrowth of an effort by the MCHB to provide states with enhanced skills for program planning and evaluation.

Decisional Framework for Setting Target Values

The first and most basic element in developing a target-setting strategy is deciding how a program will be deemed successful. Success can be defined as meeting or exceeding a target, or as making meaningful progress toward the target but not necessarily meeting it. If success is defined as meeting a target, then targets will probably be chosen more cautiously than if success is defined less strictly. If program planners and staff members wish to claim success even when a target value is not achieved, then “making meaningful progress” must be quantified, in addition to setting the target value itself. For example, the efforts to reduce infant mortality could be called “successful” based on the steady decrease in the

rate of infant mortality over the past decade, despite the fact the infant mortality rate did not reach the target specified in *Healthy People 2020*. Either definition of program success is acceptable, but the definition to be used in the later program evaluation needs to be agreed upon and, more important, made explicit to all relevant parties during the planning stage.

The way in which program success is defined influences whether targets are chosen primarily according to past or baseline indicator values or whether more emphasis is placed on the values of longer term objectives for the program. This difference in perspective can have a dramatic effect on a final target value. Referencing targets to past or baseline values is typically a more cautious approach because the target values tend to be set to a level that represents a very modest improvement in the health outcome being measured—in other words, a minimum expectation for program effectiveness. In contrast, referencing targets to longer term objectives is a bolder approach, often resulting in target values that are somewhat more difficult to reach but that challenge program managers to examine the program implementation continually and to advocate for changes if necessary. Both approaches are appropriate, and a decision needs to be made about which focus is more important to the particular program under consideration.

Once program success is defined and a consensus has been reached regarding the relative importance of past, present, or future indicator values, program planners can begin developing a specific methodology for incorporating indicator values into the target. Sometimes only current data values are used in setting a target; sometimes a combination of current values and trend data is considered; and sometimes current values, trend data, and a local or national standard are all incorporated into the target-setting process. For example, if data have already been collected over time and a well-established objective or national standard specifying a long-term outcome exists, it may be important to set a target based both on the trend in the data and on the distance between the existing standard and the desired outcome.

TABLE 6-6 shows a matrix depicting combinations of patterns over time and relationships

to a long-term objective. A different target value might be selected depending on which cell is relevant to the outcome of interest. For example, if a program is being implemented for a target population that has been experiencing worsening conditions over time and that has a current indicator value far from a long-term objective, the target value might be set more cautiously than if the program is being implemented for a target population that has been experiencing gradual improvement and that has a current indicator already fairly close to the long-term target value.

Another component of making decisions about target setting is choosing which types and sources of data will be used. A wide variety of data sources are often available and appropriate for assessing health programs and measuring objectives. Ideally, multiple data sources will be used in setting target values because each source contributes slightly different information to the target-setting process. For example, one source of data might consist of police reports on the use of guns, and another source might be medical diagnoses of gunshot wounds in emergency departments. The statistics reported by each of these sources might be similar, but they might also be different. Having access to both data sets would be useful in setting a reasonable target for the rate of intentional gun violence in Layetteville.

Both the number of data sources available and the consistency of the data across these sources influence the target-setting process. **TABLE 6-7** shows the intersection of these two dimensions. For example, if many data sources are available and their data are in reasonable agreement, then arriving at a target value is relatively straightforward because it will reflect the consistent values. In such a case, similar target values would be reached no matter which data sources are used. If many data sources are available, however, but the information is inconsistent or conflicting, then decisions must be made regarding which data source should be given precedence or which combination of data sources will be used. These decisions should be based on the strengths and weaknesses of each data source, including

TABLE 6-6 Matrix of Decision Options Based on Current Indicator Value, Population Trend of the Health Indicator, and Value of Long-Term Objective or Standard

Population Trend of Indicator	Current Value of Health Indicator in Target Audience		
	Better Than Long-Term Objective or Standard	Meets Long-Term Objective or Standard	Worse Than Objective or Standard
Improving	Set target to <i>maintain</i> current level; better than the long-term objective and limits to further improvement	Set target to <i>surpass</i> the long-term objective; continuing the improving trend	Set target to a <i>better</i> level; accelerate improving trend to approach the long-term objective
No change	Set target to slightly <i>better level</i> ; better than the long-term objective, but want to see improving trend	Set target to <i>surpass</i> the long-term objective; begin improvement in trend	Set target to <i>moderately better</i> level; begin improvement in trend
Deteriorating	Set target to <i>maintain</i> current level; stop the worsening trend	Set target to <i>maintain</i> current value; stop the worsening trend	Set target to <i>maintain</i> current level or adjust it slightly downward; stop or slow the worsening trend

TABLE 6-7 Framework for Target Setting: Interaction of Data Source Availability and Consistency of Information

	One or Only a Few Sources	Many Sources
Consistent Information Across Sources	Need to consider whether the available data are of high quality. Need to consider whether it is relevant to the program and objective target being considered.	Can use any of the data sources.
Inconsistent Information Across Sources	If the one data source is markedly different from the literature, need to either change the objective or verify the data.	Need to decide which data source to use, given the strengths and weaknesses of each data source. Need to consider which data source is most relevant to the program and the objective being considered.

its sample size, data completeness, and other aspects of data quality. The goal is to integrate the data and the information in a way that permits arriving at one target value.

The choice of data source also needs to be congruent with or to correspond to the target population or audience. For example, if a target value is being developed for effect objectives for a full-coverage, population-level program with a goal of improving birth outcomes, an appropriate data source would be vital records data. In contrast, for the Layetteville i-APP program, which is a partial-coverage program geared toward a smaller intended audience with the goal of reducing adolescent pregnancy, appropriate data sources might include medical records and surveys of the women who are program recipients, as well as the county vital records data.

The data sources for setting the target may or may not be the same as the sources of the evaluation data or the data for the community needs assessment. The choice of data source for each of these program planning and evaluation activities must always correspond to the purpose for which the data will be used. Also, if different sources of data are used, then the program planners and evaluators need to agree that the different sources yield the same information.

Stratification and Object Target Values

Another factor to consider in the decision-making process is the extent to which disparities exist across or within the intended populations. During the community health assessment, some sense of the disparities in health may be evident. If the disparities exist by different income levels, race/ethnicity, or geographic location, the data may need to be stratified by those factors. Stratification is the process by which a specific characteristic of a group is partitioned into levels, called strata, in order to see variations in the subgroups across each strata. When data are not stratified, the values are simply averages that may mask very

different outcomes for different groups within the population. For example, the unstratified rate of pregnancy among 15- to 17-year-olds in a county might appear relatively close to the *Healthy People 2020* objective of 36.2 per 1,000. When pregnancy is stratified by ZIP code, however, it may become clear that the rate of adolescent pregnancy in one area is far from the national target value and much different from the rate in another area. Target values may or may not be chosen based on stratified data, but program evaluators should incorporate stratified values into their interpretation of why targets are or are not met. With respect to the program theory, the moderating factors might also be stratified. Continuing with the adolescent pregnancy example, cognitive development or family income may be variables that can be stratified as a means of understanding which subgroup of adolescents has a disproportionately higher or lower pregnancy rate.

It is not always possible to stratify data in the way that program planners or evaluators may want. Data sources related to the indicator of interest may not include data for the variables to be stratified. For example, data from emergency department records are not likely to include information on the educational level of the patient, and data from police reports are not likely to include information on the severity of the injury.

Stratification can result in having only a few individuals being grouped within one stratum; this situation poses statistical and interpretation problems. Various approaches to addressing this problem of small numbers exist; one approach is to reduce the number of strata by combining data across strata, whether the strata are years, geographic areas, or sociodemographic characteristics. For example, while it might be desirable to stratify a population or intended audience based on age, broad (rather than narrow) age strata might be defined to ensure adequate numbers in each group. This problem of small numbers is particularly challenging for programs in rural areas or for intended audiences with rare health needs or problems.

Use of Logic Statements to Develop Targets

One technique for explicitly organizing and documenting the process of setting targets is to use logic statements. These statements can be written as the decision-making process unfolds, which is a way to keep the decisions explicit and the discussions focused. Logic statements are written in an “if, then,” or “otherwise, if, then” format. For example, for integrating different data sources, a logic statement for gun violence might be as follows:

If the emergency department data and the police department data do not agree,
then the [one or other of the data sources] will be given precedence in setting the target value.

To integrate different types of data about gun violence, a logic statement might be something like the following:

If trend data for gun violence show steady improvement, but the current value is still far from a long-term objective,
then the target value will be set to reflect an increase in the rate of improvement;
otherwise, if the trend in gun violence shows steady improvement and the current rate is already close to a long-term objective,
then the target value will be set to reflect a continuation of the existing rate of improvement;
otherwise, if no trend data are available,
then the target value will be set to reflect an improvement in the current value of X percent.

Sets of such statements can be drafted for each indicator about which decisions are being made, such as data sources, data consistency, data types, existing or perceived disparities, and resource availability. These statements should incorporate information obtained during the

community health assessment, along with input from stakeholders. Developing the if–then statements becomes part of the program planning process and helps create realistic objectives.

Although some target-setting decisions can be applied to all the program objectives, other decisions may vary depending on the objective. For each different health outcome, the objective may rely on differing pools of data sources, and have differing trends over time and differing patterns of disparities. In addition, objectives for population-based and full-coverage programs will require a target-setting strategy that differs from that developed for a direct services-level program that addresses a specific health domain within individuals. Program planners must also recognize that target setting is an iterative process, taking place over the life of a program. Rarely are target values set beyond 1 year, making it necessary to revisit the targets on an annual basis for health programs that are institutionalized or that are planned to last for a longer period of time.

Options for Calculating Target Values

Many options for calculating target values are available, each of which may be appropriate in some circumstances but not in others. The fact that target values for any one health problem can be calculated in so many ways underscores the importance of establishing a consensus on the underlying logic reflected in the if–then statements that lead to a particular value. Ten options for calculating target values are described here. The calculations can easily be done using a calculator or a spreadsheet.

FIGURES 6-3, 6-4, and 6-5 show the calculations for a program whose goal is to reduce adolescent pregnancy in Bowe County. For this program, outcome objectives are needed regarding the extent to which that goal is being met.

In planning some health programs, very limited information may be available on which to base the calculations for some target values. This can be the case, for example, for innovative programs,

	A	B	C	D	E	F	G	H	I	J	K	L
1	Current Value	% Change	Long Term Target Value	Long Term Number of Years	Population at Risk							
2						Option	Description	Formula	Target Value	Absolute Change	Percent Change	
3	37.6				16,556	1	Default	(none)	37.6	0	0.0%	
4												
5	37.6				16,556	2	Use statistically significant change (approx.)	$A5 - \text{SQRT}(2 * (A5 * (1000 - A3)) / (E5 * 2))$	33.5	4.1	-10.9%	
6												
7	37.6	-0.02			16,566	3	Use current trend as change desired	$(B7 * A7) + A7$	36.8	0.8	-2.0%	
8												
9	37.6	-0.04	30	5	16,556	4	Use meeting long-term objective					
10												
11							then, Yr 1	$(B9 * A9) + A9$	36.1	1.5	-4.0%	
12							then, Yr 2	$(B9 * J11) + J11$	34.6	1.5	-4.0%	
13							then, Yr 3	$(B9 * J12) + J12$	33.2	1.4	-4.0%	
14							then, Yr 4	$(B9 * J13) + J13$	31.9	1.3	-4.0%	
15							then, Yr 5	$(B9 * J14) + J14$	30.6	1.3	-4.0%	
16							Total improvement			7.0	18.6%	

FIGURE 6-3 Calculations of Options 1 Through 4 Using a Spreadsheet

	A	B	C	D	E	F	G	H	I	J	K	L
1	County	Size of Population at Risk	Number of Teen Births	Teen Birth Rate per 1000		Option	Description	Formula	Current Value	Target Value	Absolute Change	Percent Change
2	O	793	9	11.3					376			
3	P	2785	66	23.7			Not considering sample size:					
4	Q	859	22	25.6	5	5	Mean rate	Average (D2:D11)	376	33.8	3.8	-10.2%
5	R	2205	64	29.0	6	6	Median of the rates	Midpoint between County S and T: D6+(D7-D6)/2	376	31.0	6.6	-17.4%
6	S	1338	40	29.9								
7	T	994	32	32.2								
8	Subtotal/Rate for Approx. 50% of Population	8974	233	26.0			Considering sample size:					
9	U	708	24	33.9	7	7	Rate for "best" 50% of population	50% = B14 x 0.50 = 8,278. Sum sizes until reach 8278, O, P, Q, R, S, T				
10	V	2664	106	39.8				(C8/B8)*1000	376	26.0	11.6	-30.9%
11	W	302	15	49.7								
12	Subtotal/Rate for Approx. 75% of Population	12,648	378	29.9	8	8	Rate for 75% of population (Countries P to V)	75% = B14 x 0.75 = 12,417. Sum sizes until reach 12,417, P, Q, R, S, T, U, V, W				
13	X	3908	244	62.4				(C12/B12)*1000	376	29.9	7.7	-20.5%
14	Total/Overall Rate	16,556	622	37.6								

FIGURE 6-4 Calculations of Options 5 Through 8 Using a Spreadsheet

	A	B	C	D	E	F	G	H	I	J	K	L
1	Strata	Size of Population at Risk	Number of Teen Births	Teen Birth Rate per 1000		Option	Description	Formula	Current Value	Target Value	Absolute Change	Percent Change
2	Poverty											
3	Yes	2533	148	58.4		9	Rate for best strata	(none)	37.6	33.8	3.8	-10.1%
4	No	14,023	474	33.8								
5	Total	16,556	622			10	Overall rate based on strata specific rates	Step 1. Use Option 2 formula per strata				
6								10% decrease for Poverty = $(-0.1 * 16) + 16$	58.4	52.6	5.8	-10.0%
7								2% decrease for No Poverty = $(-0.02 * 17) + 17$	33.8	33.1	0.7	-2.0%
8												
9								Step 2. Calculate final target, weighting by % of population in each group				
10								$(J6*(B3/B5))+(J7*(B4/B5))$	37.6	36.1	1.5	-4.0%

FIGURE 6-5 Calculations of Options 9 and 10 Using a Spreadsheet

programs addressing rare health problems, and programs that are highly tailored to the location in which they are delivered. In such cases, only one piece of information from the community health assessment may be relevant—namely, a numeric value for the current level of the health problem. Four options for calculating a target value can be used in developing the outcome objective under these conditions. **TABLE 6-8** summarizes the conditions under which each of the 10 options would be best and outlines the advantages and disadvantages of each option.

Option 1 assumes that no change in the health problem will occur because of the program. This is equivalent to accepting the current level or value. As a default position, it provides a starting point, particularly if the health program is in its first year and no data are available on how much change can occur. This option may also be appropriate for health programs that are mature and are seeking to maintain the current value because it is already at an acceptable, healthy level. The formula is as follows:

$$\text{Target value} = \text{current value}$$

As an example, the target value for the birth rate per 1,000 female adolescents aged 15 through 17 is stated in the objective as 37.6 per 1,000. This target value is used in the program objective as the “how much” value.

Option 2 identifies a value that, when compared to the current value, results in a statistically significant improvement. This option would be appropriate if the data source is credible, the program has a rigorous intervention, or policy makers need to be convinced that the program is a worthwhile investment. Because change may happen by chance and not just because of the health program, planners must be able to argue that the amount of change is greater than would occur by chance alone and hence is attributable to the program. An approximate *Z*-test can be used to derive the amount of change needed to be statistically significant. Typically, the significance level is set at $p < .05$, meaning that the probability of reaching that target by chance alone is less than 5 in 100, or 5%. The .05 significance level translates

into a *Z* score of 1.96, which is used in the formula to estimate the target value. The formula is quite complex but has been simplified somewhat here so that it can be used with a spreadsheet:

$$\text{Target value} = \text{Current value} - \left(\sqrt{\frac{2 \times \text{current value} \times (\text{multiplier} - \text{current value})}{\text{population at risk}}} \times 1 \right)$$

This formula assumes that the current value is an integer—that is, it is a percentage or a number per 1,000, 10,000, or whatever the usual units are for reporting the indicator. The multiplier, then, is that unit value. In the adolescent birth rate example, the current value is 37.6 and the multiplier is 1,000. In addition, the formula is written so that the target value will be less than the current value. If improvement in an indicator translates into a target value that is larger than the current value, then the minus (–) sign in the formula will change to a plus (+) sign.

Any test for statistical significance is very sensitive to sample size. For a full-coverage program using population data, the number of recipients is typically large, so even a modest improvement in the current value can lead to a statistically significant result. If the number of participants or recipients is small, as is likely in a partial-coverage program using data for only the program recipients, it is likely that an unrealistically large target would be needed to achieve statistical significance.

In the adolescent pregnancy example, the statistical test is based on more than 16,000 adolescents, so a reasonable target value of 33.5 would result in a significant result. In contrast, suppose that this method were to be used in a program serving only 500 adolescents. In this case, a target value of 14.0 would be required to result in statistically significant improvement—clearly an impossible target to meet. Statistical testing, then, should really be used as an aid to understanding what a reasonable target value might be rather than for determining the target value per se.

Option 3 is to select a desired percentage change in the health problem and convert the percentage into a definitive target value. This

TABLE 6-8 Summary of When to Use Each Option

Option	Description of Option	Type of Program for Which the Option Is Ideal	Advantages of Option	Disadvantages of Option
1	Default, no change	Mature, stable program	Does not require historical data	Does not require improvement
2	Change based on results of statistical test	Population-based or program with large numbers of recipients	Supports argument that improvement was more than by chance	Sensitive to sample size; may result in unreasonable target; requires some statistical knowledge
3	Percentage change in health problem based on current trend, literature, or hopeful guess	Stable program; stable target population	Very straightforward and easy to understand; can easily take into account trend data if available	Must know the trend
4	Use existing benchmark or standard to project target values for several years	Program must show improvement	Comparable programs can be compared	Requires existence of long-term objective or standard; requires long-term program
5	Mean rate across geographic areas	Population-based	Easily understood	Requires having data for each area
6	Median rates across geographic areas	Population-based	Easily understood	Requires having data for each area
7	Overall rate for best 50% across geographic areas	Population-based or multisite	Takes into consideration the best and worst values in the target population; moves entire target population to an achievable value	Requires having data for each area; may be more difficult to understand; overlooks sample size

(continues)

TABLE 6-8 Summary of When to Use Each Option (*continued*)

Option	Description of Option	Type of Program for Which the Option Is Ideal	Advantages of Option	Disadvantages of Option
8	Overall rate for best 75% across geographic areas	Population-based or multisite	Takes into consideration the best and worst values in the target population; moves entire target population to an achievable value	Requires having data for each area; may be more difficult to understand
9	Rate for best stratum using sociodemographic groupings	Population-based or diverse target audience with evidence of disparities	Takes into consideration the best and worst values in the target population; moves entire target population to an achievable value	Requires having data for each group; may be more difficult to understand
10	Overall rate based on differential targets for each stratum	Population-based or diverse target audience with evidence of disparities	Program must show improvement; more intense program intervention aimed at group with the most need for improvement	Requires having data for each group; may be more difficult to understand

option is the most straightforward approach and can be understood intuitively by stakeholders. It can be used with health programs that are situated at any level in the public health pyramid and in any health domain. The percentage change used in the formula can be chosen based on information gained from published literature, or it may merely be a hopeful “guesstimate.” The formula is as follows:

$$\text{Target value} = (\% \text{ change desired} \times \text{current value}) \pm \text{current value}$$

If trend data exist for the health outcome, then the percentage decrease (or increase) can be refined based on past and recent experience. The percentage change can be chosen to reflect

either a continuation of the observed trend or a change in the trend (i.e., either an acceleration of improvement or a slowing of deterioration, depending on the health outcome of interest).

In the example of the adolescent birth rate, trend data indicate an average 2% annual decrease. Using this percentage in the formula for option 3, the target value for the birth rate per 1,000 female adolescents is 36.8 per 1,000 (see Figure 6-3). If program planners decide that a 4% decrease is more appropriate—that is, if they assume that the program can accelerate improvement—the target value for the birth rate per 1,000 female adolescents would be 36.1. The target value chosen for this calculation is then used in the program objective as the “how much” value.

This exercise reveals that, although a 4% decrease in adolescent births may require considerable programmatic resources to achieve, the reduction in the rate may be barely noticeable. It may be useful to repeat the calculation with slightly different percentage changes and consider which elements in the organizational plan and service utilization plan would need to be modified to achieve those other percentage changes.

Option 4 is used when programs are ongoing, are multiyear projects, or are expected to have long-term effects. For such programs, the time-frame portion of the objective may be several years into the future. In this case, it becomes necessary to have annual target values that cumulatively reach the desired long-term target value. Essentially, the total amount of change is dispersed across the time period for the program. For this reason, the target values for each year will be affected by the anticipated length of the program and the starting or current value. *Option 4* can be used for programs at any level of the public health pyramid, but it is appropriate only for objectives related to a long-term goal. To use *option 4*, health program planners' first decision is to select an existing benchmark or standard, such as a *Healthy People* objective that identifies the desired target value for the health problem that is to be achieved over the long term.

Calculating annual target values requires first estimating the amount of annual change needed to get close to the long-term target. This annual percentage change is then used in calculations like those in *option 3* to find the target value for each subsequent year. The following set of formulas is used in sequence to carry out *option 4*:

$$\text{Annual \% change} = [(\text{Long-term objective} - \text{current value}) / \text{current value}] / \text{number of years}$$

$$\text{Next-year target value} = (\text{Annual \% desired change} \times \text{current value}) \pm \text{current value}$$

$$\text{Subsequent-year target value} = (\text{Annual \% desired change} \times \text{past year value}) \pm \text{past year value}$$

As shown in Figure 6-3, an annual 4% decrease results in an adolescent birth rate of 30.6 per 1,000 at the end of 5 years, for an overall decrease of 7 births per 1,000. This decline represents an 18.6% decrease in the birth rate among adolescents. In this example, a rate of 30 births per 1,000 was the long-term objective; thus, using the method described here, the final target value was not exactly met because the rate of improvement was maintained at 4% each year. To reach the long-term objective, the rate of improvement would have to increase slightly each year. The question for discussion among the planning team, however, is whether a 4% decrease every year for 5 years is possible for the program, and whether the 18.6% decrease over 5 years will be acceptable to funding agencies and other stakeholders. Program planners must also consider whether the change can be identified using the methods currently chosen for use in the effect evaluation.

Options 5 through 10 are relevant for population-based and multisite programs when the data can be stratified, either by geographic area or by some characteristic such as age, race/ethnicity, or income. The adolescent pregnancy example is a population-based program, using data from all high schools in Bowe County as well as data on whether the adolescents' family income is below or above the poverty level. When stratification is used for target setting, the planners often assume that some sites may already have reached a very desirable level and therefore would not be expected to improve dramatically when the program is implemented. A corollary is that some sites will likely be drastically far from any target that might be set, which means they must make radical improvements to reach any reasonable target. The extent to which a site may already be at an ideal level warrants attention from the planning team and ought to be reflected in the logic statements and the subsequent decisions about selecting target values.

Option 5 sets the target value as the mean of the rates across the sites, and *option 6* sets the target value as the median of the rates across the sites. Options 5 and 6 are likely to give very similar target values, especially if the rates of

the health outcome across the sites are normally distributed. Conversely, if there is not a normal distribution, they may not yield similar values. In the adolescent (females aged 15–17 years) birth rate example (see Figure 6-4), the county birth rates range from 11.3 per 1,000 adolescent females to 62.4 per 1,000 adolescent females, with the mean of all high school rates being 33.8 and the median being 31.0. A disadvantage of these two options is that they do not take into account the differing sizes of the target population in each group, such as in each county or in each clinic. If the sites have very different adolescent birth rates and target population sizes, and if those data are not normally distributed, then options 5 and 6 may not be the optimal approaches.

Options 7 and 8 take into account the population sizes in the area targeted by the program. The overall rate of 37.6 births per 1,000 adolescent females for all schools combined (the value used in options 1 through 4 as the current value) is the mean for the whole population. Nevertheless, because it combines the data for all schools, it obscures the school-by-school information. Options 7 and 8, by contrast, also calculate overall current values; rather than using all schools, however, each option uses only a portion of the population with the “best” outcomes. Options 7 and 8 are based on the idea that the rate achieved by a certain portion of the target audience should be reachable by the whole target audience, so the target value ought to be set based on that existing rate.

Options 7 and 8 use what is called a “pared means method” (Kiefe et al., 1998). This approach reinforces the idea that the target value for a program should aim to move the entire target population to a value already achieved by a portion of the target population. In other words, for options 7 and 8, the target value for the objective would be for the adolescent birth rate in the schools to improve to match the birth rate already achieved by the schools encompassing 50% or 75% of adolescents.

The difference between options 7 and 8 lies in the proportion of the target population that is used to calculate the target value: option 7 is based on 50% of the target population, whereas

option 8 is based on 75% of the target population. The pared means method can actually be used with any proportion of the target population. The higher the proportion, the easier it will be to reach the target; the lower the proportion, the more difficult achieving that the target will be. Choosing 50% means that half of the target population has already achieved the target but the other half will have to improve; choosing 75% is a more conservative approach because improvement will have to occur in only 25% of the population.

Continuing with the example, to calculate the target value according to option 7, program planners would take as many schools as necessary to incorporate 50% of all the adolescents in the 10 schools that have the lowest (best) birth rates. They would then calculate the overall birth rate for this subset of schools. In this example, the calculation must include 6 of the 10 schools and thus include 50% of female adolescents. The calculation is as follows:

$$\text{Target value} = \frac{\text{Number with the health outcome in the top 50\%}}{\text{Number in target population in the top 50\%}}$$

Using these data gives a target value of 26.0 births per 1,000, which is a 30.9% decrease from the overall, current rate of 37.6. In comparison, if the counties that have 75% of the adolescents are used in the calculation (option 8), then the target value is 29.9, or a 20.5% change.

Options 9 and 10 are examples of approaches to using stratified data (see Figure 6-5). If data are available on the health status or rates of groups within the target population, then it is possible to use those rates to calculate target values for those groups. Option 9 is simply an extension of the pared means method used in options 7 and 8; it uses the “best” rate of the two groups as the overall target. In contrast, option 10 starts with two separate targets, based on the two strata, by choosing different percentage decreases or increases for each. The different percentages chosen may reflect a more intense programmatic effort aimed at the group with the most urgent need for improvement. Thus, distinct,

stratum-specific targets are calculated, but these can then be combined into a single target value for the whole population by calculating an average weighted by the size of the population in each group. The formulae for this calculation follow:

$$\text{Target value}_{\text{Group 1}} = (\% \text{ change desired} \times \text{current value}) \pm \text{current value}$$

$$\text{Target value}_{\text{Group 2}} = (\% \text{ change desired} \times \text{current value}) \pm \text{current value}$$

$$\text{Overall target} = (\% \text{ of population in group 1} \times \text{target value}_{\text{Group 1}}) + (\% \text{ of population in group 2} \times \text{target value}_{\text{Group 2}})$$

An evidence-based approach to establishing objective targets can be determined based

on information in published evaluations of the intervention. The studies are reviewed during the planning stage and include statistics that indicate how much difference the intervention made for the experimental group. By utilizing the size of the effect found in those evaluations of the intervention, the amount of possible improvement can be estimated. Similarly, the effect size statistic, which is reported in meta-analyses, can be used to estimate a target value, albeit with the help of a statistician.

In summary, a variety of techniques can be used to calculate the target value to be used in the effect objectives. Each calculation technique results in a different value (TABLE 6-9). In the example cited here, the potential target value for the rate of births to adolescents ranges from a low of 26.0 births per 1,000 using option 7 to a

TABLE 6-9 Range of Target Values Derived from Options 1 Through 10, Based on the Data from Figures 6-3 Through 6-5

Option	Description	Resulting Target Value
1	Default, no change, overall rate	37.6
2	Result of statistical testing	33.5
3	Percentage change in health problem: based on trend data	36.8
4	Use existing benchmark or standard to project target values for several years: first-year target	36.1
5	Mean of rates across geography/sites	33.8
6	Median of rates across geography/sites	31.0
7	Overall rate for best 50%	26.0
8	Overall rate for best 75%	29.9
9	Rate for "best" stratum (i.e., adolescents not living in poverty)	33.8
10	Overall rate based on stratum-specific rates	36.0

high of 37.6, which is the current value, using option 1. This range of possible and reasonable target values underscores the importance of having a decisional framework for target setting, including developing explicit logic statements, to define realistically what constitutes the success or effectiveness of a health program.

Which target value is ultimately chosen depends on the program theory, the availability of resources, and the strength of the intervention. Although options 2 and 5 through 10 are best suited to population-based programs, they can be adapted to very large programs at the direct services and enabling services levels of the public health pyramid. To use these options, sufficient data must be available for each site, and enough sites must have reasonable numbers in the groups. Options 3 and 4 are straightforward and can be used for any program.

► **Caveats to the Goal-Oriented Approach**

All aspects of planning, implementing, and evaluating a program are open for critique and reflection. Although goals and objectives may seem noncontroversial, at least three caveats highlight the need to be self-critical during these processes: the tenuous effectiveness of using objectives to guide work, the need for spontaneity, and the messy interface of objectives and performance measures.

First, the high degree of emphasis placed on having objectives—both process and effect—warrants some reflection on its history and a cursory review of the evidence of the usefulness of objectives. The concept of management by objectives (MBO) was developed and popularized in the 1950s by Peter Drucker, a management scholar, based on a theory of goal setting as being motivational for workers. MBO was subsequently adopted by the Nixon administration (Dahlsten, Styhre, & Willander, 2005), and it has since become an

entrenched expectation, particularly in federal and state governmental agencies. The purpose of MBO is to motivate individuals to work toward a common organizational goal, thereby keeping the organization focused and internally well coordinated.

Since its introduction, relatively little research has focused on the effectiveness of MBO. Poister and Streib (1995), in a survey of governmental agencies, found that MBO is widely used but inconsistently implemented. Dinesh and Palmer (1998) commented that MBO is effective but only when it has been implemented as intended. Specifically, they contended that MBO is unsuccessful in practice because it is partially implemented, and the human relations aspect of MBO has been forgotten over the years. As a consequence, MBO is seen as an older fad (Gibson & Tesone, 2001), albeit with some potential relevance for the present. The use of objectives as the basis for decision making, whether for a program or an organization, must be complemented and supplemented with attention to gaining staff acceptance of and support for the objectives, deploying human resources in a manner that supports the activities necessary to achieve the objectives, and using a reward system that makes achieving the objectives meaningful.

Second, being guided by goals and objectives is a logical, linear systems approach to planning a health program. By contrast, the growing body of scientific literature on complexity theory suggests that, in many workgroup settings, greater flexibility and spontaneity lead to more productive work teams. If a health program is being designed and provided in response to community needs and a controlled evaluation of the program is deemed unnecessary, then allowing for self-organizing teams could lead to a “better” program. This approach certainly will not be feasible for health programs that are, say, federally funded. Nonetheless, health program planners and managers may find situations in which less control—in other words less reliance on objectives—leads to a better program.

The list of objectives, once created, is too often forgotten and becomes disconnected from

the ongoing program oversight. This is more likely to be the case for programs that emerge in response to community demands or are provided inconsistently. For the most part, in health care, health outcome objectives are linked to performance measures, which are used for quality improvement. The relationships among objectives, performance measures, and program monitoring are discussed in detail elsewhere in the text.

► Across the Pyramid

At the direct services level of the public health pyramid, process objectives are likely to focus on how providers interact with program participants and how the program supports those providers in their involvement with the program. Effect objectives for programs at the direct services level will focus on individual client behavior or health status change. Setting targets for direct services programs may involve translating national objectives into local program objectives. Although national targets may or may not be appropriate for local programs, the national targets need to be considered at least as an accepted benchmark or goal.

At the enabling services level, in addition to the foci at the direct services level, process objectives are likely to include a focus on the involvement of community resources in the program, as well as emphasis on interagency collaboration and cooperation. Effect objectives for the enabling services level are likely to address changes in the behavior or health status of families and other aggregates, such as students in a school or residents of a public housing project. Setting targets for enabling services can be more challenging because national or state data regarding the problem being addressed likely will not exist. For enabling services, past experience, experiences of similar programs, and data from the community needs assessment may be the only data available for health program planners to use with a rational approach to setting the target numbers.

At the population services level, process objectives will need to include an emphasis on

the coordination of efforts necessary to implement the health program, and on the garnering of adequate and appropriate resources to provide a population-based health program. The effect objectives can have either an outcome or impact focus, and the “who” portion will consist of the community or a specific population. Processes for setting targets for population-based services—particularly those provided to state and metropolitan populations—will draw heavily on national data.

At the infrastructure level, process objectives will dominate. The infrastructure, by virtue of its nature, emphasizes developing and sustaining an organization and obtaining and managing the resources needed to implement a health program. Nonetheless, effect objectives can be written in relation to the infrastructure, most probably about the effectiveness and efficiency of services. For example, Allison, Kiefe, and Weissman (1999) proposed using a pared means benchmark method to arrive at a target value for the best-performing physicians in terms of patient outcomes. Effect objectives may also apply more directly to the infrastructure itself. For example, effect objectives might address impacts from and educational training for staff, or outcomes from employee screening programs.

► Discussion Questions and Activities

1. The organizational plan and the service utilization plan may include many elements and processes. What would you use as criteria for developing a set of objectives about the process theory? Would you set targets for process objectives?
2. For effect objectives at each level of the public health pyramid, which sources of data might be commonly used for establishing targets?
3. Imagine that you have been asked to explain to your colleagues in 10 minutes how to set targets for program objectives. Develop an outline of the steps involved.

4. Which of the 10 options for setting targets would be best suited to developing targets for the other four health problems in Layetteville and Bowe County? Provide some rationale for your choice. Try using your choice of target setting based on the following data: the adult immunization rate is 30%, with 25 hospitalizations per 1,000 adults for influenza-related pneumonia; for children younger than 1 year old, the incidence of maltreatment is 9.4 per 1,000 children (University of California at Berkeley Center for Social Services Research, http://cssr.berkeley.edu/ucb_childwelfare/RefRates.aspx); 1.2 adolescents are discharged with a diagnosis of gunshot wounds for every 1,000 hospital admissions; the neural tube defects rate is 18 per 100,000 (National Center for Health Statistics); and new diagnosis of diabetes is 7.5 per 1,000 (Centers for Disease Control and Prevention).

► Internet Resources

Taxonomy of Educational Objectives

Bloom's Taxonomy (Bloom, B., et al., [1956]. *Taxonomy of educational objectives: Handbook I: Cognitive domain*. New York, NY: Longman) is a classic and worth knowing if the intervention relates to knowledge: www.personal.psu.edu/bxb11/Objectives/ActionVerbsforObjectives.pdf

Rapid BI

This business-oriented page on writing SMART objectives has some useful tips: <http://rapidbi.com/created/WriteSMARTobjectives/>

References

- Allison, J., Kiefe, C. I., & Weissman, N. W. (1999). Can data-driven benchmarks be used to set the goals of *Healthy People 2010*? *American Journal of Public Health*, 89(1), 61–65.
- Dahlsten, F., Styhre, A., & Willander, M. (2005). The unintended consequences of management by objectives: The volume growth target at Volvo Cars. *Leadership and Organization Development Journal*, 26, 529–541.
- Dinesh, D., & Palmer, E. (1998). Management by objectives and the balanced scorecard: Will Rome fall again? *Management Decision*, 36, 363–369.
- Friedman, V. J., Rothman, J., & Withers, B. (2006). The power of why: Engaging the goal paradox in program evaluation. *American Journal of Evaluation*, 27, 201–218.
- Gibson, J. W., & Tesone, D. V. (2001). Management fads: Emergence, evolution, and implications for managers. *Academy of Management Executive*, 15, 122–133.
- Kiefe, C. I., Weissman, N. W., Allison, J., Farmer, R., Weaver, M., & Williams, O. D. (1998). Identifying achievable benchmarks of care: Concepts and methodology. *International Journal for Quality in Health Care*, 10(5), 443–447.
- Kotelchuck, M. (1997). Adequacy of prenatal care utilization. *Epidemiology*, 8, 602–604.
- Maternal and Child Health Bureau of the Health Resources and Services Administration. (n.d.). National Outcome Measures. Retrieved from, <https://mchb.tvisdata.hrsa.gov/PrioritiesAndMeasures/NationalOutcomeMeasures>
- Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.
- Poister, T. H., & Streib, G. (1995). MBO in municipal government: Variations on a traditional management tool. *Public Administrative Review*, 55, 48–57.
- Rosenberg, D. (2012). Performance and outcome measurement: Strategies for target setting. Retrieved from www.amchp.org/Calendar/Conferences/amchp-conference/Handouts2012/Documents/B01%20-%20Performance%20and%20Outcome%20Measurement%20-%20Strategies%20for%20Target%20Setting.pdf

SECTION IV

Implementing and Monitoring the Health Program

CHAPTER 7

Process Theory for Program Implementation

Implementation of a health program is done ideally in a manner consistent with the program theory and the purpose of the health program. Such implementation requires both acquiring adequate resources and then ensuring that those resources are used appropriately. The amount of effort needed at each stage of the planning and evaluation cycle varies throughout that cycle. Implementation of the health program requires the most and longest sustained effort of all the phases of a health program (**FIGURE 7-1**).

This chapter introduces the logistics associated with managing a health program, with special attention paid to general managerial issues. These logistics fall within the organizational plan and the services utilization plan portions of the process theory (**FIGURE 7-2**). The operational logistics of implementing a program tend to be straightforward but do require attention to ensure proper implementation of the health program.

Health programs are projects that can be viewed as miniature organizations. In the management literature, the organizational plan and the services utilization plan would be considered elements of the tactical plan. A common frame of reference for thinking about organizations and health programs is to consider inputs, throughputs, outputs, and

outcomes (Turnock, 2009). In the process theory, various inputs into both the organizational plan and the services utilization plan are specified. In addition, specific outputs of both the organizational plan and the services utilization plan are expected. Distinguishing between the inputs and outputs of these plans aids in acquiring the appropriate resources and in being able to communicate both programmatic needs and successes. The chapter concludes by relating the various process theory elements to the logic model format.

► Organizational Plan Inputs

The organizational plan encompasses the program inputs and resources, as well as the way in which those resources are organized. The type and amount of resources required for a health program vary with the interventions to be used. Nonetheless, the expertise of personnel, the characteristics of the intended audience, and the degree of attention to acquiring and managing resources all affect the potential for having a successful program. The organizational plan objectives serve as a

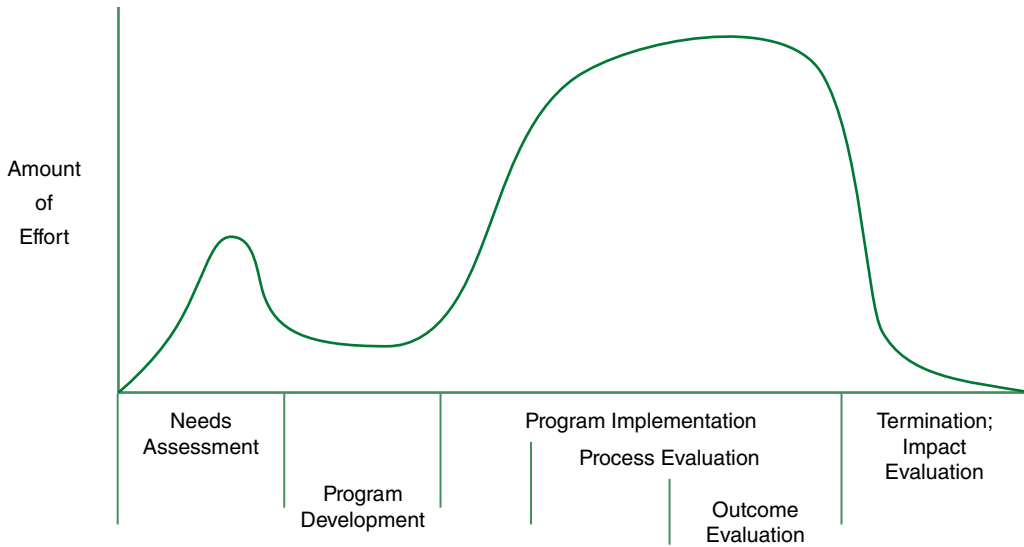


FIGURE 7-1 Amount of Effort Across the Life of a Health Program

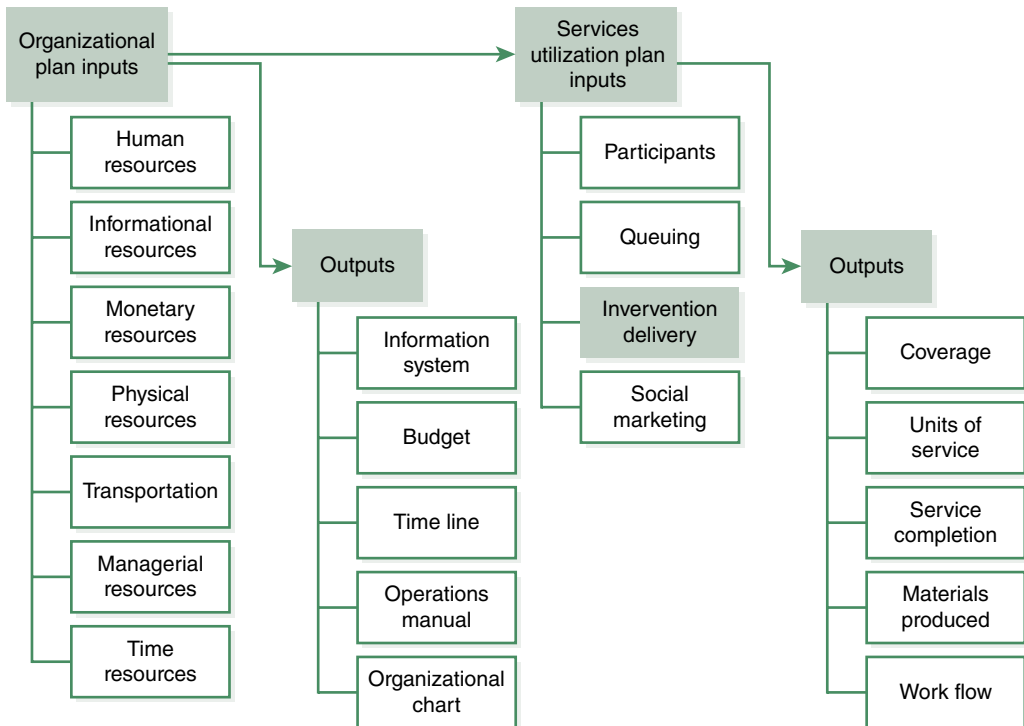


FIGURE 7-2 Diagram of the Process Theory Elements Showing the Components of the Organizational Plan and Services Utilization Plan

guide for the most critical organizational plan activities in implementing a health program. Many aspects of the organizational plan will not be in the organizational plan objectives, yet they still need to be addressed. This section presents an overview of the key inputs and outlines the rationale for considering them as key.

Human Resources

Human resources encompass the quantity and quality of personnel needed to carry out the program, in terms of their expertise, experience, and capabilities. Human resources come at a cost, of course, and personnel costs are almost always the largest portion of any

program budget. The dollar cost of personnel includes not only wages (amount paid hourly) or salaries (amount paid monthly), but also fringe benefits as a percentage of the wage or salary. Estimating the dollar cost of personnel is a rather straightforward arithmetic problem, as explained in Chapter 8 in the discussion about budgeting.

Licensure and Statutory Regulation of Health Professionals

TABLE 7-1 indicates whether individuals from each of the health disciplines are subject to legal and

TABLE 7-1 List of Health Professionals with a Summary of Typical Legal and Regulatory Considerations

Health Discipline	State Licensure Required	State Regulation of Scope of Practice	Professional Certification Exists
Dentistry	Yes	Yes	Yes
Community health	No	No	Certificate programs
Diet and nutrition	Yes, as registered dietician	Yes	Yes
Health administration	No (except for long-term care administration)	No	Yes
Health education	No	No	Yes
Industrial hygiene	Yes	Yes	Yes
Medicine	Yes	Yes	Yes, for sub-specialties
Nursing	Yes	Yes	Yes, for sub-specialties
Physical therapy	Yes	Yes	Yes
Social work	Varies	Yes, for those with licensure	Yes

state statutory regulation with regard to their practice. The legal parameters of practice for each of the licensed health disciplines ought to influence who is hired into which position. Some familiarity with the state regulations of scope of practice is important for matching the program needs with the qualifications of program personnel. In addition, for some health disciplines, professional associations oversee certification of individuals. For some health programs, it may be wise or important to ensure that the program services are provided by individuals with specialty certification, which serves as an indicator of advanced or specialized knowledge and skills.

Training for Program Delivery

Training staff members and volunteers for their roles in the program implementation is a key aspect of human resources management. Training costs include both staff time for this activity and the trainer's time. In addition, costs are incurred for the materials used during training, such as handouts, equipment, or audiovisual materials. These costs may mount quickly, but such training is nevertheless necessary for ensuring the standardized delivery of program interventions. Without this standardization of the intervention through training, the actual intervention delivered can deviate from what was intended by the program, depending on the personal preferences of the individuals providing the intervention. In other words, training helps align the theory-in-use and the espoused theory with the effect theory. This has serious implications for achieving the desired outcomes and subsequently for ensuring the long-term sustainability of the program.

Training also helps maintain the morale and self-efficacy of the staff members. Receiving training contributes to program staff members feeling valued, trusted, and capable of providing the interventions. These feelings influence the sense of having a higher ability to carry out the intervention as designed—in other words, they promote greater task self-efficacy. This is an especially important ingredient in the delivery of a

complex intervention to an intended and recipient audience that is resistant or difficult to change.

A separate consideration with regard to human resources is workforce cultural diversity. As health programs are increasingly intended for audiences with specific cultural and racial characteristics, it becomes important that program staff members are not only culturally competent with regard to the intended audience but—ideally—are members of the intended audience. While having program personnel who are culturally diverse may help in the delivery of the intervention, it can sometimes lead to other problematic issues among staff members. Program managers must be attentive to signs of interracial or cross-cultural tensions among program personnel and address these issues as soon as they surface.

Volunteers as Human Resources

Community involvement in the development and implementation of health programs can occur through the establishment of advisory committees and boards, councils, or consortia. Through such groups, community members have a formal venue through which to participate in the program delivery and are best considered as being volunteers and a human resources input.

Volunteers are widely used to deliver health program interventions. For example, volunteers have delivered a physical activity program for women in Iran (Pazoki, Nabipour, Seyednezami, & Imami, 2007), served as community health workers in Nepal (Glenton et al., 2010), and provided cardiac health peer education in Ontario (Karwalajtys et al., 2009).

The use of volunteers in health programs makes them a key human resource requiring particular attention, especially if they create a synergy between themselves and the program recipients (Hintgen, Radichel, Clark, Sather, & Johnson, 2000). In other words, volunteers gain from the experience even as participants benefit from their service in terms of personal attention (Chapin et al., 2013). Evidence suggests that volunteers may experience improved health outcomes themselves (Hong & Morrow-Howell,

2010). This could be used as an incentive to recruit volunteers.

Volunteers are motivated by factors that are different from those that motivate paid employees. Altruism is a major motivation to be a volunteer (Bowen, Kuniyuki, Shattuck, Nixon, & Sponzo, 2000). Other motivators include personal gains in terms of feeling a sense of belonging; assistance and allowances from the project, and pathways to paid work (Topp et al., 2015); and translating health knowledge into action (Singh, Cumming, Mohajer, & Negin, 2016). Older adults are likely to perceive the opportunity to engage in a new endeavour as the strongest reason to volunteer, although a sense of life satisfaction, a need to be productive, and social interaction are also important reasons to volunteer (Morrow-Howell, Kinnevy, & Mann, 1999).

Recruitment of volunteers can be accomplished through a variety of techniques. In a study of volunteers for one cancer prevention program, Bowen and associates (2000) found that half of the volunteers were recruited through media, such as television, radio, or newspapers, and from work-related sources. Although these methods are similar to those used to recruit employees, the messages for recruiting volunteers will be different. Specifically, these messages will focus on motivating factors such as those mentioned previously.

Volunteers bring variable relevant backgrounds in trainings and experiences with the focal population. This makes training volunteers particularly critical to the success of the health program, including clear written guidance and ongoing support tailored to the volunteer's needs (Allen et al., 2016). Volunteers may need not only the skills to implement the intervention effectively but also emotional support in dealing with challenging clients and knowledge about the clients and the health condition. Volunteers in programs for high-risk clients may be particularly vulnerable themselves, including being subject to police harassment (Bluthenthal, Heinzerling, Anderson, Flynn, & Kral, 2008). Another human resources issue arises from the fact that the source of satisfaction and work-related stress is different

for volunteers than for employees (Ferrari, Luhrs, & Lyman, 2007). And volunteers for programs involving disasters and violence may experience emotional distress and exhaustion (Clukey, 2010; Thornton & Novak, 2010).

One issue germane to community health programs and the use of volunteers was identified in the study by Trettin and Musham (2000), which found a difference between program developers and community volunteers in their perceptions of the role of volunteers. This finding hints at an undercurrent in the relationship between program planners and the community—namely, the perception that individuals are being exploited under the guise of being volunteers (Maes, 2012). This undercurrent will be more evident and influential in determining the success of a health program in some communities.

Physical Resources

Physical resources include material resources, facilities, supplies, and equipment (Kettner, Moroney, & Martin, 1999).

Material resources are those tangible items that are needed to provide the intervention and to provide program support. Intervention equipment might include, for example, blood pressure cuffs, syringes, or imaging machines. Space—usually called facilities—is another material resource needed for both the intervention and the program support. The costs associated with having or renting a classroom, auditorium, community meeting room, waiting room, examination room, and offices need to be taken into account. Supplies constitute another material cost. They encompass miscellaneous office supplies (e.g., paper, stationery, pens, clips, computer supplies), specific resources for program staff members (e.g., journal subscriptions, resource or reference manuals), and any items related to the intervention (e.g., tote bags, clipboards, stopwatches, Band-Aids, flip charts). Physical resources may be needed for the evaluation as well—generally office supplies and storage space for data collected. Maintaining adequate supplies, without hoarding, is the function of

inventory control; having too many unused supplies drains monetary resources that may be needed for other expenses.

A large category of physical resources is office equipment: computers, printers, cables, fax machines, and photocopiers. If the costs of these items exceed an amount set by the organization, the purchase of the item will be considered a capital expense and will require special approval and possibly special ordering procedures. Also, some funding agencies look unfavorably on the purchase of standard office equipment exceeding a certain amount. For these reasons, many program managers minimize purchases that are considered capital expenditures.

Transportation

Transportation is a separate type of resource and expense. Transportation must be thought of from the perspective of program staff members and program participants. For staff, the issue is one of reimbursement for any travel related to providing the program. For example, if the health program includes an outreach component, then staff members need to document carefully their mileage or keep track of expenditures for public transportation. From the perspective of program participants, the issue is accessibility of the program site given the usual mode of transportation used by potential program participants. Thus, in a rural program, the transportation issue might be one of travel time required to get to the program site, whereas in an urban program, the issue might be proximity to a major mass-transit stop.

Informational Resources

Computer hardware and software costs are always included in the budget, usually as physical resources. By contrast, information is a key intangible resource. The knowledge and expertise of the staff members must be viewed as a resource to be managed. Information possessed by individuals is considered an asset for the organization, and managerial attention is devoted to managing knowledge as resource (Castro, Delgado-Verde,

Navas-López, & Cruz-González, 2013; Rossi, Cricelli, Grimaldi, & Greco, 2016).

Information as a resource can be present in the form of professional networks (Gulati, 1999), street smarts that affect program implementation, and professional knowledge and experience. Staff members bring their professional knowledge to the program, and they gain additional knowledge through the training sessions. In terms of health programs, this means the knowledge held by employees involved in the program is valuable and deserves management attention.

As staff members become more qualified for other positions based on their knowledge and experience, it can become difficult to retain those personnel. Consequently, retention of such employees is a key issue for health programs and represents another reason to make efforts to keep the program personnel satisfied with their jobs and the organization. Loss of staff is not just an issue of replacing personnel; rather, it is more appropriately thought of as replacing knowledge and expertise—hence the importance of maximizing staff retention and minimizing unplanned staff turnover.

Time

Time generally translates into personnel costs, but it must also be thought of as a separate resource. Time is relevant to the overall time line for program design and implementation. The sequencing of events, especially in the start-up phase of a program, is very time dependent. Any delay in accomplishing one step could easily affect the overall time line. If meeting a start date is essential, then additional personnel resources may be required during the start-up phase, which in turn will affect the budget. Time also affects the budget in less obvious ways through depreciation, inflation, and interest. These points are discussed later in this chapter as part of planning via budgets.

Managerial Resources

The qualities and characteristics of the managerial personnel are also key resources. When

selecting the project manager, Posner (1995) suggests considering the skills required of a person in the position—namely, organizational abilities, communication skills, team-building skills, leadership qualities, coping skills related to managing complex and ambiguous tasks and environments, and technical skills related to the health program intervention.

Organizational abilities include the logistics of juggling multiple tasks and persons as well as the ability to keep track of important dates and information. The ideal program manager will be able to manage the project in terms of structuring the relationships among project personnel as well, as is usually reflected in organizational charts that identify the flow of communication, delegation, and responsibilities. A somewhat different but equally crucial aspect of organizational abilities is the understanding of how the health program fits within the parent organization, and the knowledge of how to position the program for success within the parent organization.

Communication skills encompass not only the ability to communicate verbally with peers but also the ability to guide discussions; communicate nonverbal messages that are consistent with verbal messages; and, perhaps most important, listen and appreciate various points of view. In addition, written communication skills are essential for a health program manager to function in the larger context of grant writing, report generation, and e-mail communication. Another facet of communication skills is the ability to understand who needs what and when as prerequisites to the creation and manipulation of information to have an effective and efficient health program.

Negotiation is a major component of communication skills required of program managers. During this process, two or more parties reach a shared decision despite having different preferences with regard to the options. Much research about negotiation has revealed what we know about how people think and respond to information. Three pieces of information are needed in any negotiation: each party's best

alternative to the negotiated agreement, analysis of all the interests of each party, and the relative importance of each party's interests. Armed with this information, the negotiator enters into a process in which strategies are used to arrive at an agreement by breaking out of the existing definitions of differences. These strategies include many actions commonly encountered in good communication—specifically, building good relationships and starting with defined goals (Halpert, Stuhlmacher, Crenshaw, Litcher, & Bortel, 2010). Program managers will negotiate not only with funding agencies and with program staff members and community stakeholders. Thus, a rudimentary understanding of negotiation techniques is important.

Team building is another essential managerial skill, especially for participatory, community-based, or multidisciplinary health programs. Team building includes mobilization of a group, fostering group cohesion, and facilitation of teamwork. Information about team building and teamwork can be found in both the lay managerial literature available at most bookstores and in professional management journals. Managers must consciously attend to the quality of interactions within teams and the quantity of teamwork. Problems can arise quickly, of course, and they must be addressed expeditiously. Otherwise, the team can become dysfunctional and drain resources rather than being a resource itself.

An indicator that team building has been successful is the use of stories. As program staff members begin to develop as a team, stories will be told about clients, managers, and events. These stories, which are ubiquitous in most organizations, serve the function of building team norms and establishing communication among the team members.

Leadership is the ability to inspire and motivate others into action that is purposeful and organized. Program managers are rarely thought of as leaders, yet they must motivate staff members, serve as role models of productivity, and generate enthusiasm for the health program. As a consequence, program managers

need to be knowledgeable about the motivational process in regard to both staff members and program participants. Motivation of program staff members results in all members doing what they were hired to do in a timely, efficient, and high-quality manner.

The technical skills required of program managers consist of skills related to the health program intervention. In most health programs, the manager must have basic scientific or practice knowledge about the health problem and the type of interventions that are provided. In some situations, program staff members, program participants, or funding agencies may not view the health program favorably if the program manager does not have some professional credibility. Technical skills are also necessary to supervise the program staff adequately and ensure the integrity of the program intervention.

Fiscal Resources

Fiscal resources encompass money, investments, and income. Existing fiscal resources are generally listed on the income side of a balance sheet or budget because they comprise funding and income generated through fees and billing. Monetary donations are another resource that need to be tracked and included in accounting records. However, if participants are given any cash for participation, then monetary resources are placed on the expense side of the budget.

Budgets are mechanisms for planning and tools for communicating and refining priorities. As projections of dollar amounts, budgets enable the program planner to assess the fiscal feasibility of doing a project. Developing a budget for a program highlights which programmatic changes may be needed for the health program to be fiscally responsible and efficient. For these reasons, budgets are an input to the health program organizational plan. Each organization will have its own particular format for developing budgets and may use special software for this task. The financial officer in the organization will set forth the rules and accounting specifics used across programs and departments within the organization.

► Organizational Plan Outputs

Not all inputs into the organizational plan will be directly linked to specific organizational plan outputs; some will be linked to services utilization plan outputs. The outputs discussed in the following subsections of the chapter are examples that could be measured or at least documented as effective use of the organizational plan inputs.

Time Line

A *time line* is a way to represent graphically the dates, time span, and sequence of events involved in planning, initiating, sustaining, and evaluating the health program. Time lines vary from simple to highly intricate. At a minimum, the time line ought to reflect the necessary activities, be related to the expenses detailed in the budget, and be easily understood by those who will use it (**EXHIBIT 7-1**). A time line can be created by using the table function in a word processing program, a spreadsheet, or specific project management software. This communication tool conveys deadlines, helps keep activities coordinated and sequenced, communicates accountability for assigned tasks, and helps program managers estimate personnel and material costs.

Operations Manual

The *operations manual* contains the policies, procedures, guidelines, and protocols related to the health program. It also includes job descriptions and workplace policies and procedures.

One element of many health program operations manuals is a section on safety. For both staff members and participants, safety is an issue—safety not only in their cars as they are coming to or from the program site but also on the street surrounding the program site. In particular, if program staff members are outreach workers, their safety must be considered. In such a case, safety strategies, including monitoring, become an important management responsibility (Reese, 2015).

EXHIBIT 7-1 Example of an Abbreviated Time Line for a Short-Term Health Program

Activity	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Convene program planning group												
Conduct community needs and asset assessment												
Translate assessment information into health problem statement and objectives												
Formulate program theory; articulate process theory												
Initiate institutional process regarding human subjects' protection												
Advertise for program personnel												
Hire and train program personnel												

(continues)

EXHIBIT 7-1 Example of an Abbreviated Time Line for a Short-Term Health Program (*continued*)

Activity	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Advertise the health program (social marketing)												
Deliver the health program												
Conduct process evaluation												
Conduct outcome evaluation												
Analyze evaluation data												
Produce reports												
Disseminate findings												

Organizational Chart

Most organizations construct graphic representations to depict the relationships among work units, departments, and individuals. As a result of having personnel designated for the health program and having program accountability, health programs have their own organizational chart. One aspect of the health program organizational chart that may be different from the organizational charts of other work units is the inclusion of a community-based consortium or

council that serves as an advisory committee to the health program. Inclusion of such groups is important because it reflects, in a legitimate and visible fashion, the involvement of the community in the health program.

Also, the health program ought to be identified specifically within the appropriate organizational chart. For health programs that are population-based, such as state child health insurance, the program ought to appear somewhere in the organizational chart of the state agency. The program also ought to be included

in the organizational charts of the community organizations involved in promoting or providing the program.

Information System

The information system has outputs that are part of the organizational plan. The programs and templates created for the program are outputs, as well as the analyzed data. New report capacities constitute information system outputs. To the extent possible, and without losing sight of program objectives, evaluators should try to use existing data. Whatever the data source, it is wise to analyse a small batch of pilot data as early as possible, ideally before the program begins. This can enable program staff members to improve data quality or enable evaluators to adjust evaluation plans based on what is actually available.

► Inputs to Service Utilization Plan

Social Marketing

Kotler and Zaltman (1971) were the first authors to advocate social marketing as a method to reach a wide audience with health or social messages. Subsequently, Walsh, Rudd, Moeykens, and Moloney (1993) defined *social marketing* as the design, implementation, and control of the program calculated to influence the acceptability of social ideas.

Social marketing adapts the four P's of classical marketing: product, price, place, and promotion (Cheng, Kotler, & Lee, 2009). *Product* refers to the services, tangibles, or ideas that are delivered with the intent of being beneficial to the intended audience. Social marketing focuses on the beneficial aspect and understanding from the perspective of the recipient how those benefits are valued and perceived. *Price* in social marketing is the cost, of any type, that poses a barrier to accessing or using the product. Price

considerations focus not only on the charge for the service but also on the secondary costs, such as transportation or loss of peer group status. *Place* refers to where the product is available, whether it is in a clinic, on a billboard, or in a convenient location within a store. Product accessibility, convenience, and visibility are qualities of place that merit attention. *Promotion* is the more visible, publicity type of activities, including paid media, public service media, and word-of-mouth sharing by opinion leaders. Social marketing goes beyond these classic four P's to include other P's, such as partnership and policy.

Social marketing can be combined with health communications, which is sometimes called health marketing (Centers for Disease Control and Prevention [CDC], 2011). Principles from social marketing have been used successfully in a variety of health educational programs to deliver the health program to the intended audience (French, 2010). In this way, the development and implementation of the social marketing strategy serve as an input to the service utilization plan because that strategy enables the health program to reach the intended audience.

Eligibility Screening

One of the first decisions facing program planners is to define for whom the program is designed. Although general agreement may exist about who is the focus of the program based on the needs assessment and the logic model, further specificity is required. The *intended population* is the entire population in need of the program, whereas the *intended audience* is the segment of the population for whom the program is specifically envisioned and planned. The term *recipient* refers to those individuals who actually receive or participate in the program.

For programs at the population level of the public health pyramid, the intended population is also the intended audience and—ideally—the recipients. At the direct services and enabling services levels of the pyramid, however, no program can hope to accommodate all those persons in need; thus, the program may be designed for

a subpopulation. For such programs, *intended audience* is a better term.

The distinction between the intended population or audience and the recipients is critical in terms of both budgetary issues and program implementation and evaluation issues. The program can have an impact only on the recipients, so the evaluation will focus primarily on this group. Nevertheless, planners also need to quantify the broader intended audience to estimate underinclusion and overinclusion in the program and describe how these variations may influence the evaluation.

Reach: Underinclusion and Overinclusion

Ideally, only members of the intended audience would receive the program. In reality, this ideal can be difficult to achieve, resulting in overinclusion or underinclusion of individuals in the program.

Overinclusion occurs when some participants in the program are not part of the intended audience. It can be minimized by developing procedures to exclude correctly individuals who are not members of the intended audience. For example, in a dental sealant program, children who are younger than 2 years of age, as well as children who are between the ages of 5 and 14 and who already had sealant treatment, might be excluded to avoid overinclusion.

Underinclusion occurs when some members of the intended audience do not receive the program. It can be minimized by developing procedures to include correctly members of the intended audience for whom the program is designed. Underinclusion in the dental sealant program would be having fewer children of the appropriate age receive the dental sealant than the number of children of that age who need the sealant and who are within the catchment area of the dental clinic. Underinclusion can occur if the program is not well publicized, if some characteristic of the program is unappealing to the intended audience, or if a barrier prevents members of the intended audience from accessing the program.

In terms of program expenditures, overinclusion can result in a shortage of funds, whereas underinclusion can result in unspent funds that may need to be returned to the funding agency. Underinclusion and overinclusion are also undesirable from the perspective of the program evaluation. Providing services to individuals who do not need the program (overinclusion) can decrease the measurable effect of the program on participant outcomes. That is, the extent of change experienced by those individuals who do not need the program is likely to be less than the extent of change experienced by those individuals who do need the program. This diffusion of the program's effects translates into a decrease in the average amount or degree of change found when all participants in the program are considered. In addition, overinclusion may artificially inflate the normative need for the program. If current enrollment or requests for participation in the program are used for future planning of the program, overinclusion would falsely increase the apparent number in the intended audience. Overinclusion also can lead to a decreased availability of funds to include true intended individuals in the program. This result is particularly likely if members of the true intended audience are more likely not to be the first persons to enroll in the program.

Underinclusion can also affect evaluation results, particularly for programs designed for and delivered at the population level. Having too few members of the intended audience in the program: (1) could make it difficult to find significant effects due to small numbers, (2) could increase the amount of program services received by individual participants and thereby falsely inflate the program effects, and (3) definitely increases the cost per participant. At any level of the public health pyramid, underinclusion can lead to biased evaluation results if members of the intended audience who do and do not participate in the program differ from one another in ways that are related to the program's effectiveness.

Several steps can be taken to help minimize overinclusion or underinclusion. The first step comes in developing the process theory in terms

TABLE 7-2 Relationship of Test Sensitivity and Specificity to Overinclusion and Underinclusion

Sensitivity	Specificity	
	High	Low
High	Ideal inclusion and coverage; minimal overinclusion and underinclusion	Overinclusion
Low	Underinclusion	Overinclusion and underinclusion

of specifying how those in need of the program get into the program; this is part of the services utilization plan. Another step is to have a solid, thoughtful marketing plan, which is another element in the services utilization plan.

Once the intended population or audience has been clearly specified, screening tests that are both highly sensitive and specific can be used to minimize both overinclusion and underinclusion (**TABLE 7-2**). Test *sensitivity* refers to the probability that the screening test is positive when an illness, need, or existing risk factor is actually present. Using a highly sensitive screening test to identify individuals who are eligible for the program increases the likelihood that more individuals who actually need the program will be in it, thereby reducing underinclusion. Test *specificity* refers to the probability that the test is negative when there is no illness, need, or risk factor. Using a highly specific screening test to identify individuals who are not eligible for the program results in fewer individuals in the program who do not need it, which reduces overinclusion.

In practice, it is never possible to have a screening test that is both 100% sensitive and 100% specific. Typically, a trade-off must be made between sensitivity and specificity, with the screening test being either more sensitive or more specific. Nonetheless, a screening mechanism is often the best way to minimize overinclusion and underinclusion.

Scope: Full and Partial Coverage

The distinction between intended population, intended audience, and recipient is also critical in determining whether the program has partial or full coverage (Rossi & Freeman, 1993). This distinction between partial- and full-coverage programs has implications for public health and health policy.

Partial-coverage programs are designed to serve some portion of the intended population, and participation in the program is based on a set of criteria that focuses recruitment strategies and takes into account limited resources. During the planning stage, the decision to have a program that provides only partial coverage generally stems from having limited capacity to serve all those in the target population. Partial-coverage programs are likely to occur at the direct care or enabling services levels of the public health pyramid. Examples of such programs include early childhood intervention programs for children at developmental risk, and hospice care for those dying who choose that service.

Full-coverage programs are delivered, or are intended to be delivered, to the entire target population. By definition, these programs are more likely to occur at the population services level of the public health pyramid. Examples of full-coverage programs include seat-belt laws and water fluoridation.

TABLE 7-3 Examples of Partial- and Full-Coverage Programs by Level of the Public Health Pyramid

Pyramid Level	Partial Coverage (for Segment of Intended Population)	Full Coverage (for Entire Intended Population)
Individual—direct services	Dialysis for portion of those with kidney failure, early-childhood-intervention programs for children at developmental risk, hospice care for those dying who choose the service	Ambulance and emergency medical care for all individuals, immunization clinics available to all individuals
Aggregate—enabling services level	Needle exchange programs for some substance abusers, Medicare services for the disabled and homebound, Head Start for low-income children	Medicaid coverage for dialysis for those with kidney failure
Population-based level	WIC program for low-income families, state child health insurance plans (SCHIP) for low-income families	Seat-belt laws, Medicare coverage for all individuals older than age 65, fluoridation of the water supply for all residents
Infrastructure level	Portable electronic devices for nurses making home visits	Licensure for all physicians, nurses, and dentists; national cancer registry

Some programs are less readily identified as providing full or partial coverage because, although the program is designed with the population level in mind, the intended population is restricted to those meeting the criteria for participation, such as income level for state child health insurance programs; federal Women, Infants, and Children (WIC) nutrition programs; or Medicare. Because these programs are intended to serve the entire intended population, they could be considered full-coverage programs, although the eligibility criteria are such that they are partial-coverage programs. Across the public health pyramid, both partial- and full-coverage programs reflect whether the program is designed primarily to make changes at the individual, aggregate, or population level (**TABLE 7-3**).

Making a distinction between full- and partial-coverage programs during the planning phase may not be so easy. Stakeholder issues may arise, and advocacy positions may be taken

regarding whether a program ought to provide full or partial coverage. Frontline staff members may hesitate to decline referrals from colleagues with whom they have ongoing interactions, regardless of budgetary or logistical issues. The expectations of these partners must be taken into account and reconciled with issues of feasibility if the program is to be successful in gaining their support or endorsement. Also, if the health program is to provide partial coverage, program planners must establish eligibility criteria and procedures for prioritizing and enrolling potential recipients. This issue, of course, can lead to considerable debates over the particulars of the cutoff criteria chosen for program eligibility. The other reason to consider whether the health program will provide full or partial coverage is that the scope of the program affects the design of the evaluation and potentially the cost of conducting an effect evaluation of the program.

In addition, the scope of the program needs to be considered with respect to underinclusion

or overinclusion. Overinclusion is much more difficult to detect in a full-coverage program than in a partial-coverage program because the presumption in the former type of program is that the intended program recipients are the members of an entire intended population. Given that a full-coverage program, by its nature, is likely to have a large number of recipients, it will be difficult to identify those few recipients who are not members of the intended population. By contrast, in a partial-coverage program, overinclusion is more likely to occur than underinclusion.

The interaction of the scope of a program with the appropriateness of inclusion may have implications that need to be considered, especially when developing the program marketing plan and the program budget. It will also be a consideration in establishing the eligibility criteria and screening procedures. In addition, an awareness of the potential for this interaction may help explain later findings during evaluations of either the process or program effect.

Screening

Program eligibility screening is another services utilization plan input. Such a procedure is necessary to ensure that the program is provided to members of the intended audience, thereby minimizing underinclusion or overinclusion. Despite the inclination to want to provide the program to anyone interested, screening enhances the efficiency and the effectiveness of the program. Efficiency is enhanced by providing the program only to those participants to whom the intervention is tailored, thereby making it less likely that the program intervention will need to be individually tailored. Effectiveness is enhanced because only those who need the program, and thus are more likely to experience the benefits of the program, are included in the outcome and impact evaluations.

Queuing

Waiting to be seen for services, being on hold, and having to wait until services become available

are all aspects of being put in a queue. Waiting lines and wait times reflect the degree of match between the capacity to provide the service and the demand for the service. The services utilization plan ought to include a plan for handling wait-lists and such.

For example, if an immunization clinic is being held, the services utilization plan ought to balance the anticipated number of individuals seeking immunizations against the rate at which individuals can be processed for the immunization, the length of time needed to give the immunization, and the number of program staff members available to implement the immunization clinic. An imbalance will result in either people waiting for long periods or staff members not having work.

Because the particulars of studying queues can be complex, large health programs—and particularly those that are ongoing—may find it valuable to hire an operations specialist to study the issue of queuing. Consultation with such experts can ensure that the health program is provided in the most timely and efficient manner possible.

Intervention Delivery

The fact that the discussion of intervention delivery appears halfway through this text is no accident. The actual delivery of the intervention, though it takes the most effort (see Exhibit 7-1), is relatively easy if the planning has been well done and the process theory has been thought out thoroughly.

Delivery of the intervention ought to follow the protocols and procedures developed specifically for the health program. Such adherence to the plan ensures that the intervention is delivered in a standardized and consistent manner. The level of detail included in the protocols and procedures will vary across programs, of course. For example, if the health program intervention involves secondary prevention of breast cancer with mammography screening, then the intervention protocol should include considerable details about the procedure for

doing the actual mammogram, the taking of a history before the mammogram is performed, and notification and referral of those screened. In contrast, if the health program is a metropolitan-wide mass-media awareness campaign about the value of adult immunizations, then the intervention protocol should allow for flexibility in accessing and communicating with media contacts while providing guidance in who to approach, which topics to address and to avoid, and which products (e.g., public service announcements, video clips, flyers) to distribute to which media sources.

Implementing evidence-based interventions requires managers' commitment to ensure the successful uptake of these practices in diverse settings. For example, researchers studying five evidence-based practices in community mental health settings found more actively engaged leadership, attention to workflow dynamics such as policy and documentation, and reinforcement such as monitoring and feedback to be associated with better implementation fidelity (Torrey, Bond, McHugo, & Swain, 2012).

One aspect of the intervention that makes common sense, but is often neglected, is pilot-testing the program. Prior to full program implementation, the program may be

pretested, with the program developers paying attention to the various program components. Pretesting can take many different forms, such as having focus groups review materials and make comments, providing the program for free to a small group, or having experts comment on the program design and materials. In particular, marketing materials and media messages ought to be pretested. Characteristics to be assessed include the materials' attractiveness, comprehension, acceptability, and persuasion. Any materials that will be read by program participants should also be tested for readability. The rule of thumb is the lower the level, the better. No one wants to struggle with technical language, complex sentences, or large words. Generally, an eighth-grade reading level is recommended for materials. **EXHIBIT 7-2** shows this paragraph as rewritten at an eighth-grade level.

Pretesting of the interventions ought to include pretesting of the evaluation instruments that will be used during the program and after the program has been completed. As more health programs are required by funding agencies to document their success, the evaluation is becoming more integral to the actual intervention and overall program delivery.

EXHIBIT 7-2 Chapter Text Paragraph Rewritten at an Eighth-Grade Reading Level

It makes common sense to try out a program before it begins, but this is often not done. Before starting a program, try out the handouts and the program parts. There are many ways to see if the people in the program will understand the handouts and program. One way is to have a focus group look over the materials and make comments. Another way is give the program to a small group of people for free and see if there are any problems. Also, experts can help by making comments on the program and the handouts. Advertisements and media messages need to be tried out before they are used, too. Look at how attractive they look, how easy it is for people to understand them, whether people will accept the message, and how good the messages are at convincing people to participate in the program. Also, handouts need to be checked for how easy they are to read. Today, word processing software can check them and show the grade level they are written at. The rule of thumb is that lower grade levels are better because no one wants to work at reading technical words, long sentences, or large words. For example, this paragraph is written a little below the eighth-grade level.

► Services Utilization Plan Outputs

Outputs of the services utilization plan include the number of units of service provided and the quantity of service completions. *Units of service* is a term used to refer the agency- or program-specific quantification of what was provided, such as hours per client, number of inpatient visits, number of educational sessions, or number of hours of client contact. Because what constitutes a unit of service can vary widely, each health program must specify what it considers a unit of service. Another service utilization output is the number of services that have been completed. For health programs, this might be the number of immunization clinics held, the number of completed referrals for medical follow-up, or the number of health educational courses provided.

Ensuring that a mechanism is in place to track the service plan outputs is a critical managerial responsibility. **TABLE 7-4** is a template for tracking the interventions and the units of service outputs. Fuller et al. (2007) used this format to show the amount of reach across different intended populations.

One other services utilization output exists as well: the materials developed and produced as part of the effort to provide the health program, such as public service announcements (PSAs), educational videos, annual reports, or curricula. Another key output is the workflow—that is, the extent to which program staff members have work over a given time period or that work is done in a coordinated manner.

Keeping a record of the various outputs as the health program is implemented makes it possible later to conduct a process evaluation of activities focused on program implementation. These evaluation activities are discussed in detail elsewhere in the text.

TABLE 7-4 Template for Tracking Services Utilization Outputs Using Example Interventions and Hypothetical Activities

Intervention Component	Intended Audience A: Persons with a Health Problem	Intended Audience B: Clinics	Intended Audience C: City
Individual education	100 persons at risk	76 providers visited	(Not applicable)
Individual screening	600 persons	30 providers	1,000 persons screened at health fairs
Group education	15 groups of persons at risk	8 groups at 3 clinics	15 groups at library; 6 groups at schools
Population education	2,000 flyers; 600 stickers	50 posters	25 public service announcements (PSAs); 10 health fairs
Individual support for behavior	125 persons at risk	32 providers	(Not applicable)

Summary: Elements of Organizational and Services Utilization Plans

This chapter has explained the elements of the organizational plan and services utilization plan, which collectively make up the process theory. In designing and planning for a health program, attention to the process theory becomes a reality check that the needed resources and processes are in place to initiate the program.

For example, continuing with the Layetteville example, **FIGURE 7-3** gives a possible process theory for a neural tube defect prevention program. However, the process theory needs to correspond to and support the intervention theory. When the process theory is combined with the effect theory, a comprehensive view of the program becomes possible (**FIGURE 7-4**).

Having a diagram such as Figure 7-4 allows planners to determine what might be missing as part of the effort to address the health problem and to deliver the intervention, as well as what has been added that might not be essential to achieving the desired program effect.

Alternative Plan Formats

In summary, the program theory shows the inputs and outputs of the organizational plan and the service utilization plan. Other formats for summarizing this information exist. Each format has advantages and disadvantages, and will resonate with different audiences. Regardless of which format is used, the intention is to anticipate, to the fullest extent, the needs and

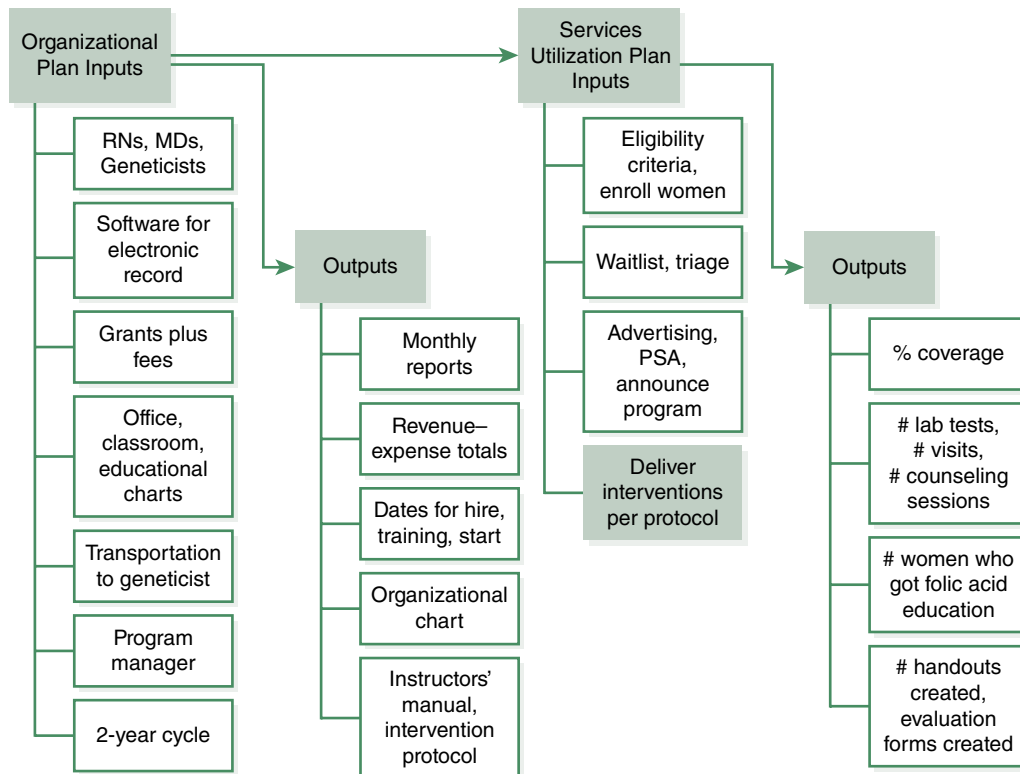


FIGURE 7-3 Process Theory for Neural Tube Defects and Congenital Anomalies Health Problem

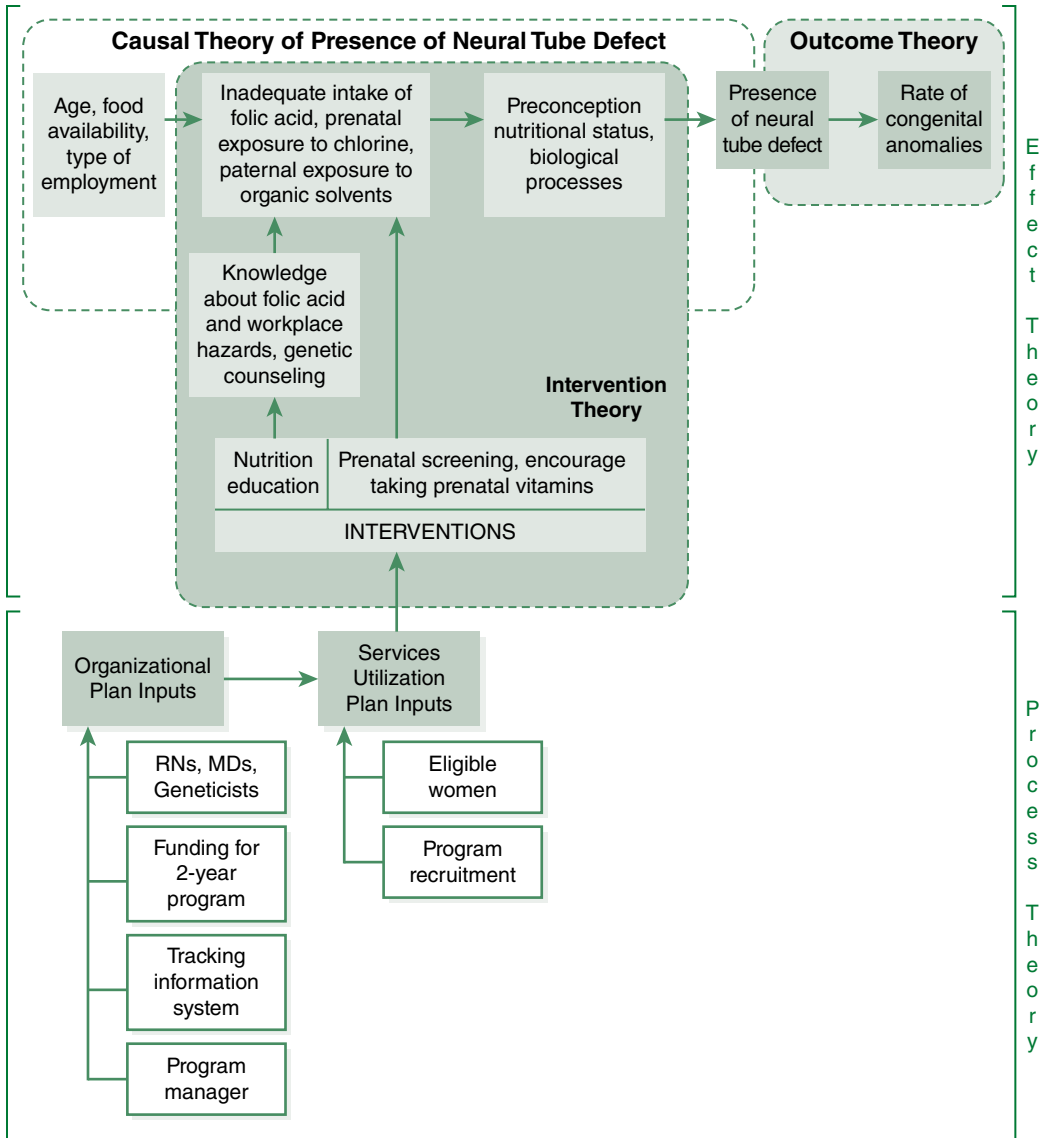


FIGURE 7-4 Effect and Process Theory for Neural Tube Defect Prevention Program

activities that will make the health program as successful as possible. The two most commonly used formats are reviewed here.

Logic Models

As a tool for program planning and evaluation, a *logic model* is a diagram in a table format that shows the relationship of inputs and activities

to outputs, immediate outcomes, and long-term outcomes. Since their introduction in the late 1980s, logic models have become ubiquitous and are frequently required by funding agencies. The CDC has used logic models for its programs, and the Public Health Foundation includes them as one of the quality improvement tools. A logic model is one-page tabular summary of the program theory. The logic model comes in

TABLE 7-5 Hypothetical Logic Model of a Program for Reducing Congenital Anomalies

Assumptions	Inputs	Activities	Outputs	Immediate Outcome	Long-Term Outcome
Health problem is important to community of providers and residents. Residents have access to health care and food sources.	Program staff members. Location to hold the screening. Donation for printing.	Create a time line and tracking system. Conduct staff training. Create program protocol and manual. Create PSAs and recruitment materials.	Number of educational materials; program manual. Number of staff trained. Number of women screened for type of employment and serum folic acid. Percentage of women who start folic acid preconception.	Improved knowledge of folic acid. Percentage of births with neural tube defects (anencephaly, spina bifida), cleft lip or palate.	Rate of congenital anomalies.

many variations. **TABLE 7-5** is an example of a logic model for the neural tube defect health problem and the interventions chosen to address the problem.

Comparing the process and effect theories for a congenital anomalies health problem (Figure 7-4) with the logic model for the same problem (Table 7-5), the differences and areas of overlap become clear. The logic model lists activities but does not distinguish between the actual intervention activities and the various activities that are part of the organizational plan inputs. In contrast, the process theory clearly distinguishes the actual intervention from other activities associated with the organizational plan and service utilization plan inputs. The logic model can be thought of as a different representation of the overall program theory. Both logic models and the program theory models function as tools to help anticipate needs and gaps, to engage in thinking about the entire program, and ultimately to have a plan of what

the program encompasses that can be used to guide process and effect evaluations.

Kaplan and Garrett (2005) drew on experience with three comprehensive, multisite programs when they reviewed the corresponding logic models to identify the value of logic models. They found that development of a logic model not only fostered collaboration for already strong coalitions but also proved challenging for diverse and underresourced organizations or coalitions. The logic models they reviewed did reveal assumptions that led to changes in staffing for the program, but often the assumptions were unstated or unrealistic. These authors also found that the lack of attention to the rationale for the program—in other words, the intervention theory—was related to the heavy emphasis that the logic model put on activities and outcomes. A key benefit of the logic model was that it facilitated communication with internal and external stakeholders, including

fundlers. Overall, logic models are complex and time consuming to develop well, but they can pay off through improved alignment among intended population health needs, programs, and subsequent process evaluation (Gervais, Montigny, Lacharité, & Dubeau, 2015).

Business Plans

A *business plan* is a document that summarizes the analyses behind the development of a product, service, or program; substantiates the development of a program based on a wide range of crucial considerations; and provides details for implementation of the program. Business plans, in contrast to the diagrammatic layout of the process theory and logic models, are written documents. Nevertheless, as with the process theory diagram and logic models, the overriding intent is to communicate the details of the program and to provide evidence that the details have been thoroughly thought out and actions are well planned and coordinated.

Research has shown that business plans facilitate decision making and accelerate development of the product or service (Burke, Fraser, & Greene, 2010). The use of business plans is advocated in medicine (Cohn & Schwartz, 2002), nursing (Hamric, Hanson, Tracy, & O'Grady, 2014), and public health (Orton, Umble, Zelt, Porter, & Johnson, 2007). Although development of health programs is rarely thought of as an entrepreneurial activity, the development of health services is very appropriately viewed from this perspective. As Ellis, Embree, and Ellis (2015) note, business cases are especially relevant for more costly initiatives. They advocate clarifying implementation plans and outcomes measures in order to demonstrate the thoughtful development of new initiatives.

TABLE 7-6 lists the basic elements of a business plan, along with the corresponding elements of the organizational plan, services utilization plan, and logic model. As the table makes clear, the business plan, because it is a written document, has greater detail in many areas. By comparison, logic models provide the least amount of information

and the least specific information. More detail, particularly about the financial characteristics of the proposed program, is often needed to communicate effectively and convincingly the need for the proposed health program.

► Across the Pyramid

At the direct services level of the public health pyramid, implementation of a health program is very similar to the implementation of other health services. However, particular attention should focus on tailoring the human resources to the programmatic intervention. The social marketing plan is also targeted to individuals and their behaviors.

At the enabling services level, a match between providers and the health program intervention is necessary. If the program intervention is an enabling service, the providers (human resources) are more likely to have a background or expertise in social services. Enabling service programs are also more likely to use volunteers, a decision that has implications for the managerial resources needed. The social marketing plan is also needs to be tailored to the aggregate targeted by the program. Because many enabling services either require a referral or are accessed via a referral, the social marketing of the program may focus as much on the providers making the referrals as on the intended audience of clients.

At the population services level, as at the other levels of the public health pyramid, a match must exist between the abilities and skills of the providers and the health program intervention. Social marketing to a population should have a broad base of appeal and should almost certainly use mass media.

This chapter has dealt mostly with the infrastructure level of the public health pyramid. Having a highly specified, comprehensive organizational plan and services utilization plan provides a strong foundation for implementing an effective health program. Also, having the necessary and appropriate resources for the

TABLE 7-6 Generic Elements of a Business Plan, with Their Purpose and Corresponding Element of the Process Theory and Logic Model

Business Plan Format	Purpose	Process Theory Element	Logic Model Element
Title/cover page	Gives a first impression		
Executive summary	Gives a first impression		
Business concept	Describes the program design, with goals and objectives	Entire program theory	Entire logic model
Market analysis	Analyzes the demand, need, competition, and effect on existing services and health status	Community needs assessment	
Financial analysis	Projects revenues and expenses; states the fiscal assumptions used in analyses	Organizational plan inputs: monetary resources, budget	
Risk and competitive analysis	Discloses the sources and types of possible failures, with alternatives to avoid those failures; balances failure risks with merits of the program		
Operational plan	Shows how personnel, management, space, and equipment come together for the program; delineates resource requirements	Organizational plan inputs: human, informational, physical, managerial, time resources, transportation	Inputs, activities
Marketing plan	Describes strategy to reach the intended audience, branding, distribution, price, promotion	Service utilization plan: social marketing, participants	Activities
Milestones	Gives time frame for accomplishment of key tasks and outcomes	Process and outcome objectives	Immediate and long-term outcomes

health program is an indicator of the quality of the infrastructure. Even the most creative ideas and the most scientifically sound programs will fail if they lack an adequate infrastructure. Too

often attention is focused on the health program interventions and clients, without the prerequisite attention and effort being devoted to developing and maintaining the programmatic infrastructure.

For this reason, the organizational plan must identify resources for the infrastructure as well as resources for the services utilization plan activities. In addition, if the health program is intended to increase the capacity of the infrastructure—for example, to improve workforce capacity—an organizational plan and services utilization plan for such a program is still warranted as a step to ensure the success of the program.

► Discussion Questions and Activities

For all of the questions below, please pick one of the following health issues identified as salient in Bowe County's most recent community health needs assessment: vaccination, adolescent pregnancy, gunshot wounds, or type 2 diabetes.

1. Give examples of how the outputs of the organizational plan and the services utilization plan relate to the process theory objectives.
2. What type of social marketing would you use?
3. How might you minimize underinclusion or overinclusion in such a health program?
4. Draft a logic model for this health program.

► Internet Resources

There is a wealth of resources on developing various types of logic models and program theory models. The following list represents diverse perspectives and approaches.

Reading Level of Written Materials

A Google search using “reading statistics” will yield instructions for assessing the reading grade level of text highlighted in Word. Also, “CMS’s Toolkit for Making Written Material

Clear and Effective” (<https://www.cms.gov/outreach-and-education/outreach/written-materialstoolkit/downloads/toolkitpart07.pdf>) contains a subsection on using reading formulas.

Kellogg Foundation Evaluation Manual

This text, which is available as a PDF file, is a comprehensive publication on evaluation. Chapter 5 (<https://www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook>) relates to the development of program theory. You will be asked to type in an e-mail address in order to download a PDF copy.

ReCAPP BDI Logic Model Course

For an audio experience, cruise through the highly recommended online walkthrough on logic models found at <http://recapp.etr.org/recapp/documents/logicmodelcourse/index.htm>. The material is presented in PowerPoint.

Centers for Disease Control and Prevention (CDC)

The CDC has a metasite for logic models with some good links. Find it at www.cdc.gov/eval/resources/index.htm. The CDC also provides scholarly papers on marketing at <http://www.cdc.gov/healthcommunication/>

Tools of Change

The Canadian website at <http://www.toolsofchange.com/en/home/> has links to social marketing as well as numerous examples of programs using various techniques for motivation.

Social Marketing Institute

Working papers and published articles at the Social Marketing Institute website (<http://socialmarketing.org/>) are scholarly resources.

Business Plans

Numerous resources on developing business plans can be found on the Internet, including some sites that provide how-to guides and

templates. Entrepreneur.com has an entire section dedicated to developing business plans at www.entrepreneur.com/businessplan/. The U.S. Small Business Administration (SBA), created by Congress to help, counsel, and protect small businesses, provides an online guide to writing a business plan at <https://www.sba.gov/starting-business/write-your-business-plan?interiorpage=2015>. The nonprofit My Own Business offers an entire online course about developing a business; the section on business plans (www.myownbusiness.org/s2/) includes basics, formatting, examples, and templates.

References

- Allen, R. S., Azuero, C. B., Csikai, E. L., Parmelee, P. A., Shin, H. J., Kvale, E., . . . Burgio, L. D. (2016). "It was very rewarding for me . . .": Senior volunteers' experiences with implementing a reminiscence and creative activity intervention. *The Gerontologist*, *56*(2), 357–367. doi:10.1093/geront/gnu167
- Bluthenthal, R. N., Heinzlerling, K. G., Anderson, R., Flynn, N. M., & Kral, A. H. (2008). Approval for syringe exchange programs in California: Results from a local approach to HIV prevention. *American Journal of Public Health*, *98*(2), 278–283. doi:10.2105/AJPH.2005.080770
- Bowen, D. J., Kuniyuki, A., Shattuck, A., Nixon, D. W., & Sponzo, R. W. (2000). Results of a volunteer program to conduct dietary intervention research for women. *Annals of Behavioral Medicine*, *22*, 94–100. doi:10.1007/BF02895173
- Burke, A., Fraser, S., & Greene, F. J. (2010). The multiple effects of business planning on new venture performance. *Journal of Management Studies*, *47*(3), 391–415. doi:10.1111/j.1467-6486.2009.00857.x
- Castro, G., Delgado-Verde, M., Navas-López, J., & Cruz-González, J. (2013). The moderating role of innovation culture in the relationship between knowledge assets and product innovation. *Technological Forecasting & Social Change*, *80*, 351–363. <http://dx.doi.org/10.1016/j.techfore.2012.08.012>
- Centers for Disease Control and Prevention. (2011). Health marketing. Retrieved from www.cdc.gov/healthcommunication/HealthBasics/WhatIsHC.html
- Chapin, R. K., Sergeant, J. F., Landry, S., Leedahl, S. N., Rachlin, R., Koenig, T., & Graham, A. (2013). Reclaiming joy: Pilot evaluation of a mental health peer support program for older adults who receive Medicaid. *Practice Concepts*, *53*(2), 345–352. doi:10.1093/geront/gns120
- Cheng, H., Kotler, P., & Lee, N. R. (2009). *Social marketing for public health: An introduction*. Sudbury, MA: Jones and Bartlett.
- Clukey, L. (2010). Transformative experiences for Hurricanes Katrina and Rita disaster volunteers. *Disasters*, *34*(3), 644–656. doi:10.1111/j.1467-7717.2010.01162.x
- Cohn, K. H., & Schwartz, R. W. (2002). Business plan writing for physicians. *American Journal of Surgery*, *184*, 114–120. [http://dx.doi.org/10.1016/S0002-9610\(02\)00921-2](http://dx.doi.org/10.1016/S0002-9610(02)00921-2)
- Ellis, R. J. B., Embree, J. L., & Ellis, K. G. (2015). A business case framework for planning clinical nurse specialist–led interventions. *Clinical Nurse Specialist*, *29*(6), 338–347. doi:10.1097/NUR.0000000000000162
- Ferrari, J. R., Luhrs, T., & Lyman, V. (2007). Eldercare volunteers and employees: Predicting caregiver experiences from service motives and sense of community. *Journal of Primary Prevention*, *28*, 467–479. doi:10.1007/s10935-007-0108-6
- French, J. (2010). *Social marketing and public health: Theory and practice*. Oxford, UK: Oxford University Press.
- Fuller, C. M., Galea, S., Caceres, W., Blaney, S., Sisco, S., & Vlahov, D. (2007). Multilevel community-based intervention to increase access to sterile syringes among injection drug users through pharmacy sales in New York City. *American Journal of Public Health*, *97*, 117–124. doi:10.2105/AJPH.2005.069591
- Gervais, C., Montigny, F., Lacharité, C., & Dubeau, D. (2015). The Father Friendly Initiative within Families: Using a logic model to develop program theory for a father support program. *Evaluation and Program Planning*, *52*, 133–141. <http://dx.doi.org/10.1016/j.evalprogplan.2015.04.006>
- Glenton, C., Scheel, I. B., Pradhan, S., Lewin, S., Hodgins, S., & Shrestha, V. (2010). The female community health volunteer program in Nepal: Decision makers' perceptions of volunteerism, payment, and other incentives. *Social Science & Medicine*, *70*, 1920–1927. <http://dx.doi.org/10.1016/j.socscimed.2010.02.034>
- Gulati, R. (1999). Network location and learning: The influence of network resources and firm capabilities on alliance formation. *Strategic Management Journal*, *20*, 397–420. doi:10.1002/(SICI)1097-0266(199905)20:5<397::AID-SM1097>2.0.CO;2-K
- Halpert, J. A., Stuhlmacher, A. F., Crenshaw, J. L., Litcher, C. D., & Bortel, R. (2010). Paths to negotiation success. *Negotiation and Conflict Management Research*, *3*(2), 91–116.
- Hamric, A. B., Hanson, C. M., Tracy, M. F., & O'Grady, E. T. (2014). *Advanced practice nursing: An integrative approach*. 5th edition. Elsevier/Saunders. St. Louis, MO.
- Hong, S.I., & Morrow-Howell, N. (2010). Health outcomes of Experience Corps: A high-commitment volunteer program. *Social Science & Medicine*, *71*, 414–420.
- Hintgen, T. L., Radichel, T. J., Clark, M. B., Sather, T. W., & Johnson, K. L. (2000). Volunteers, communication, and relationships: Synergistic possibilities.

- Topics in Stroke Rehabilitation*, 7(2), 1–9. <http://dx.doi.org/10.1310/4R0L-ULJM-MEP1-XBGK>
- Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning*, 28, 167–172. <http://dx.doi.org/10.1016/j.evalprogplan.2004.09.002>
- Karwalajtys, T., McDonough, B., Hall, H., Guirguis-Younger, M., Chambers, L. W., Kaczorowski, J., . . . Hutchison, B. (2009). Development of the volunteer peer educator role in a community cardiovascular health awareness program (CHAP): A process evaluation in two communities. *Journal of Community Health*, 34, 336–345. doi:10.1007/s10900-009-9149-5
- Kettner, P. M., Moroney, R. M., & Martin, L. L. (1999). *Designing and managing programs: An effectiveness based approach* (2nd ed.). Thousand Oaks, CA: Sage.
- Kotler, P., & Zaltman, G. (1971). Social marketing: An approach to planned social change. *Journal of Marketing*, 35, 3–12. doi:10.2307/1249783
- Maes, K. (2012). Volunteerism or labor exploitation? Harnessing the volunteer spirit to sustain AIDS treatment programs in urban Ethiopia. *Human Organization*, 71, 54–64. <http://dx.doi.org/10.17730/humo.71.1.axm39467485m22w4>
- Morrow-Howell, N., Kinnevy, S., & Mann, M. (1999). The perceived benefits of participating in volunteer and educational activities. *Journal of Gerontological Social Work*, 32(2), 65–80. doi:http://dx.doi.org/10.1300/J083v32n02_06
- Orton, S., Umble, K., Zelt, S., Porter, J., & Johnson, J. (2007). Management Academy for Public Health: Creating entrepreneurial managers. *American Journal of Public Health*, 97, 601–605. doi:10.2105/AJPH.2005.082263
- Pazoki, R., Nabipour, I., Seyednezami, N., & Imami, S. R. (2007). Effects of a community-based healthy heart program on women's physical activity: A randomized controlled trial guided by community-based participatory research (CBPR). *BMC Public Health*, 7(147), 216. doi:10.1186/1471-2458-7-216
- Posner, B. Z. (1995). What it takes to be a good manager. In J. R. Meredith & S. J. Mantel, (Eds.) *Project management: A managerial approach* (3rd ed., pp. 146–149). New York, NY: John Wiley & Sons.
- Reese, C. D. (2015). *Occupational health and safety management: A practical approach* (3rd ed., pp. 35–46). Boca Raton, FL: CRC Press.
- Rossi, C., Cricelli, L., Grimaldi, M., & Greco, M. (2016). The strategic assessment of intellectual capital assets: An application within Terradue Srl. *Journal of Business Research*, 69(5), 1598–1603.
- Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th ed.). Thousand Oaks, CA: Sage.
- Singh, D., Cumming, R., Mohajer, N., & Negin, J. (2016). Motivation of community health volunteers in rural Uganda: The interconnectedness of knowledge, relationship and action. *Public Health*, 136, 166–171. <http://dx.doi.org/10.1016/j.puhe.2016.01.010>
- Thornton, L. A. & Novak, D. R. (2010). Storying the temporal nature of emotion work among volunteers: Bearing witness to the lived traumas of others. *Health Communications*, 25, 437–448. doi:10.1080/10410236.2010.483340
- Topp, S. M., Price, J. E., Nanyangwe-Moyo, T., Mulenga, D. M., Dennis, M. L., & Ngunga, M. M. (2015). Motivations for entering and remaining in volunteer service: findings from a mixed-method survey among HIV caregivers in Zambia. *Human Resources for Health*, 13(72), 1–14. doi:10.1186/s12960-015-0062-y
- Torrey, W. C., Bond, G. R., McHugo, G. J., & Swain, K. (2012). Evidence-based practice implementation in community mental health settings: The relative importance of key domains of implementation activity. *Administration and Policy in Mental Health and Mental Health Services Research*, 39(5), 353–364.
- Trettin, L., & Musham, C. (2000). Using focus groups to design a community health program: What roles should volunteers play? *Journal of Health Care for the Poor and Underserved*, 11, 444–455.
- Turnock, B. J. (2009). *Public health: What it is and how it works* (4th ed.). Sudbury, MA: Jones and Bartlett.
- Walsh, D. C., Rudd, R. E., Moeykens, B. A., & Moloney, T. W. (1993, Summer). Social marketing for public health. *Health Affairs*, 104–119. doi:10.1377/hlthaff.12.2.104

CHAPTER 8

Monitoring Implementation Through Budgets and Information Systems

Two key elements exist behind the scenes that contribute to the success of a program and to the ability to evaluate a program. One element is the ever-important budget and the accompanying skills to develop and monitor the program's budget. The language and skills needed to create and monitor the *operating budget*, the budget that includes all the costs and revenues related to delivery of the program, are introduced. The other element, equally important, is the information system used by the agency for tracking and recording program participation and outcomes of participants. The informatics landscape continues to evolve rapidly, so only the basics as they relate to health program development and monitoring are presented.

► Budgets and Budgeting

Program managers need a mechanism for monitoring the fiscal status of the program during implementation. This requires understanding the

basic concepts related to budgeting and having basic costing and budgeting skills, allowing for the development of a sound fiscal plan for the health program. General budget principles that apply across budget formats and specific organizational requirements are presented in this chapter. The intent is to introduce basic program budget concepts, particularly as they relate to the operating budget, so that health program planners and managers can communicate with financial personnel, funding agencies, and administrators to present the program in a positive light.

Budgets have the appearance of being factual and neutral. The reality is otherwise. Budgets are inherently political, are contextually influenced, and reflect values and preferences of the program developers. One only needs to consider the annual debate over the federal budget to appreciate the politicized nature of budgets. The consequence of defunding health programs can have direct health effects, even for noncontroversial health programs. For example, preventive care

decreased and emergent care increased after an orthopedic trauma service for indigent patients was defunded (Forsh, Amanantullah, Coleman, & Wolinsky, 2014). Underfunding of projects can result in not meeting project goals (Hazzard, Moreno, Beall, & Zidenberg-Cherr, 2012). Political interference in program administration, despite wide public support, has implications for achieving population health goals (Petticrew & Krishnaratne, 2014). Such issues will likely weigh on the minds of health program planners and managers as they prepare and monitor their budgets.

Budgeting Terminology

The broadest categories within budgets are expenditures and revenues. Expenditures are classified as either fixed or variable and as either direct or indirect.

Fixed costs do not vary with the number of clients served. Examples of fixed costs include rent, salaries of administrative personnel, and insurance costs. In contrast, *variable costs* do vary with the number of clients served. Examples of variable costs include copying program handouts, program advertising, and refreshments for participants. Depending on how the health program is designed and implemented and how the organization does its accounting, certain costs may be counted as fixed or variable. For example, if program staff members are paid regardless of how many clients show up, then personnel wages and salaries represent a fixed cost. Conversely, if the program uses staff members on a part-time, as-needed basis, then personnel costs are variable. Budgets prepared based on the distinction between variable and fixed costs are more likely to be useful for program management via fiscal monitoring and for later conducting of economic and accounting analyses of the program.

Another way to think about costs is as direct or indirect. In the purest sense, *direct costs* reflect those resources that are used directly in the delivery of the program. Generally, the wages and salaries of staff members providing the intervention are a direct cost, as are materials

or supplies used with clients. Similarly, in the purest sense, *indirect costs* are those costs not associated with the delivery of the program but, more generally, with supporting the program. Utility bills, telephone charges, and legal and human resources staff are all examples of indirect costs. Indirect costs associated with *overhead expenses* (e.g., rent, utilities, facilities management, shared clerical support staff, office equipment) are typically estimated based on a standard rate that is set by the program funding agency or the organization's financial officer. Although the federal government can negotiate indirect costs as a percentage of direct costs to be over 50%, funders can also set much lower caps on this crucial source of support for general expenses such as facilities and administration. Given this wide range, it is important to obtain the correct rate at which indirect costs should be applied to the expenditure side of the budget.

In developing a budget for planning a program, Foster, Johnson-Shelton, and Taylor (2007) remind us that costs associated with time are very important. They estimated the costs associated with participant recruitment time, staff training time, intervention time of staff members and participants, and time of volunteers. The need for this level of detail can be anticipated and the actual expenditures tracked as the program is delivered.

A critical distinction exists between the cost of a service and the charge for that service. The *cost* of the service is the simple sum of all resources required to provide the service, including indirect costs. However, clients may be asked to pay more than the cost; *charges* typically include the cost plus a profit margin. When budgeting, the program planners must consider both the cost of the service and the charges, given that ultimately the charges influence participation and acceptance of the program.

Identifying revenues is a bit simpler. Funds for health programs come from grants from federal funding agencies or foundations, fees collected from program participants, reimbursements from third-party payers, or charitable fund raising. Revenues might also be matched from state or

federal agencies for local dollars allocated to the health program.

One source of revenue is often invisible: in-kind donations. Those services are provided to the program free of charge, but the program would have to pay for them if they were not donated. A common example is printing costs given as an in-kind donation; volunteer time is another in-kind donation for staff time. In some nonprofit agencies, the in-kind donations can be substantial. One study found in-kind donations to be a key resource during the program start-up phases (Tangka et al., 2008). It is important to track in-kind donations for two reasons. Funding agencies look at in-kind donations favorably as an indication of community support for the program. Also, if adequate in-kind donations are not received, a contingency plan for paying for those services must be developed and implemented.

Opportunity costs refer to purchases that cannot be made because of having spent the money on something else. In our daily lives, we instinctively calculate opportunity costs; for example, if I buy a ticket to fly home for the holiday, then I cannot buy a new smartphone. The opportunity costs accrued by program participants depend on the type of program. Two examples of opportunity costs used in cost analyses include loss of regular class time for a school-based health program delivered during the school day (Kilmer, Burgdorf, D'Amico, Miles, & Tucker, 2011), and parents' work loss due to a child's medical visits (Tai & Bame, 2011). Other examples of opportunity costs include losses or gains in work productivity, the family burden related to the health problem or program, out-of-pocket expenses related to participation, and cost of participating in the program (Drummond, Sculpher, Claxton, Stoddart, & Torrance, 2015). In terms of participants, because they are paying, in some form, for being in the program, they no longer have that money to spend on something else. For example, if a mother takes her child for immunizations, the cost of getting to the clinic and the copayment may cost her the opportunity (the money) to purchase new clothes for that child. Opportunity costs also

exist for the agency; for example, providing program A may mean that program B cannot be delivered. In summary, opportunity costs to the participants and the organization must be considered during the program planning stage.

Two other budget concepts deserve to be mentioned. *Discounting* is the process of converting future dollars, and future health benefits, into the present dollar value. It involves decreasing the current value by a rate, usually between 1% and 3%, on an annual basis. In a sense, discounting is the reverse of interest. Through this approach, expenses and health benefits that might be expected in 2 or 20 years can be valued on a par with one another. This facilitates interpretation of the dollar values and makes it simpler to see the costs so that decision makers can make choices based on data that are, so to speak, standardized. In practice, discounting is rarely used for operating budgets but is critical in budgets related to capital improvement or long-term financing.

One other concept related to costs is externalities. Secondary effects, whether anticipated or not, may occur, and the effects can spill over to individuals not participating in the program; these are called *externalities*. Externalities may be indirect or even unanticipated consequences of the program, whether beneficial or harmful. Including them can be important in comprehensive economic analyses that attempt to monetize a broad range of program-related effects, such as from the societal perspective. Externalities are identified by conducting a so-called thought experiment in which one imagines possible programmatic effects on the intended audience as a whole, regardless of participation. Identifying externalities is important because they then become effects (if beneficial) or costs (if harmful or in some way costly). Some externalities for participants may not be easy to monetize. For example, participation in a substance abuse program may cost the participant older friendships based on substance use activities. Being able to list and discuss the externalities can be important in justifying a program to various stakeholders.

► Budgeting as Part of Planning

Monetize and Compute Program Costs

Most costs are related to the program implementation and include the resources utilized by the program staff members and participants. Detailed program expenditures are used as the basis for computing the indirect and direct costs associated with providing the program. The process theory continues to be a guide for identifying which program costs to include in the budget (FIGURE 8-1) and for attaching a dollar

value to the inputs listed in the process theory. Costing out the elements of the process theory can feel odd or uncomfortable, but it must be done if the budget is to capture accurately and fully the costs in monetary terms.

Most program participants want to know what it will cost them to participate. The charge for the program or service is the most obvious cost. Other costs for program participants may exist, such as new or special equipment (e.g., handrails or cooking utensils), educational materials (e.g., books or magazine subscriptions), transportation, or child care while attending the program. Being able to provide this information to potential participants can be helpful in recruiting them into the program.

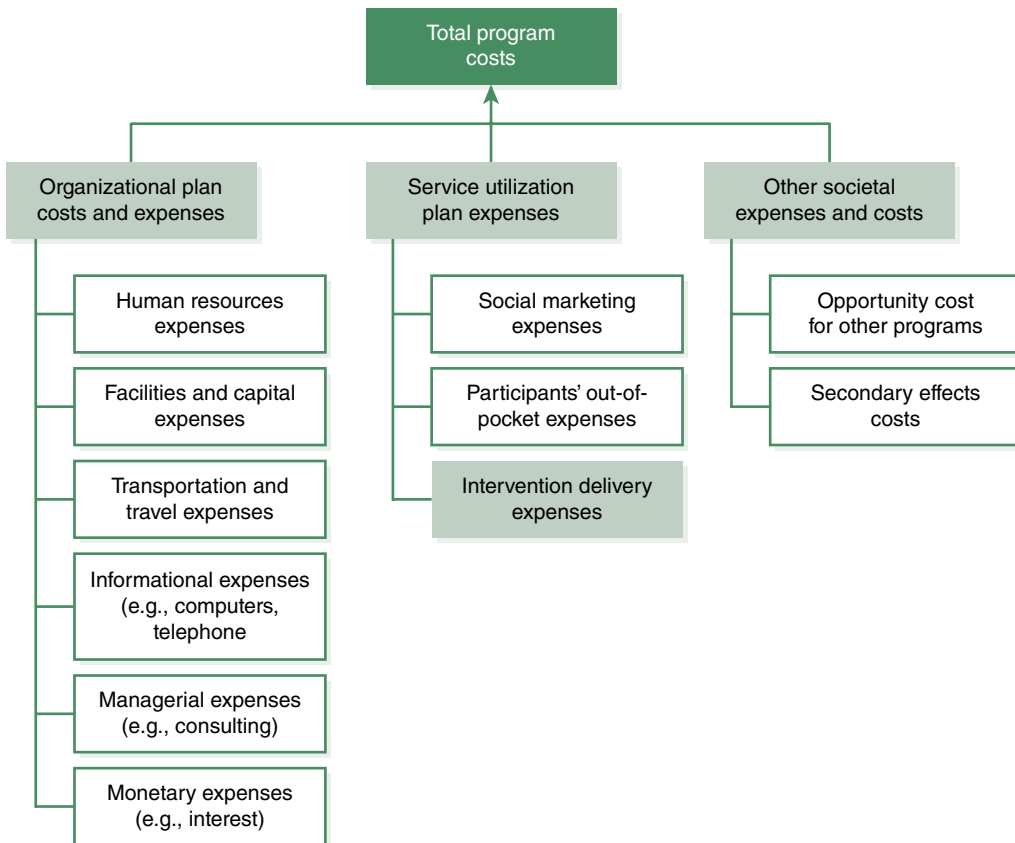


FIGURE 8-1 Relevance of Process Theory to Economic Evaluations

Budget for Start-Up and Evaluation Costs

For large-scale health programs that are being implemented for the first time, some costs will be related specifically to the start-up phase and scaling up. These costs can include the one-time purchase of large or small equipment, the extra time needed to find locations or negotiate contracts, or other one-time purchases needed to launch a program. In one study, the cost of scaling up a large adolescent sexual health program was approximately 20% of the total program costs (Terris-Prestholt et al., 2006). Cost effectiveness analyses of community-based rapid HIV testing included varying start-up costs across communities (Shrestha et al., 2008). Another example of start-up costs for large-scale programs relates to the 2014–2015 Ebola virus outbreak in West Africa. The outbreak resulted in the Centers for Disease Control and Prevention (CDC) recommending preparing 55 hospitals for Ebola care, at an average cost of over \$1 million per hospital, to become an Ebola treatment center (Herstein et al., 2016).

The budget must also include anticipated expenses related to the program evaluation. Whether the program staff members will be involved in the evaluation activities or whether a consultant is hired, funds must be allocated to support the evaluation before the program begins. Retrospectively acquiring funds to conduct an evaluation can be difficult. In general, program grant proposals and budgets that do not include evaluation funds receive lower-priority scores. Evaluation expenses generally fall into the same categories as program expenses, although material expenses are typically limited to supplies and copying costs. Incentives given to individuals to participate in the evaluation can account for a substantial portion of the evaluation budget. Personnel will usually be the largest expense. At a minimum, a meaningful evaluation cannot be done for less than 10% of the direct program costs.

Break-Even Analysis

After the program budget is complete and nearly final, it is possible to do a break-even analysis. A break-even analysis is the mathematical determination of the point at which the expenses related to providing the program are equal to or less than the revenues generated for or from the program. This type of analysis uses the price of the service (the charge), the variable costs of program, and the fixed costs of the program. The rather straightforward formula for a break-even analysis (Finkler, Ward, & Baker, 2007) follows:

$$\text{Quantity of services} = \text{fixed cost} / (\text{charge per client} - \text{variable cost per client})$$

When the total fixed costs associated with the program are divided by the difference between the amount charged per participant and the variable cost per participant, the result gives the number of services—in other words, per paying participant—needed to be provided to break even financially. When the break-even analysis is done for classes rather than for one-on-one services, the number of participants per class needs to be taken into account in calculating the variable costs. This can be a little tricky, but it is not complicated. The following exhibits show the step-by-step process. **EXHIBIT 8-1** gives a narrative scenario in which a break-even analysis is needed. The data from the narrative are then shown in a budget format in **EXHIBIT 8-2**. The formulas in **TABLE 8-1** are applied to the budget information, using a spreadsheet program, Microsoft Excel. Sometimes it is not immediately clear where the break-even point might be. **EXHIBIT 8-3** shows, in a table format, a range of the costs and revenues, with the break-even point noted.

Even programs provided by nonprofit agencies would be wise to conduct a break-even analysis. This rudimentary process provides useful insights into the amount of funding needed. Unfortunately, public agencies and public programs neglect their fiscal accountability and efficiency accountability all too often by not conducting

EXHIBIT 8-1 Example of a Scenario Needing a Break-Even Analysis

Layetteville's Lighthouse Agency has decided to implement an updated evidence-based program, Bright Light, to address self-care for elderly diabetes patients, based on the needs identified during the community health assessment. Operational fixed expenses for Bright Light are budgeted at \$1,500 per year.

Because Bright Light is an intensive educational program and Lighthouse tries to have a consistent ratio of teachers to clients, the number of teachers varies with the number of clients. For every 10 clients, Lighthouse employs 1 teacher, at a salary of \$300, but Lighthouse will not hold a class with fewer than 5 clients. Lighthouse can provide up to 1 class per month.

Based on this information, how many clients must be served, billed, and pay for Bright Light classes if the program is to break even? What recommendation would you make to Lighthouse?

The team created a budget (Exhibit 8-2) and applied the formulas (Table 8-1) using Excel. Conclusion: Assuming that the classes are full with 10 clients, Lighthouse loses money if it chooses Option A, but it makes a little money with Option B.

EXHIBIT 8-2 Example of a Budget Used for a Break-Even Analysis for Bright Light on an Excel Spreadsheet

A	B	C	D
1	Fixed costs (annual)—TOTAL		\$1,500
2	Rent	\$1,000	
3	Clerical support	\$100	
4	Cleaning service	\$100	
5	Telephone (base charge)	\$300	
6			
7	Revenue		
8	Charge per class	\$100	
9			
10	Variable costs per class—TOTAL		\$350
11	Teacher per class of 10	\$300	
12	Materials cost per student	\$50	

TABLE 8-1 Formulas Applied for Options A and B

	Option A	Option B	Using data from Exhibit 8-2
	7 classes	8 classes	
Fixed costs	\$1,500	\$1,500	= D1
Variable costs	\$5,600	\$6,400	= (number of classes × D10)
Expenses	\$7,100	\$7,900	= Fixed costs + variable costs
Revenue	\$7,000	\$8,000	= (number of classes × 10 students × C8)
Balance	(\$100)	\$100	= Revenue - expenses

a break-even analysis. If clients are not paying for services, as is often the case in public health programs or mass-media campaigns, the price an individual might be willing to pay for the service or other such information can be used in place of the charge or price. For example, if residents of a community are willing to pay only \$0.10 for information about how to prevent sexually transmitted diseases, then a safe-sex mass-media campaign will need to reach a very high number of persons theoretically to break even.

Thinking in terms of break-even analysis may seem unethical or contrary to the public health ethic. A break-even analysis would not be appropriate for health services that are required to ensure the health and safety of a population, such as a program to ensure that persons infected with tuberculosis take their medications or a national infant immunization program. In programs such as these, program participants are not expected to pay and thus the revenue portion of the break-even equation is zero. However, a break-even analysis could be conducted to understand the fiscal implications of these programs. In reality, use of a break-even analysis is a fiscally responsible way to make

decisions among programs or programmatic options. It also provides a quantifiable rationale for proceeding or for modifying a health program. A break-even analysis may reveal that additional funding is required to provide the program as intended; it is far better to identify this potential problem before the program is initiated.

Budget Justification

Budget justification is a requirement for almost all grant proposals, although the degree of detail expected varies by funding agency. A safe rule of thumb is to provide a very detailed budget justification; more detailed budget justifications demonstrate a more thorough program implementation and evaluation plan. Most budget justifications involve some narrative explanation of why the dollar amounts are requested, but they must also include fairly detailed arithmetic formulas that show the derivation of specific costs. For example, budget narratives typically show the cents per mile paid to staff members for travel, the estimated number of miles staff members are expected to travel, and the number of staff members traveling those miles. Even if a

EXHIBIT 8-3 Break-Even Table Shows Number of Paying Students Needed to Break Even*

E	F	G	H	I	J	K	L
	Number of Classes	Total Number of Students	Fixed Costs	Variable Costs	Total Costs	Total Revenue	Balance
			= D1	= (C11 × F17) + (C12 × G17)	= H17 + I17	= G17 × C8	= K17 - J17
17	1	10	\$1,500	\$800	\$2,300	\$1,000	(\$1,300)
18	4	40	\$1,500	\$3,200	\$4,700	\$4,000	(\$700)
19	4	45	\$1,500	\$3,450	\$4,950	\$4,500	(\$450)
20	5	50	\$1,500	\$4,000	\$5,500	\$5,000	(\$500)
21	5	55	\$1,500	\$4,250	\$5,750	\$5,500	(\$250)
22	6	60	\$1,500	\$4,800	\$6,300	\$6,000	(\$300)
23	6	65	\$1,500	\$5,050	\$6,550	\$6,500	(\$50)
24	7	70	\$1,500	\$5,600	\$7,100	\$7,000	(\$100)
25	7	75	\$1,500	\$5,850	\$7,350	\$7,500	\$150
26	8	80	\$1,500	\$6,400	\$7,900	\$8,000	\$100
27	8	85	\$1,500	\$6,650	\$8,150	\$8,500	\$350
28	9	90	\$1,500	\$6,900	\$8,400	\$9,000	\$600

* Using data from Exhibit 8-2.

health program is being sponsored by the parent organization, a budget justification is typically presented to departmental administrators or advisory boards when requesting their support.

Most grant proposal budgets focus on the major categories of direct costs. However, some federal funding agencies have begun to ask for budgets that are linked more directly to the program objectives, while others ask for budgets broken out by levels in the public health pyramid. Such budgets enable the funding agencies and program managers to determine the merit of the budget in terms of what is planned and which outcomes are anticipated. While creating such a budget can be challenging and requires some degree of speculation, assigning costs per program objective can be a powerful motivational and managerial tool.

► Budget as a Monitoring Tool

A budget developed during the planning stage becomes a tool to monitor implementation after the program has started. In addition, the careful attention to tracking expenses and income can be helpful in doing different types of cost analyses as part of a monitoring and subsequent planning cycle. The main approach to using a budget as a monitoring tool is tracking the budget variance.

Budget Variance

On an ongoing and regular basis (usually monthly), the program manager should determine the extent to which current expenditures exceed (or not) the projected program expenditures. The difference between the budgeted and actual expenditures or income is called the *budget variance*.

The variance is calculated for each category or line item as a simple subtraction of expenditures from the budgeted amount, which can be easily calculated with the use of spreadsheets. When the program budget is developed during the planning stage, it would be wise to have the

spreadsheet set up so that makes it easy to include columns for variance (**EXHIBIT 8-4**). Because of delays associated with billing, timing of organizational fiscal reports, and the various dates selected by the organization at the beginning of a fiscal year, the actual expenses and income will always be based only on what is available at any given time. As bills are posted and income recorded, the expenditures and income must be updated.

Severe negative variances can occur because of overexpenditures or inadequate income, whereas severe positive variances can arise due to underexpenditures or greatly increased income. Either way, the presence of a significant variance alerts program managers that the program may not be delivered as planned and that some aspect of the program may need further scrutiny and modification.

Monthly updating of the budget variance usually suffices. As the end of a grant-funded program nears, at which time the balance needs to be near zero, more frequent updating would be wise. During the program implementation, the overall variance ought to be no more than 10% to 20% of the projected budget for an expense category. Some funding agencies specify the degree of budget variance that is acceptable and indicate at what point the budget needs to be renegotiated with the funding agency. Program managers, by monitoring the budget variance on at least a monthly basis, can make needed adjustments to spending throughout the fiscal year so that the year-end variance is within an acceptable range.

Types of Cost Analyses

Several different types of analyses related to program costs are possible. The four main types of economic analyses are briefly reviewed as an introduction to the various approaches. The types can be classified along two dimensions (Drummond et al., 2015). One dimension is whether one or more programs are under consideration, and the other is whether costs only or costs and effects are included in the analysis (**EXHIBIT 8-5**).

EXHIBIT 8-4 Example of a Budget Showing Year-to-Date Variance							
A	B	C	D	E	F	G	H
Category	Total Budget	Cumulative Expenses	Expected	Dollar Variance	Amount Left to Spend	Percentage Unspent	Percentage Variance
			b X (number of months into FY/12)	= d - c	= b - c	= f/b	= (c/d)/100
1	Salaries (professionals)	\$77,174	\$76,474	\$(700)	\$24,791	24%	1.0%
2	Assistant salaries	\$6,878	\$6,962	\$(84)	\$2,405	26%	1.0%
3	Fringe benefits	\$24,454	\$24,358	\$96	\$8,023	25%	1.0%
4	Fixed Subtotal	\$108,506	\$107,794	\$(688)	\$35,219	25%	
5							
6	Materials/supplies	\$1,164	\$1,200	\$(36)	\$436	27%	1.0%
7	Travel	\$2,357	\$2,250	\$107	\$643	21%	1.0%
8	Services	\$578	\$600	\$(22)	\$222	28%	1.0%
9	Printing	\$838	\$905	\$(67)	\$369	31%	0.9%
10	Subcontracts	\$1,689	\$3,314	\$(1,625)	\$2,730	62%	0.5%

11	Equipment	\$1,500	\$2,637	\$1,125	\$1,512	\$(1,137)	-76%	2.3%
12	<i>Variable Subtotal</i>	\$12,526	\$9263	\$9394	\$(131)	\$3263	26%	0.0%
13								
14	Indirect costs (12%)	\$18,750	\$13,996	\$14,063	\$(67)	\$4,755	25%	
15								
16	EXPENDITURE TOTAL	\$175,000	\$131,764	\$131,250	\$(886)	\$43,236	25%	0.0%
17								
18	Category	Total Budget	Cumulative Revenue	Expected	Dollar Variance	Amount Left to Generate	Percentage Unfunded	Percentage Variance
19	Fees	\$78,500	\$59,583	\$58,875	\$708	\$18,917	24%	1.0%
20	Donations	\$6,500	\$3,928	\$4,875	\$(947)	\$2,572	40%	0.8%
21	Grant funds	\$90,000.0	\$75,000.0	\$67,500	\$7,500	\$15,000	17%	1.1%
22	REVENUE TOTAL	\$175,000	\$138,511	\$131,250	\$7,261	\$36,489		
23								
24								
	BALANCE		\$6,747					

EXHIBIT 8-5 Types of Cost Analyses

	One Program	Two or More Programs
Costs Only	Cost description Cost analysis Cost minimization	Cost comparison
Costs and Effects	Cost analysis	Cost effectiveness Cost–benefit Cost utility

Classifying economic analyses of programs along these two dimensions helps to discern the types and the subsequent requirements for conducting each type of analysis.

Cost-Only Analyses

When only one program is considered and effects are not included, the type of analysis is termed a *cost description*. Cost description is the simplest form of economic analysis in that it is a straightforward presentation of expenses related to the delivery of the health program. Most program managers routinely prepare cost descriptions, particularly for any annual reports that require an accounting of expenses by category or line item. A cost description is best thought of as part of process monitoring, particularly with regard to accountability and the budget aspect of the organizational plan. However, more sophisticated cost analyses might focus on a specific stage in the life cycle of a program, such as the start-up stage. Tangka et al. (2008) did just that by estimating the costs associated with starting a colorectal cancer program.

When the cost description includes a breakdown of total expenses by an analytic factor, such as time periods, staff activities, or funding source, it becomes a *cost analysis*. A cost analysis is also useful as a process monitoring tool. It essentially analyzes costs by other elements of either the

organizational plan or the service utilization plan. In this way the findings of a cost analysis begin to provide information about the efficiency of a program, using, for example, dollars spent per program participant. These findings can be compared to published reports, benchmarks, or the original program plan to interpret the extent to which the program is more or less efficient than similar programs. Cost analyses can also focus on the costs or savings that result from the health program. For example, Lairson, Huo, Ball Ricks, Savas, and Fernández (2013) used call logs and additional expenditure data to quantify the staffing and material costs associated with the addition of conducting cancer screenings to existing 211 help center services.

Cost minimization is an analysis to determine the best ways to provide the program at the lowest cost. This type of analysis ought to be part of the basic budgeting process because it allows program managers to select the lowest reasonable expense for a category or to comparison-shop for expenses such as supplies and capital equipment.

Published cost descriptions and cost analyses can be used during the planning stage to estimate program expenses for different intervention options. Clearly, cost descriptions and cost analyses are important for process monitoring and for understanding the basic cost parameters of a program. The basic information generated through these analytical approaches is easy for

program staff members and stakeholders to understand, so it facilitates decision making about the program's future.

Costs and Effects Analyses

Cost comparison compares the costs of two or more programs without looking at outcomes or impacts. These comparisons may focus on the costs per participant for the programs being considered or the revenues generated by various programs. Cost comparisons might be done by a single agency with multiple health programs as the basis for deciding which program to continue. For example, Jerrell and Hu (1996) reported on a comprehensive analysis of costs associated with clients in three types of mental health and substance abuse prevention programs administered by one agency: a 12-step program, a behavioral skills program, and case management. An agency might undertake a cost comparison if it provided different intervention programs for the same health problem and needed to eliminate one of the programs. Cost comparisons can include standard practice in comparison to a new program. Cost comparisons can also be done via simulations (e.g., Yonas et al., 2013), which do not require actual program implementation or actual cost data.

Cost-effectiveness analysis always compares the costs of two programs against one type of impact that is measured the same way in both programs. Using this approach, the programs are then compared on the basis of cost per unit of outcome. During the program planning phase, a review of cost-effectiveness reports could help program planners decide which interventions to implement. After the program has been implemented, program personnel can perform a cost-effectiveness analysis, with some help from a program evaluator. As with cost comparisons, simulations can be conducted to compare both the costs and various potential outcomes of alternative programs and implementation choices. Among this set of approaches, system dynamics modeling examines feedback loops that can lead to unintended effects (Lich, Urban, Frerichs, &

Dave, 2017). Simulation can yield insights into direct and indirect positive and negative potential effects of alternatives before starting a new program. This can help decision makers avoid costly mistakes in program selection and implementation. However, developing such models requires specialized expertise and labor-intensive reviews of prior research, potential new data analyses, consultation with informed stakeholders, and iterative programming and stakeholder review. Hence, the practical application of these methods for most health program evaluators may be using previous simulations conducted in reasonably comparable contexts.

Like a cost-effectiveness analysis, a *cost-benefit analysis* compares two programs. In this type of analysis, however, the programs need not address the same health problem. The program effects are compared based on larger societal benefits in addition to the outcomes and impacts of the program. The two dissimilar programs are compared on the basis of cost per dollar value of benefits achieved. Often the program under consideration is compared to a do-nothing option. Again, a review of existing cost-benefit analyses can help in choosing a program during the planning stage.

A *cost-utility analysis* measures the outcome of health programs in terms of the potential participants' preference for the health outcome. This makes it the most complex and theoretical economic analysis. Programs are compared on the basis of their cost per unit of preference, called utility. For all practical purposes, only researchers perform such calculations and cost-utility analyses.

► Information Systems

Since 2009, several federal acts and initiatives have been passed that are dramatically changing the health information landscape and will require the ongoing attention of public health program managers. The strategic goals of these pieces of legislation, essentially, are to have: (1) a nationally adopted standard terminology

for health data elements, (2) secure mechanisms for electronically sharing and transmitting health information, and (3) a data repository that can be used to monitor in real-time health events and trends. Each organization has its own information system, but for many health programs, that system needs to interface with the information systems of the funder and regulatory bodies. The following introduction to health informatics and public health informatics forms a base for understanding the issues and processes relevant to health program managers and developers.

Health Informatics Terminology

Knowing how to use a computer and the Internet is not the same as understanding health informatics. In the broadest sense, an *information system* encompasses both the hardware and the software used to collect, store, and exchange digital information for a work unit, whether it is a healthcare organization or a utilities company. *Information technology* refers the products used within the information system, such as Microsoft Office and Ethernet Internet connections. Most organizations have a *management information system* (MIS), with a subset of the information system designed to address managerial, fiscal, and human resources processes and data. Appointment scheduling software might be a component of the MIS for a healthcare organization, as would be any accounting software.

Geographic information systems (GISs) are used increasingly to collect, analyze, and display public health data using maps. John Snow, who mapped new cases of cholera in London in 1854, was essentially the first to use GIS in public health. With the advance of software specifically designed to map data points, GIS can be used at each phase of the program planning cycle. For example, GIS has been applied to tracing environmental health risk exposure, disease spread, and primary care access (Craglia & Maheswaran, 2016). ZIP codes are another convenient unit for mapping both participation and need and for assessing the correlation between the two (Hsu et al., 2010).

Within the healthcare industry, *health information technology* (HIT) broadly encompasses all of the technology in use within healthcare organizations, from computerized imaging to tweets of health messages. Until the passage of the American Recovery and Reconstruction Act (ARRA) in 2009, focus centered on medical technology linked to imaging or devices or on the documents related to the medical record. The ARRA provided financial incentives for physicians and healthcare organizations to speed up the adoption and use of HIT. Thus, emphasis has shifted from the *electronic medical record* (EMR), the more traditional paper medical record maintained by a physician, to the *electronic health record* (EHR), which has been defined as the longitudinal electronic record of a patient's health information generated by one or more encounters in any care delivery setting (Healthcare Information and Management Systems Society [HIMSS], 2016).

If the health program is connected to a hospital or a clinic and the participants are patients of those healthcare organizations, then some form of an EHR exists. A concern from the perspective of a health program is the potential need to interface with the existing EHR. If the health program is based in an agency or organization that does not provide direct medical services, however, such as a community parks district or a church, then linkages to an EHR may not be of concern.

Information Systems Considerations

Throughout the program planning and evaluation cycle, the information system plays an important role as the means to retrieve existing or collect new data, store data and other program data elements, and then utilized the data elements in the form of reports. **FIGURE 8-2**, a simplified representation of an information system for a health program, shows varying arrow sizes to indicate the major direction of flow of data. At the center of the information system is a

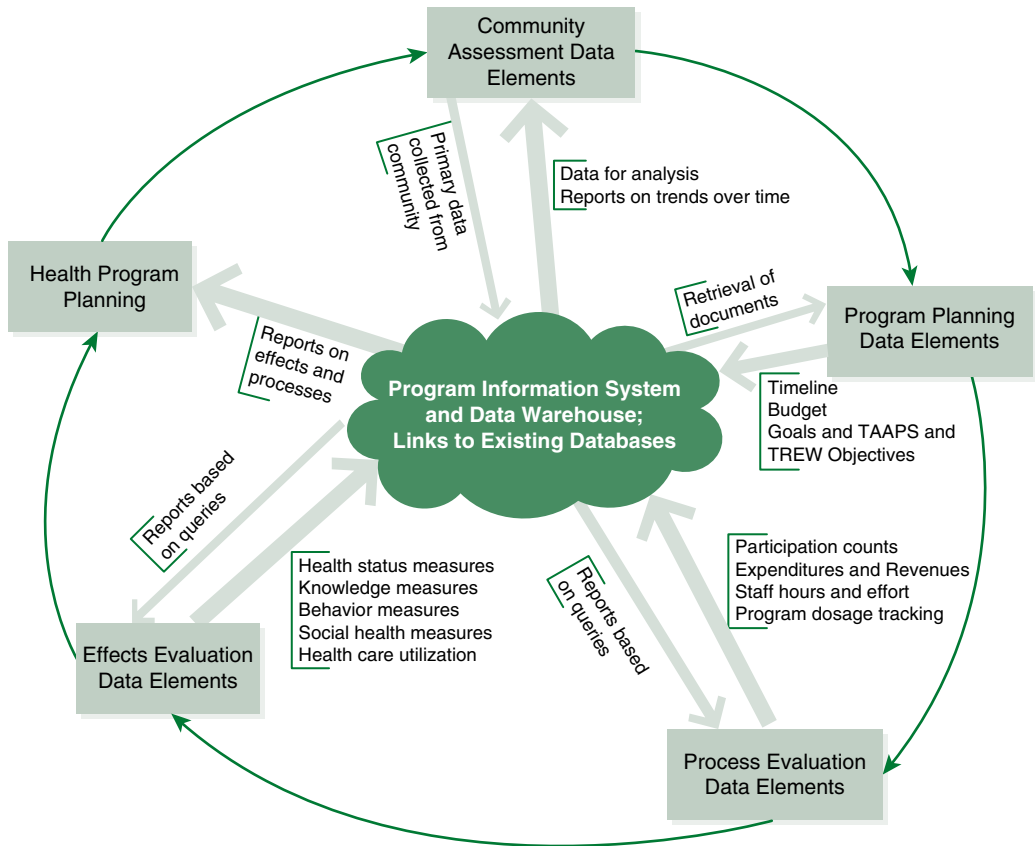


FIGURE 8-2 Information System Processes Throughout the Program Planning Cycle

computer server or data warehouse, implying that the data submitted and retrieved are compatible with the various types of hardware and software used throughout the organization. In larger healthcare organizations, the issues of ensuring compatibility fall within the domain of the information officer. In smaller organizations providing health programs, the program manager may be the one facing compatibility problems. Nonetheless, informatics systems can be used to generate data for making program decisions (Paneth-Pollak et al., 2009).

Developing and using a *data dictionary*, a list of the names of data elements to be collected, with definitions of those data elements, helps ensure a standardized definition and consistency in use (American Health Information Management

Association [AHIMA] e-HIM Work Group on EHR Data Content, 2006). Once the program objectives have been established, then a data dictionary can be created for the data elements needed to conduct the process and the effect evaluations. Thus, data elements in the data dictionary for the program range from specific expenditure items to participant demographic characteristics. For example, if participant race/ethnicity is a data element, the dictionary defines the categories of race/ethnicity. The task of creating the data dictionary can feel tedious and mundane, but having it facilitates communication with the information officer or technology staff members, and minimizes having unreliable or missing data.

As the availability and flexibility of handheld electronic devices increases, so will their usefulness

in monitoring the implementation of a health program. The options and potential applications seem endless, from using cell phones for texting reminders about health behaviors to providing links to existing web-based health information. Mobile health (mHealth) and wearable health information sensing devices allow for continuous monitoring at the individual and population levels (Kumar et al., 2013). These can be particularly helpful to programs that are not connected to EHRs. In addition, social media websites can be used to recruit participants and to disseminate key health information across a population.

Health organizations continue to expand the use of a web-based system for entering data in real time via a form at a secure website. Online surveys have nearly replaced pen-and-paper questionnaires. Web-based systems have greater flexibility for generating multiple outputs. For example, a web-based system for tracking the amount of time staff members spent on given tasks can be used to determine the true costs of staff activities (James et al., 2011). Mobile healthcare applications are becoming increasingly popular for monitoring and delivering healthcare interventions as well as collecting data. Advanced applications include monitoring vital signs and blood glucose levels (Baig, GholamHosseini, & Connolly, 2015). Program managers need to decide whether the size of the health program warrants the development of a web-based data collection system or whether use of paper forms is more appropriate given the intended program audience. The decision must fit the anticipated number of participants, the planned duration of the intervention program, and available resources.

Data entry issues are substantial. Accuracy at the point of data entry is paramount. The adage of “garbage in, garbage out” holds so true for health information. The responsibility for ensuring accurate data entry falls to the program manager. Accuracy depends on entering the data as intended and not having the cursor move, which would cause an incorrect answer to

be selected from a drop-down menu. Accuracy also depends on everyone entering data having the same understanding and definitions of the data being entered—for example, not having two people interpret “mostly followed instructions” in different ways.

Backing up data must not be ignored. Following organizational procedures for saving and backing up data ensures that the data will not be lost. For small programs, the program manager needs to establish that policy for a frequent routine and then monitor whether the policy has been strictly followed by staff members.

► Across the Pyramid

At the direct services level of the public health pyramid, participants experience the cost of the program, which becomes program revenue. Thus, good budgeting and creative generation of revenue from other sources may influence the level of participation because it reduces the out-of-pocket cost to participate in the health program. Program participants should not be distracted by the presence of an information system or the data entry of their information into such a system. Programs may be intentionally focused on increasing the individual’s use of his or her own EHR as a health awareness and health promotion tool.

At the enabling services level, as with the direct services level, budgeting and information systems should be transparent and not noticeable. Budgeting for programs at the enabling services level includes the same key budgeting categories as with a direct services health program. Information systems at this level of the pyramid are more difficult to conceptualize and implement, and thus rely heavily on data from individuals and families receiving the enabling services.

At the population services level, budgets tend to be those of the state or federal governments and earmarked for programs that serve the population. Large healthcare systems, such as Kaiser Permanente or the U.S. Department

of Veterans Affairs, may also have budgets for programs that address their enrollees. Nationally, various efforts are underway to utilize the EHR along with other data sources to conduct public health surveillance for sentinel events foreshadowing possible communicable disease outbreaks.

This chapter has dealt mostly with the infrastructure level of the public health pyramid. Being fiscally responsible in planning and monitoring a program has ramifications for receiving subsequent program funding and for being viewed as credible by the program stakeholders.

► Discussion Questions and Activities

1. Speculate how fixed and variable costs might change for a health program at the direct services level, the enabling services level, or the population services level of the public health pyramid. Discuss the implications of these changes on the results of a break-even analysis.
2. Imagine that Layetteville's Lighthouse Agency has learned that the class leader will be available for only 6 months. It now needs to adjust its budget so that it will break even after the sixth class. It remains committed to having no more than 10 students per class. How would you go about generating a recommendation for the agency?
3. Create a data dictionary with at least seven data elements related to program implementation and five data elements related to the participants of an adult immunization health program being provided through community-based clinics. Which data elements were most difficult to define and why?
4. Imagine that your healthcare organization is starting a health program for diabetes prevention. You have been charged

with leading a team to address the health information technology issues related to that program. What steps would you take? What would be your top three concerns?

► Internet Resources

There is a wealth of online resources on developing budgets and data systems. The following list represents diverse perspectives and approaches.

Break-Even Analysis

Online and software calculators make break-even calculations much simpler. Dinkytown.net offers free trials for its financial calculators with options to purchase. Microsoft Excel comes with a built-in financial calculator (http://www.tvmcalcs.com/index.php/calculators/excel_tvm_functions/excel_tvm_functions_page1).

Public Health and Health Informatics

For those with an interest in gaining an overview of the complex world of public health informatics, the CDC offers an introductory course on this subject (<http://www.cdc.gov/publichealth101/informatics.html>). In addition, the Health Information and Management Systems Society (HIMSS) provides a wealth of information, including a resource library and professional development opportunities (www.himss.org).

Budget Justification

Generating the budget justification for a grant is simplified by using the checklist and examples developed by the University of Colorado Boulder (<http://www.colorado.edu/ocg/sites/default/files/attached-files/Budget%20Justification%20Checklist%2011%202014.pdf>). An example of a national budget justification is presented by the US Department of Health and Human Services (HHS) for the 2017 fiscal year on the public health and social services emergency fund (<http://www.hhs.gov/sites/default/files/fy2017-budget-justification-phssef.pdf>).

References

- American Health Information Management Association (AHIMA) e-HIM Work Group on EHR Data Content. (2006). Guidelines for developing a data dictionary. *Journal of AHIMA*, 77(2), 64A–64D.
- Baig, M. M., GholamHosseini, H., & Connolly, M. J. (2015). Mobile healthcare applications: system design review, critical issues and challenges. *Australasian Physical & Engineering Sciences in Medicine*, 38(1), 23–38. doi:10.1007/s13246-014-0315-4
- Craglia, M., & Maheswaran, R. (2016). *GIS in public health practice*. Boca Raton, Florida: CRC Press.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford, UK: Oxford University Press.
- Finkler, S. A., Ward, D. M., & Baker, J. J. (2007). *Essentials of cost accounting for health care organizations* (3rd ed.). Sudbury, MA: Jones and Bartlett.
- Forsh, D. A., Amanantullah, D. F., Coleman, S., & Wolinsky, P. R. (2014). The financial impact of the loss of county indigent patient funding on a single orthopedic trauma surgery service. *Journal of Trauma and Acute Care Surgery*, 76(2), 529–533.
- Foster, E. M., Johnson-Shelton, D., & Taylor, T. K. (2007). Measuring time costs in interventions designed to reduce behavior problems among children and youth. *American Journal of Community Psychology*, 40, 64–81. doi:10.1007/s10464-007-9125-0
- Hazzard, E. L., Moreno, E., Beall, D. L., & Zidenberg-Cherr, S. (2012). An evaluation of the California Instructional School Garden Program. *Public Health Nutrition*, 15(2), 85–90. doi:10.1017/S1368980011001248
- Healthcare Information and Management Systems Society (HIMSS). (2016). Electronic Health Record (EHR). Retrieved from www.himss.org/ASP/topics_ehr.asp
- Herstein, J. J., Biddinger, P. D., Kraft, C. S., Saiman, L., Gibbs, S. G., Smith, P. W., . . . Lowe, J. J. (2016). Initial costs of Ebola treatment centers in the United States. *Emerging Infectious Diseases*, 22(2), 350. doi:10.3201/eid2202.151431
- Hsu, C. E., Watson, K., Boekeloo, B., Shang, N., Metzger, C., & Downer, G. (2010). The use of emerging informatics techniques to evaluate the delivery of NMAETC programs to address HIV and AIDS disparities. *Journal of the National Medical Association*, 102(12), 1116–1122.
- James, P., Bebee, P., Beekman, L., Browning, D., Innes, M., Kain, J., . . . Waldinger M. (2011). Creating an effort tracking tool to improve therapeutic cancer clinical trials workload management and budgeting. *Journal of the National Comprehensive Cancer Network*, 9(11), 1228–1233.
- Jerrell, J. M., & Hu, T. W. (1996). Estimating the cost impact of three dual diagnosis treatment programs. *Evaluation Review*, 20(2), 160–180.
- Kilmer, B., Burgdorf, J. R., D'Amico, E. J., Miles, J., & Tucker J. (2011). Multisite cost analysis of a school-based voluntary alcohol and drug prevention program. *Journal of Studies on Alcohol and Drugs*, 72(5), 823–832.
- Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., . . . Swendeman, D. (2013). Mobile health technology evaluation: The mHealth evidence workshop. *American Journal of Preventive Medicine*, 45(2), 228–236. doi:http://dx.doi.org/10.1016/j.amepre.2013.03.017
- Lairson, D. R., Huo, J., Ball Ricks, K. A., Savas, L., & Fernández, M. E. (2013). The cost of implementing a 211 call center-based cancer control navigator program. *Evaluation and Program Planning*, 39, 51–56. doi:http://dx.doi.org/10.1016/j.evalprogplan.2013.04.001
- Lich, K. H., Urban, J. B., Frerichs, L., & Dave, G. (2017). Extending systems thinking in planning and evaluation using group concept mapping and system dynamics to tackle complex problems. *Evaluation and Program Planning*, 60, 254–264. http://dx.doi.org/10.1016/j.evalprogplan.2016.10.008
- Paneth-Pollak, R., Schillinger, J. A., Borrelli, J. A., Handel, S., Pathela, P., & Blank, S. (2009). Using STD electronic medical record data to drive public health program decisions in New York City. *American Journal of Public Health*, 100(4), 586–590. doi:10.2105/AJPH.2009.175349
- Petticrew, M., & Krishnaratne, S. (2014). The Fag Lady, revisited: Margaret Thatcher's efforts on behalf of the tobacco industry. *Public Health*, 128(10), 904–910.
- Shrestha, R. K., Clark, H. A., Sansom, S. L., Song, B., Buckendahl, H., Calhoun, C. B., . . . Heffelfinger, J. D. (2008). Cost-effectiveness of finding new HIV diagnoses using rapid HIV testing in community-based organizations. *Public Health Reports*, 123 (suppl 3), 94–100.
- Tai, T., & Bame, S. I. (2011). Cost-benefit analysis of childhood asthma management through school-based clinic programs. *Journal of Community Health*, 36(2), 253–260. doi:10.1007/s10900-010-9305-y
- Tangka, F. K., Subramanian, S., Bapat, B., Seeff, L. C., DeGroff, A., Gardner, J., . . . Royalty, J. (2008). Cost of starting colorectal cancer screening programs: Results from five federally funded demonstration programs. *Prevention and Chronic Disease*, 5(2), A47.
- Terris-Prestholt, F., Kumaranayake, L., Obasi, A. I., Cleophas-Mazige, B., Makokha, M., Todd, J., . . . Hayes, R. J. (2006). From trial intervention to scale-up: Costs of an adolescent sexual health program in Mwanza, Tanzania. *Sexually Transmitted Diseases*, 33(10), S133–S139. doi:10.1097/01.olq.0000200606.98181.42
- Yonas, M. A., Burke, J. G., Brown, S. T., Borrechach, J. D., Garland, R., Burke, D. S., & Grefenstette, J. J. (2013). Dynamic simulation of crime perpetration and reporting to examine community interventions. *Health Education and Behavior*, 40(10), 87S–97S. doi:10.1177/1090198113493090

CHAPTER 9

Implementation Evaluation: Measuring Inputs and Outputs

Once the health program has started, stakeholders, funding agencies, and program staff members want to know if the program was implemented successfully and as planned. Answering this question becomes possible by devoting attention to its implementation. This chapter covers techniques and issues related to documenting, monitoring, and evaluating the implementation of the program. These topics are addressed in relationship to the components of the program theory, including the process of creating data that can be used to assess the achievement of the process objectives established during the planning phase. Documenting and assessing the extent to which the process objectives have been achieved, and at what cost, are important aspects of providing a program. Maintaining a focus on the quality and quantity of implementation helps identify gaps between program accomplishments and process objective targets. As with the development of the program interventions and plan, obtaining input from recipients and stakeholders while gathering

and interpreting data enriches the understanding about the health program implementation.

► Assessing the Implementation

Questions that focus on elements of the organizational and service utilization plans are essentially questions about implementation, rather than about effects of the program. These questions tend to fall into one of three levels of sophistication in regard to program implementation. Although slightly different in focus and emphasis, all three categories of questions are concerned with assessing the elements of the process theory portion of the program theory (**FIGURE 9-1**). Thus, the measures and data collection for each are likely to be quite similar—in fact, depending on the programmatic circumstances, they may even be the same. Therefore, this chapter covers these three types of questions in an integrated manner.

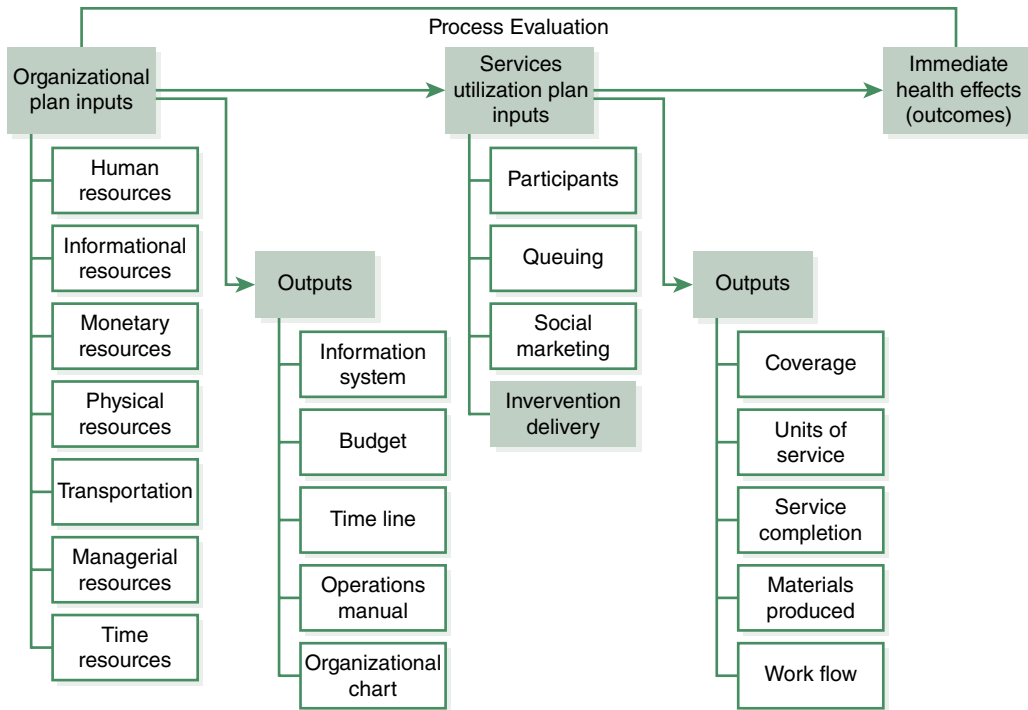


FIGURE 9-1 Elements of the Process Theory Included in a Process Evaluation

Unfortunately, terms describing implementation-focused questions are not used consistently in the literature. The following description of the three levels of questions provides a framework for understanding the various terms, such as process monitoring, program monitoring, process evaluation, and implementation monitoring.

Implementation Documentation

Implementation documentation refers to the simple tallying of activities and processes carried out as implementation activities of the program. Questions of this sort represent the simplest level of sophistication. As simple documentation, this line of questioning does not lead to subsequent interpretations of what was documented and therefore does not prompt program managers to take actions to change or improve the program.

Implementation documentation certainly involves collecting the data specified in the process objectives. Accurate, complete, and timely implementation documentation is the foundation for the next levels of questions about implementation. In other words, although implementation documentation is necessary, it is not sufficient.

Implementation documentation is carried out in part to meet the requirements of funding agencies—specifically, to demonstrate the extent of program implementation. Required reporting often entails a predetermined set of data that will need to be collected and used in the report to the funding agency. Thus, not all aspects of implementation documentation are under the control of the program administrators or evaluators. The data collection requested by the funding agency will ideally be consistent with and built into the process objectives.

Implementation Assessment

The next level of sophistication focuses on implementation assessment. *Implementation assessment* is the ongoing, nearly real-time activity of collecting data about the implementation of the program for the purpose of making timely corrections or modifications to the implementation through changes to elements of the process theory. Implementation assessment is called either program monitoring or process monitoring.

Process monitoring, the term used in this text to refer to this level of implementation assessment, is the ongoing, real-time assessment of the implementation of the program. It is integrated into the implementation of the program as a managerial or oversight tool to ensure that the intervention is being delivered within the parameters established during the planning stage. Ideally, process monitoring could lead to a favorable subsequent implementation evaluation.

Program monitoring is concerned with the elements described in this chapter. In addition, some of its aspects are similar to quality improvement techniques. Each program is associated with a set of process objectives, and the implementation assessment activities are focused on achieving those objectives. This linkage between the implementation assessment activities and the process objectives keeps the monitoring activities reasonable in scope and focused on key or critical processes for having a successful program.

One purpose of implementation assessment is to provide managerial guidance and oversight of the implementation of the program. Program monitoring, like the quality improvement methodologies, can provide data on which to base changes or corrections in the delivery of the program. The interactive and iterative nature of planning and implementing a health program inevitably requires some degree of flexibility, particularly for programs that are continuous, ongoing, or repeated. This state of flux may be why community coalition program monitoring systems sometimes need to be modified (Chalmers et al., 2003).

The process monitoring information can inform decision making regarding which aspects of the organizational plan or the service utilization plan are ineffective in accomplishing the process objectives. The actualization of this flexibility and the willingness to make changes amount to corrective managerial practice. Through such actions, program managers address issues of accountability and quality. In addition, process monitoring can function as a warning system for problems in delivery of the program and can provide a basis for modifying the process theory for subsequent revisions of the program.

Implementation Evaluation

The most sophisticated level of questioning comprises implementation evaluation. *Implementation evaluation* is a comprehensive, retrospective determination of the extent to which the program was delivered as designed and whether the variations might have had significant or important implications for the program effects. Implementation evaluation is generally called process evaluation. *Process evaluation* entails systematic research to assess the extent to which the program was delivered as intended. Thus, it is the systematic examination of programmatic coverage and delivery. Patton (2008, p. 324) defined process evaluation as focusing on “the internal dynamics and actual operations of a program in an attempt to understand its strengths and weaknesses.” Although the term is not used consistently in the literature, *process evaluation* is used in this text to refer to implementation evaluation.

Process evaluation has at least two purposes. One purpose is to gather data about the delivery of the interventions so that the results of an effect evaluation can be interpreted within the context of the program delivery. That is, if the intervention does not have effects on the health problem, it will be important to verify that the program was in fact delivered and to quantify the degree of intensity and faithfulness to the interventions design. Process evaluations are intended to demonstrate that program

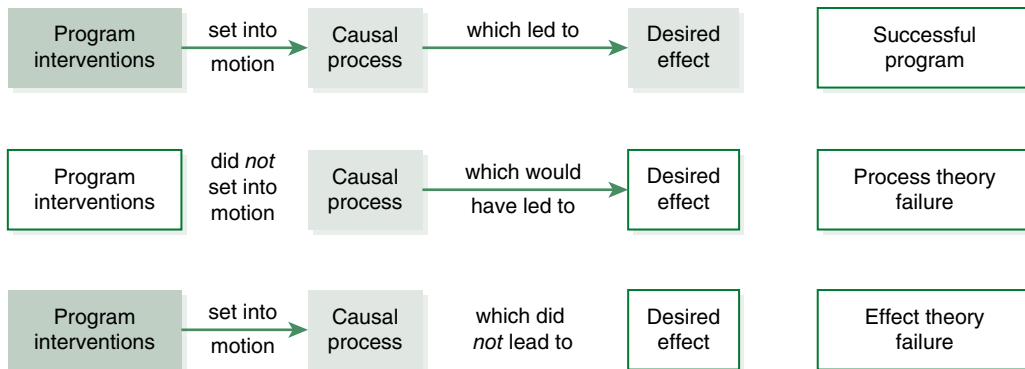


FIGURE 9-2 Roots of Program Failure

Weiss, C. H. (1998). *Evaluation*, (2nd ed.). Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.

specifications are met and, as such, are useful for ensuring that the work being done by program staff members is consistent with the program's process objectives and plan. Process evaluations help identify whether the implementation of the program contributed to the program's failure, as distinct from whether the effect theory was incorrect (**FIGURE 9-2**).

The second purpose of process evaluation relates to the dissemination or replication of the health program. Put simply, the process evaluation provides comprehensive operational information to the new sites so that the program can be replicated successfully.

► Efficacy, Effectiveness, and Efficiency

In evaluating programs, three key concerns are often whether the program was efficacious, effective, and efficient. These three terms are often used as synonyms, with minimal attention being paid to the important differences among the concepts. In fact, these differences have implications for the role of the process evaluation, and they reveal the underlying reasons for the different evaluation activities needed to quantify each.

Of the three terms, *efficacy* is probably the one most likely to be misused, and it is usually used instead of the term *effectiveness*. *Efficacy* is the maximum potential effect under ideal conditions. Because ideal conditions are difficult to create, the efficacy of an intervention is determined through rigorous studies—usually clinical trials, especially randomized clinical trials. By controlling many of the potential influences, randomized clinical trials provide a context in which the greatest possible effect of a treatment or intervention can be directly attributed to that treatment or intervention. Because of the costs and ethical considerations involved in clinical trials, efficacy evaluations of health programs are seldom done unless they are performed as part of evaluation research. In efficacy studies, the role of a process evaluation is to establish that the health program intervention is being carried out precisely according to the intervention protocol.

Effectiveness is the realistic potential for achieving the desired outcome when a treatment or intervention is used in real life. The degree of effectiveness of an intervention reflects what can be expected given the normally messy situations that occur when health programs are delivered to real-world clients. Data from outcome assessments and outcome evaluations, as well as from evaluation research, provide

practical experience with the anticipated level of intervention effectiveness. The degree of effectiveness may be reflected in several different statistics, depending on the evaluation design and methods. Any statistic that denotes the degree of difference related to having received the programmatic intervention—whether a difference score, a correlation coefficient, or an odds ratio—provides information on the degree of effectiveness. In this way, existing studies of the effectiveness of interventions provide a benchmark against which to gauge the success of subsequent or similar programs.

As a consequence, such studies have value beyond their specific findings. The role of a process evaluation in effectiveness studies is not only to ensure that the program is being provided according to protocol but also to document situations, events, and circumstances that influenced the delivery of the intervention.

Efficiency is, generically, the relationship between the amount of output and the amount of input, with higher outputs that are achieved with fewer inputs being deemed more efficient. With regard to health programs, efficiency is generally thought of in terms of the amount of effect from the program intervention—that is, the ultimate output—compared to the amount of inputs that went into providing the intervention. Data are collected as part of implementation documentation on the expenditures for all types of inputs into the program as well as on the outputs from the organizational and service utilization plans. Efficiency is then calculated as the cost per unit of output, where the unit of output is selected from the various outputs of the program. Using this basic notion of cost per output, efficiency could also be calculated per process objective.

► Data Collection Methods

Along with the question of what to measure is the question of how to collect data. At least

seven categories of methods of data collection are appropriate for process evaluations: activity logs, organizational records, client records, observations, questionnaires, interviews, and case studies (**TABLE 9-1**). This list of data collection methods does not preclude disciplined creativity in developing and using data collection methods or tools that are uniquely tailored to the program, including checklists (Simbar, Dibazari, Saeidi, & Majd, 2005). The use of activity logs, whether paper and pencil or web-based (Turner, Yorkston, Hart, Drew, & McClure, 2006), and organizational records tends to be more specific to process evaluation.

The choice of a data collection method needs to be congruent with the indicators in the process objectives and the best method for arriving at a conclusion about whether the process objective target was reached. Each method has its own advantages and disadvantages, all of which need to be considered carefully when choosing a data collection method.

For all data collection methods, high standards must be met in terms of the quality of the data collected, the reliability of the tools, and the accuracy of data entry. The need for reliability extends to assessing and establishing an acceptable degree of interrater reliability in the use of checklists and activity logs. Having a reliable means of assessing delivery of the intervention is especially important when the consistency of intervention delivery is problematic or when the program is a pilot that is being evaluated as an experiment.

► Quantifying Inputs to the Organizational Plan

During program development, specific organizational resources are identified as being keys to implementing the health program. Both inputs and outputs of the organizational plan are included in an implementation evaluation (**FIGURE 9-3**), which determines the extent to

TABLE 9-1 Methods of Collecting Process Evaluation Data

Method	When to Use	Examples of Measures	Pros	Cons
Activity log	Have a list of actions that are discrete, and a common understanding for what those are, need quantitative data	Number of sessions, number of participants, time received inquiry phone calls, date of press release	Can tailor activity log to the program activities, easy to use, easy to analyze data, applicable across pyramid levels, can develop for use on handheld devices	May become too long, may not be completed on a regular basis, easy to falsify
Checklist	Have a list of actions or behaviors that can be observed	Set up room for session, gave supportive comments, distributed program materials	Simple to use, can be developed to include a time frame, data entry is straightforward, applicable across pyramid levels, can develop for use on handheld devices	Difficult to narrow list items, challenging to write items for consistent interpretation, reliability needs to be established
Organizational records	Have existing records that capture information needed and can legally access those records, need quantitative data	Length of time on waiting list, number of computers bought or upgraded, number of hours worked	Accessibility to the information, applicable across pyramid levels, can be electronic databases	Need a data abstraction form, records may not include what is needed, may require complex data linking and data analysis
Client records	Have existing records that capture information needed and can legally access those records, need quantitative data	Program attendance, client compliance with program elements	Accessibility to the information, likely to be electronic record	Need a data abstraction form if paper records are used, records may not include what is needed, may require complex data linking and data analysis

Method	When to Use	Examples of Measures	Pros	Cons
Observation	Need to have data on interpersonal interactions or sequences of events	Number of staff–participant interactions	Data may reveal unexpected results, naturalistic, can quantify observations, can be recorded on handheld devices	Time intensive, need observation checklist, complex data analysis
Questionnaire	Need to collect data quickly from reliable respondents and have a reliable and valid questionnaire, need quantitative data	Degree of satisfaction with program, degree of compliance with program interventions	Can collect pencil-and-paper or web-based versions for many programs, applicable across pyramid levels	Respondent must have good reading skills and motivation to complete the questionnaire, will gather useless data if not well written, can be expensive for population-level programs
Interview	Have time and need qualitative data or have respondents for whom questionnaire is not appropriate	Commitment of staff members to program and intervention	Able to get detailed descriptions during one-on-one interview, possibly new insights	Time intensive, need private place for the interview, need interview question and format, more complex data analysis
Case study	Need to understand the full set of interactions around the program and the context in which it is functioning	Degree to which managerial personnel make changes to ensure fidelity of intervention	Gives very thorough picture of program and provides new insights	Extremely complex because it uses multiple methods over period of time, time intensive, very complex data analysis

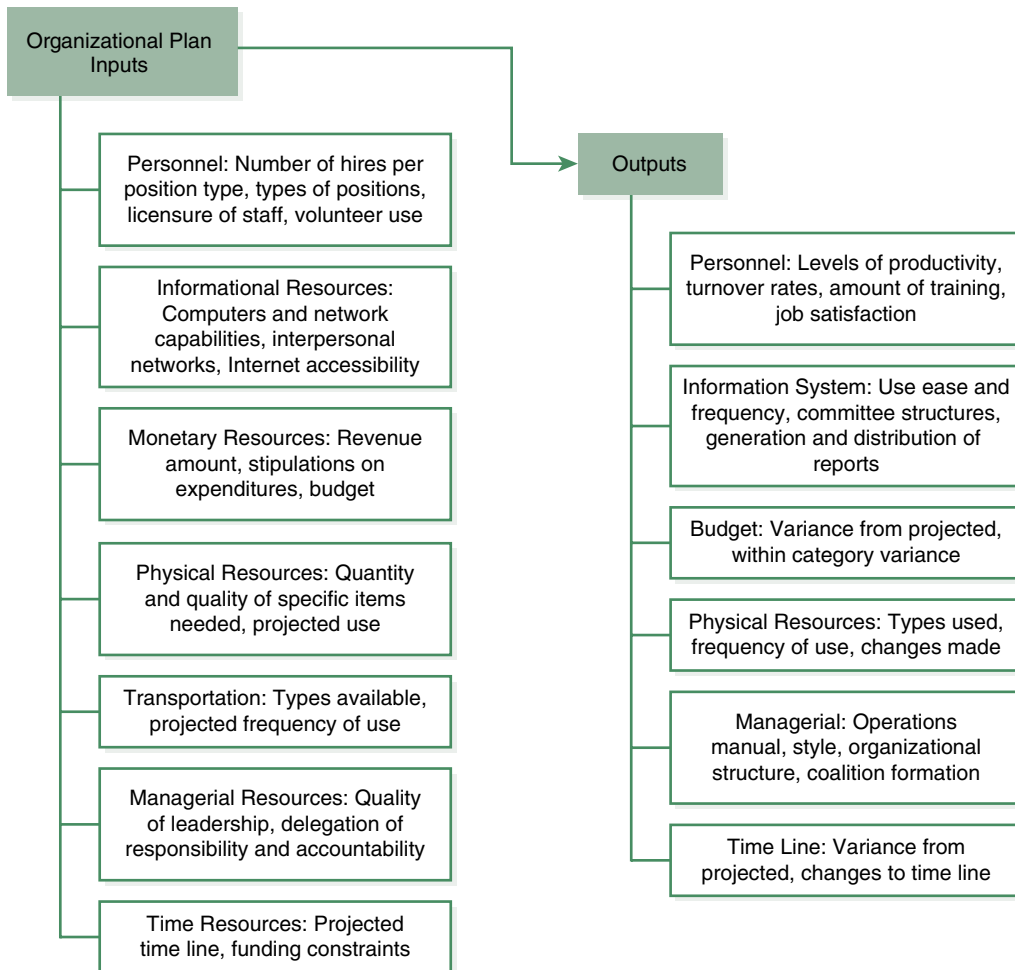


FIGURE 9-3 Examples of Organizational Plan Inputs and Outputs That Can Be Measured

which those inputs were available and used and the quantity of outputs. Although every input could be monitored, tracking all of them would be neither prudent nor feasible. Each program has a set of concerns and interests with regard to the organizational plan inputs, as reflected in organizational plan process objectives. Thus, the choice of which organizational plan inputs are evaluated is influenced by the earlier work of the program planners and the current concerns of program staff members.

A variety of measures can be used to measure each organizational plan input and output (TABLE 9-2), although the specific measures depend on the program and its objectives. At a minimum, the human resources and physical resources devoted to and utilized by the program must be monitored. These two organizational plan inputs are used as examples of considerations that would be involved in monitoring inputs. These discussions are not intended to be definitive but rather illustrative of how to approach the development of a process evaluation plan.

TABLE 9-2 Example of Measures of Inputs and Outputs of the Organizational Plan

Input	Measures of Input	Measures of Output
Human resources	Number of full-time equivalents (FTEs), number of new hires, number of volunteers, percentage of licensed personnel, percentage of personnel with certification, educational level of staff members, hours of training and orientation	Number of hours worked, staff–recipient ratio, hours of client contact per staff member, degree of job satisfaction of staff members and volunteers, degree of commitment of staff members to program and intervention
Informational resources	Number of computers bought or upgraded, number of program recruitment efforts, availability of communication hardware and software, ease of process data entry and retrieval, ease of impact data entry and retrieval	Degree to which computer and telecommunication systems facilitate delivery of the intervention; availability and accessibility of personnel, budget, operating, meeting, or other reports
Monetary resources	Amount of grant monies and donations, amount of indirect costs deducted from the program, number of proposals submitted for program funding	Dollar or percentage variance from budgeted per line item, number of grants awarded, profit or loss
Physical resources	Number and type of capital equipment, number and type of office or clinical equipment, square footage of office space	Extent to which changes are made to physical resources needed for intervention delivery, replacement of aged equipment
Transportation	Parking fees, total mileage per month, number of bus passes used, program vehicle expenses	Mileage per staff member, number of clients receiving transportation assistance, transportation cost per staff member or per program participant
Managerial resources	Place in organizational chart, years of experience, educational level, degree of ability to communicate clearly and persuasively	Extent to which managers are viewed by staff members as controlling or delegating, degree to which managerial personnel make changes to ensure fidelity of intervention
Time resources	Time line developed, presence of deadline dates	Number of days delayed, percentage of deadlines met, number of repeated requests

Human Resources

Both the quantity and the quality of the human resources used by the health program ought to be assessed, given that human resources often constitute the largest cost component of a program budget. The importance of tracking the staffing levels of a program is reflected in the results of a study carried out by Rosenheck and Seibyl (2005). Using 5 years of data on staffing levels, these authors were able to find an association between higher staffing levels and better outcomes for program participants. Similarly, nurse staffing levels have been associated with various indicators of inpatient quality, such as the number of falls and mortality rates (Minnesota Department of Health, 2015). This section does not present a definitive list of measures of human resource inputs but rather is meant to provide some guidance and ideas for monitoring related to human resources.

Not only is the quantity of personnel relevant, but those individuals' level of commitment to the program, their competencies in terms of knowledge and skills, and their attitudes are also critical considerations. The degree of commitment to the program is important to assess, given that program staff members who lose commitment are less likely to implement the program fully as designed. A change in commitment levels may reflect unanticipated challenges to the delivery of the program. For example, the program staff members may develop lower levels of commitment because of obvious failures of the intervention or because of undesirable side effects of the program. Thus, monitoring the program staff members' commitment to the program and to the interventions is one avenue for gaining insights into program implementation.

The extent to which program personnel are competent to deliver the program also provides information on the extent of program implementation. If licensed health professionals are required to deliver the intervention, but the program managers have insufficient fiscal resources to hire such individuals, then the program is not likely to be implemented as planned. Thus,

the extent to which the credentials of program staff members match the competencies needed for program implementation provides more evidence of the extent to which the designed interventions are delivered. This statement is not meant to imply that unlicensed personnel are not qualified. Indeed, many types of interventions and health programs rely heavily and appropriately on a wide range of qualifications for program staff members. Rather, the issue is the extent to which the program staff members' qualifications actually match what is required to implement the program fully and thus gain the maximum effect for program participants.

As the health program evolves over iterations of implementation, the qualifications needed by program staff members might change as well. These changing needs in relation to staff qualifications ought to be uncovered and noted through an implementation assessment or evaluation. The observed change may lead to a revision of the process theory for the program.

The participation of stakeholders continues into the process evaluation. The inclusion of volunteers and stakeholders as human resources in the organizational plan is intended to serve as a continual reminder that these individuals and groups need to be actively engaged in the various stages of the health program. Ideas from program staff members about ways to improve the program can be part of the process evaluation. Their ideas may help increase the effectiveness and efficiency of the program without compromising the integrity of the intervention as conceptualized. Their ideas may also provide insights into which aspects of the intervention were not implemented as intended. For example, staff members are very likely to pick up on and articulate the difference between what they were asked to do and what they are actually able to do with the given resources.

Another human resources input into many public health programs is the membership of a community consortium, coalition, or advisory board. The rationale for using community coalitions or consortia is based on the belief that the inclusion of such groups of individuals from the target population not only fosters the

development of culturally appropriate health programs but also influences the context of the program in ways that enhance the service utilization plan. Determining the extent to which this human resource was utilized may be mandated by various federal funding agencies. For example, as part of a Centers for Disease Control and Prevention (CDC) initiative to reduce racial and ethnic disparities across five health conditions, each grantee was required to report on the status of its community coalition. As part of a Maternal and Child Health Bureau program to reduce infant mortality, grantees had to have a community consortium and report on the functioning of that consortium.

Attendance records for community consortia or advisory board meetings are indicators of the amount of input of that type of human resource. The positions and organizations represented by the membership of the consortia or advisory board are other indicators of the human resources capability and diversity. Process monitoring data related to coalitions can reveal difficulties in mobilizing the community as a whole and elucidate the preferential interests of various groups. Process evaluation data help assess whether coalition activities have contributed any program effect. For example, members of a chronic disease prevention network reported needing to work with representatives of other sectors to address poverty-related issues effectively (Geneau, Legowski, & Stachenko, 2009). Such process information can be important for future coalition development.

Whereas commitment to and belief in the program and passion for the health problem are viewed as human resources inputs, job satisfaction is viewed as an output of the program. This distinction is important as a way to differentiate whether the correct inputs were obtained—namely, motivated individuals—from whether the human resources inputs were converted to satisfied staff members. The effort to screen and hire volunteers and staff members might also be important in this regard.

The preceding discussion hints that a wide variety of measures might be appropriate to

document, monitor, and evaluate human resources. The number of staff members per job category, the number of full-time equivalent (FTE) workers per job category, and the number of staff members with licensure and certifications are the absolute minimum data to be collected.

Physical Resources

Documenting the extent to which facilities are adequate in terms of the types of rooms needed to provide the health program or the accessibility of equipment can be particularly relevant for some health programs, especially if specific physical resources are essential for the success of the program. Lack of or failure to use the necessary physical resources may indicate that the program was not fully implemented. For example, if a health promotion program includes both an educational component requiring only a classroom and a cooking demonstration component requiring a kitchen, then the process monitoring needs to include documentation of the use of both facilities. If no kitchen facilities were used for the program, then only the educational component of the program was implemented. Knowing that facility limitations were an issue because only one component of the program was implemented provides an explanation for a weak program effect on the stated health outcomes.

The simplest measure of physical facilities would be a dichotomous variable (i.e., yes/no) as to whether the facilities or equipment specified in the process objective were used. However, for physical resources such as a classroom, it may also be useful for process monitoring purposes to collect data on whether the room was heated, well lit, and so forth. Similarly, if physical resources include items such as blood pressure cuffs or supplies for vaccination, then it may be informative to document or monitor their placement, adequacy, expiration dates, disposal safety, and such. These seemingly mundane aspects of the physical resources can influence whether program staff members utilize those resources as intended.

► Quantifying Outputs of the Organizational Plan

Just as only key organizational plan inputs are evaluated, so only key organizational outputs can realistically be evaluated. Those outputs associated with process objectives are the minimum to be included in the evaluation. For some programs, it may be important to understand the organizational structure, such as where in the organizational hierarchy the health program director is located. The program director's position in the organizational chart can indicate the relative importance of the health program and hence the ability of the program manager to garner resources for the program. Here, two organizational plan outputs—information systems and budgets—are used as examples of different approaches to the measurement of organizational plan outputs.

Information Systems

The ability to process data and generate needed reports is an output of the information systems. Given the central role played by information systems in acquiring program funds and documenting program implementation and effects, the information system must be able to generate the needed reports. The quality and timeliness with which reports can be generated can be considered an output of the information system. Tracking the number of problems, requests, and complaints would be another approach to measuring the information system outputs.

Part of the process evaluation may include an assessment of the system's capacity to handle data from the effect evaluation.

Measures of the information system output includes items such as staff members' perception of ease of use, frequency of using critical databases, amount of time clients are on hold, number of calls waiting to be answered, number and frequency of reports generated, and perceived usefulness of the reports generated. Information systems also

include the use of inter- and intraorganizational communication. Measures of interpersonal information systems include the number of active committees; the number of community meetings attended by staff members; the perceived accuracy and timeliness of information received; and the number and types of modes of intraorganizational communication used, such as e-mail, memos, electronic mailing lists, and meetings.

Monetary Resources

The monetary resources generated to support the health program are the most obvious output of the budget plan. Essentially, the output of budget-focused efforts is the total dollars generated. Thus, the number and size of grants from philanthropic, federal, or state agencies, as highly valued measures, are counted. In addition, some health programs may be funded through contracts, which are also counted as outputs of the budgeting efforts.

► Quantifying Inputs to the Services Utilization Plan

Process evaluation typically is equated with elements that are included in an evaluation of the services utilization plan—namely, data on the participants and program delivery (**FIGURE 9-4**). As mentioned earlier in regard to evaluating the organizational plan, the process objectives serve as a guide for deciding which possible inputs and outputs are crucial to evaluate. **TABLE 9-3** provides some examples of measures that are useful for service utilization plan monitoring.

Participants and Recipients

The most basic data about participants are a simple count of how many individuals inquired about the program or were served. However, other types of data about program participants are reasonable to collect.

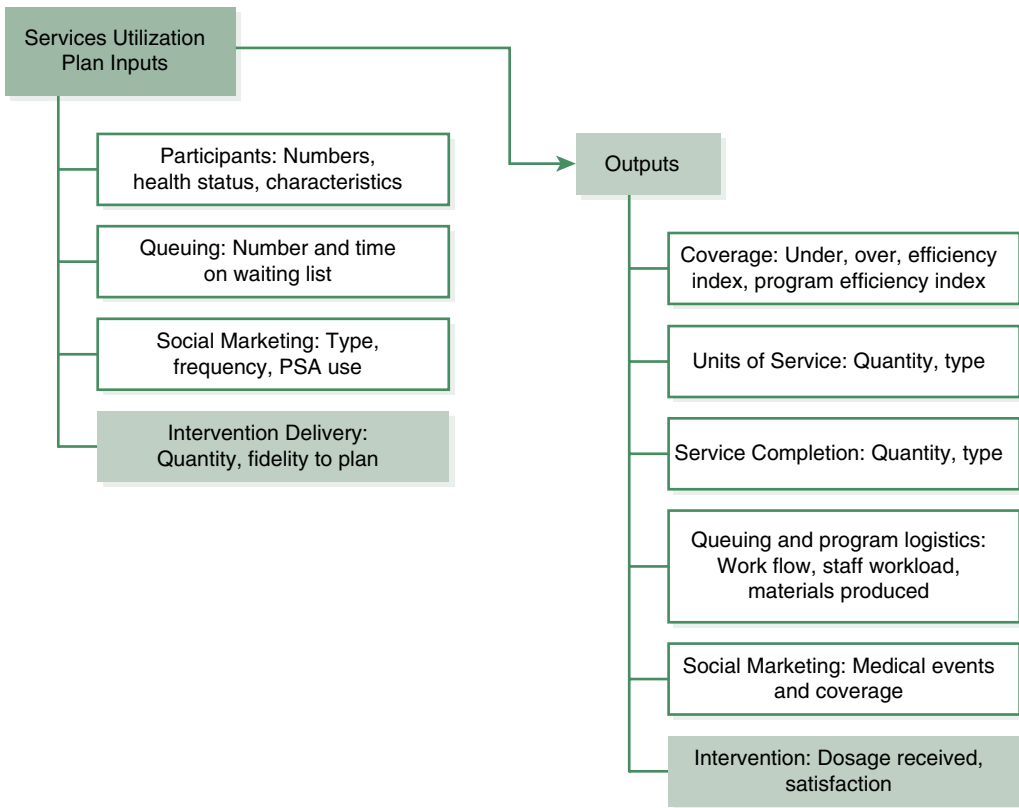


FIGURE 9-4 Examples of Services Utilization Inputs and Outputs That Can Be Measured

A simple, straightforward documentation of demographic characteristics—whether through interviews, self-report questionnaires, or clinic records—is one means to collect this type of data. Demographic data are necessary for determining the extent to which the target audience has been reached and whether the social marketing approaches were effective. These data can be readily collected during program sign-in or intake procedures, especially for direct and enabling services. Demographic data on recipients of population-focused programs may be more difficult to obtain directly given that population-based health programs do not require recipients to be physically present at a given location. For example, it would be very difficult to determine the number of recipients of health messages delivered via billboards.

Some basic information about the health status of participants can be collected and can prove important for later analyses of program effect. Measures of health status information can range from the simple, such as a single question rating of overall health, to a checklist of health conditions, to findings from a detailed medical history and physical examination. The complexity of the health measure corresponds to the level of the public health pyramid at which the program is aimed and the target audience characteristics.

Intervention Delivery and Fidelity

Intervention fidelity indicates whether the intervention was provided as designed and

TABLE 9-3 Examples of Measures of Inputs and Outputs of the Services Utilization Plan

Element	Measures of Input	Measures of Output
Program reach	Number of requests for program	Percentage undercoverage, percentage overcoverage
Participants	Number of recipients or participants, number of persons denied or unqualified for the program	Efficiency index, program efficiency index, degree of satisfaction with the program
Queuing and program logistics	Number on waiting list, presence of system to move individuals from waiting list to program or alternative programs	Length of time on waiting list, evenness of work among staff members and across time (workflow), number and types of materials produced
Mazlan program based on social marketing efforts	Type of social marketing, quality of marketing, extent of social marketing analysis	Number of advertising events, number of requests for program based on social marketing efforts
Intervention	Number of meetings to standardize program, extent of revisions based on previous cycle of intervention delivery, extent of revisions based on new research evidence	Fidelity to intervention plan, number of sessions, hours of program delivery, number of participants completing intervention (service completion), number of requests for additional program delivery, use of materials produced

planned; that is, it comprises the alignment of the intervention activities with the elements of the intervention theory. The concept of intervention fidelity applies to single, one-time programs as well as to standardized interventions adopted by numerous organizations. Unless high intervention fidelity is achieved, the program may fail to reach the desired outcomes. Achieving high implementation fidelity can be difficult for community-based organizations to achieve (Mihalic, Fagan, & Argamaso, 2008). Historically, relatively few evaluators of community-based health programs reported how they monitor implementation fidelity. Attention has been devoted more recently to the measurement of implementation fidelity in community-based prevention programs, including the use of

self-reports, video observations, and audio recordings (Breitenstein et al., 2010). Failures in implementation fidelity can discredit potentially effective interventions (Dobson & Singer, 2005). This reinforces the notion that adequate and ongoing process monitoring can help prevent the program failure due to inadequate program implementation by detecting the source of the failure to have full intervention fidelity.

Intervention fidelity may be compromised in three ways. First, an intervention can fail because of the lack of a program (nonprogram, no treatment), meaning that the program was not provided. If the intervention requires a sophisticated delivery system, it may not be delivered. For example, a program that involves coordinating services among a variety of health professionals

employed by different agencies can easily result in no program. An alternative way for no treatment to occur is that the delivery of the intervention in some way negates the intervention. Having program personnel who are not supportive, encouraging, or empathetic would negate intended program interventions that require such qualities in staff members. Similarly, physical resources—for example, a building with limited and difficult access—might negate interventions for persons with disabilities.

The second way that an intervention can fail is if an intervention other than the one designed and planned is provided, or if the dosage is drastically reduced. If an educational program is designed to cover five topics but only one topic is covered, despite holding the classes for the designated number of hours, the program will be dramatically diluted, possibly to the point of having no effect on participants. The process objectives that specify the program content and quantity can be used to assess whether the strength (dosage) of the intervention was delivered as intended. It is also possible that the intervention theory was flawed; if so, the inadequacy of the intervention as conceptualized will lead to intervention failure, although the actual program may be delivered as planned.

The third way that an intervention can fail is if the intervention is provided in an inconsistent manner, resulting in a nonstandardized treatment over time or among recipients. A standardized intervention is necessary to ensure that the intervention is responsible for the outcomes. If program personnel use their own discretion to alter the intervention, there is no assurance that the intervention as planned is responsible for the health effects found in an outcome evaluation. One approach to minimize this potential source of program failure is to incorporate the program theory in the training of program personnel, thereby ensuring that they appreciate the need to follow the guidelines for delivery of the program. Of course, the organizational plan outputs of policies, procedures, and standardized materials need to be in place and used as tools to help standardize the program.

Inconsistency in intervention delivery can also have a more insidious and hidden cause. Both program participants and program staff members have their own working explanations or theories of how a program affects participants. The espoused theory is the stated or espoused explanation for how things happen (Argyris, 1992). In a properly functioning program, the effect theory becomes known and understood by program staff members; therefore, it also should be the espoused theory. Of course, staff members inevitably know what they are supposed to say about the program interventions, regardless of whether they believe it or whether their actions match their espoused theory.

The actions that staff members take to achieve the ends constitute their theory-in-use—that is, the working theory that staff members actually use. One key aspect of process monitoring is to observe the theory-in-use because it provides insights into which intervention was provided to recipients, and then compare it to the intervention theory developed during the planning stage. It is especially important to observe the theory-in-use during the process monitoring and to compare it to the program intervention theory.

Naturally, some programs will be more prone to inconsistency than other programs. For example, the espoused theory and the theory-in-use among staff members in immunization clinics are likely to be quite congruent. In contrast, staff members working in gun violence prevention programs are more likely to demonstrate incongruence between the two theories because such a program entails complex interpersonal interactions and addresses socially sensitive problems.

Without standardized interventions, it is difficult to connect the health effects of the program to the program interventions. For this reason, the process evaluation of the intervention needs to take into account the possible sources of inconsistency in how the program intervention is delivered. Findings from the program evaluation may be confusing or misleading if the espoused theory and the theory-in-use are incongruent. Implementation of the program may be inconsistent, with some staff members providing the

program according to the espoused theory and other personnel delivering the program based on their theory-in-use.

The inconsistencies among the three theories (intervention theory, espoused theory, and theory-in-use) can be a source of *decision drift*. Although a decision may be made as a milestone in the planning process, over time the decision can evolve in unexpected ways among planners and program staff members. Decision drift is a natural process in long-term health programs, but it can also occur within a short span of time—even within a single meeting. Decision drift is detrimental only if it results in a program that lacks coherence or that no longer addresses the health problem. Through process evaluation, the extent to which decision drift has occurred can be assessed, and the specific areas in which it occurred can be pinpointed. Recognizing that decision drift has occurred provides a basis for either revising the decisions or modifying the program elements and objectives to bring them more in line with the current decisions. Which action to take depends on whether the decision drift has resulted in a more or less efficient and effective health program.

► Quantifying Outputs of the Services Utilization Plan

During a process evaluation of each of the services utilization plan outputs, one of the first decisions to be made is which time frame will be used. Because many health programs are vulnerable to seasonal fluctuations, the time frame used is often highly significant. For some programs, annual measures are more reasonable; for shorter programs, the end of the program is sufficient. Ideally, the evaluation of the services utilization plan outputs should occur as close to real time as possible rather than being retrospective to ensure that programmatic changes are made in a timely fashion.

The service utilization plan outputs discussed here are coverage, units of service, service completion, workflow, and materials produced. These outputs are assessed with regard to the extent that corresponding process objectives have been achieved.

Coverage as Program Reach

The extent to which the program reached its intended audience is called *coverage*, but it may also be referred to as reach. Coverage, in the context of state or federal health policy and subsequent health programs, has become a topic of political debate, which increases the importance of understanding coverage measures and interpretations. Studies of programs have identified attention to coverage as important to understanding why the programs did not have the expected effects (Gottfredson et al., 2006) and for determining which program effects were associated with coverage (Macinko, de Souza, Guanais, & Simoes, 2007). Means used to track program coverage have included documentation, monitoring, and evaluation. If the coverage rates are less than expected or predicted, then it is important to identify the barriers to accessing the program, which may entail a qualitative approach (Sobo, Seid, & Gelhard, 2006). There are several ways to quantify and interpret coverage.

Measures of Coverage

Monitoring the degree of participation in a health program is a basic aspect of process evaluation. All funding agencies and program managers want assurances that the program had participants. The mechanism for tracking the number of individuals served by the program must be in place before the process evaluation begins. Measures of coverage require having accurate data on the number of program participants within a given time period. Collecting data on a frequent or periodic basis allows for ongoing monitoring and still makes possible the aggregation of the numbers to get totals for a given time period. For example, if

immunization clinics are offered three times per month, managers may count the number of participants per clinic and then add the totals for all three clinics to obtain a total number of persons served per month.

Coverage is assessed with regard to undercoverage and overcoverage. Data from the needs assessment are required regarding the numbers of individuals in need and not in need of the program. These data, along with the actual number of individuals served by the program, form a matrix of undercoverage, ideal coverage, and overcoverage (**TABLE 9-4**). Calculating the different measure of coverage is simple, yet it can yield a great deal of information about which program component needs attention due to undercoverage. **EXHIBIT 9-1** lists the formulas

for coverage measures. A narrative example of coverage measures is shown in **EXHIBIT 9-2**, with the corresponding data in an Excel spreadsheet shown in **EXHIBIT 9-3**.

Undercoverage is measured as the number of individuals who are in need of the service and actually received the service divided by the number of individuals who are in need. It occurs when the program is not delivered to a large portion of the target audience. On the other side of the coin, *overcoverage* occurs when the program is being used by individuals not in the target audience. It is calculated as the number of individuals who are not in need of the service but who receive the service divided by the number of those who receive the service. Overcoverage can put a strain on existing human and physical

TABLE 9-4 Matrix of Undercoverage, Ideal Coverage, and Overcoverage

	Persons Not Served by the Program	Persons Served by the Program
Persons Not in Need of Program	Ideal coverage	<i>Overcoverage</i>
Persons in Need of Program	<i>Undercoverage</i>	Ideal coverage

EXHIBIT 9-1 Formulas for Measures of Coverage

$$\text{Percentage of overcoverage} = \frac{\text{number not in need but served}}{\text{number served}}$$

$$\text{Percentage of undercoverage} = \frac{\text{number in need and served}}{\text{number in need}}$$

$$\text{Coverage efficiency} = \frac{(\text{number served} - \text{number served but not in need})}{\text{number in intended population}}$$

resources (Osuch et al., 2015). Assessments of under- and overcoverage can be used to determine how to improve a program (Countdown Coverage Writing Group, 2008) and ultimately to justify its continuance.

A more telling indicator is coverage efficiency. *Coverage efficiency* is calculated as the number served minus the number served but not in need divided by the number in target population. If there is no under- or overcoverage, then the coverage efficiency is 100%. As shown in Exhibit 9-3, it is difficult to reach that level of coverage efficiency.

The coverage efficiency index can be applied to the program to create the program efficiency index. The *program coverage efficiency index* is calculated by summing the efficiency indices for each program component and then dividing this sum by the number of program components.

Although these measures of coverage provide information about the extent to which the program is reaching the intended audience, they do not provide information on the extent to which the program is meeting its objectives. To make this determination, another calculation is needed. That is, the percentage of coverage efficiency achieved can be compared to the percentage coverage efficiency specified as the objective target value (see Exhibit 9-3). This calculation uses the numbers from the under- and overcoverage matrix and the target value set in the objective. It represents another way for program managers and staff members to determine whether the program is reaching its objectives.

High coverage results indicate that the program has achieved good marketing and program recruitment. Taken as a set, coverage measures indicate the extent to which efforts to enroll individuals in the program are effective and the target audience is being reached. They also indicate areas warranting managerial action and further tailoring of the program.

When thinking about the sources of data needed to measure participation, one must

think in terms of intended populations and recipients. The number of individuals in need of the program—specifically, the size of the intended population—should have been determined through the community assessment. If the community needs assessment data do not include some estimation of the size of the intended audience, it becomes almost impossible to determine whether the program is dealing with undercoverage or overcoverage. At best, the level of coverage can be estimated only from alternative data sources.

Units of Service

For some programs, it will be important to distinguish between the number of program participants and the number of contacts. Outreach programs may be quite successful in making a high number of contacts with potential program participants, but the number of individuals who actually show up for the program may be a small fraction of that number. In this case, a decision must be made about whether the number of individuals contacted is equivalent to the number of outreach recipients. This is a gray area, and each program will have a slightly different answer to this question. The key is to make clear the item being counted and the definition on which it is based. In particular, the process objective should specify what is being counted.

The unit of service (UOS) must be clearly defined and articulated before the count can begin. A UOS is a predetermined unit, such as number of contact hours, number of individual clients seen, number of educational sessions offered, or size of caseload. This measure is used primarily for programs at the direct or enabling services levels of the public health pyramid. For some programs, the UOS is specified by the funding agency, in which case, data for that UOS become a minimum of what is included in the process evaluation of services delivered.

Participant-Related Issues

Dosage Measurement

Of the five dosage elements (frequency, duration, strength, route of administration, and administration credibility), frequency and duration have the greatest relevance for ongoing monitoring. Frequency of the intervention—whether hourly, daily, weekly, or monthly—and duration of the intervention—whether one session, 8 weeks of classes, or 6 months of exposure—are the elements of dosage that may vary from the objectives laid out in the plan because of either program or participant factors. Thus, to ensure that the planned dosage was received by the participants, it is important to determine the degree to which program participants completed the health program—in other words, service completion. The inverse of service completion is the dropout rate of the program.

Singular event programs, such as screening clinics, are likely to have different completion rates than programs with longer-term involvement with the participants, such as substance abuse counseling, Meals on Wheels, or exercise classes. For some enabling services, service completion is the achievement of the service plan or care plan. Counting the number of participants who have completed their service plan provides different information on the extent to which the intervention was implemented and the dosage received by the average participant. Service plan completion can be estimated only if staff members keep good records of enrollment and record participant attendance on an individual basis.

Drastic changes in completion rates may signal problems with program staff members or with the design of the program. The process objectives should include a threshold for what is an acceptable program completion rate.

Level of participation is a corollary of service completion. If the intervention or impact theories are predicated on a certain level of participation, then data on the level of participation need to be collected as a means of determining whether participants received the appropriate “dose” of the intervention. For some health programs, level

of participation might be a simple yes or no on attendance at, say, an immunization clinic, or it might comprise a weighted calculation of the percentage of time during a set of sessions in which the participant was actively engaged with the group. The measure chosen depends on the nature of the program and the target value and indicator specified in the process objective. A low level of participation can result if program personnel are not skilled in engaging participants (a managerial issue) or if the intervention is not appealing to the participants (a process theory issue).

The dosage received by a program participant can be calculated if accurate and complete data have been collected on each participant and the quantity of intervention received. Continuing with the congenital anomalies prevention program in Layetteville, **EXHIBIT 9-4** shows the dosage for three women. Suppose that the program has four components—screening, prenatal counseling, use of prenatal vitamins, and a cooking class—and each component has a different planned amount specified in the intervention theory and the corresponding process objectives. If the amount of each component received by each woman is recorded, then the program dosage can be calculated for each woman. Using the program average dosage for each component, a weighted average dosage can then be calculated. The advantage of calculating the weighted average dosage is that it takes into account any variation in planned dosage for components.

The main challenge in calculating the dosage is the collection of data in a manner that enables its calculation. If they know the program average per program component, the program manager and planners can make more informed decisions about program modifications and changes, such as which components need revision or additional support.

Satisfaction Measurement

Participant satisfaction with the program is an element of process evaluation and not a health outcome of the program. This perspective stands in contrast to the general vernacular of classifying

EXHIBIT 9-4 Examples of Calculating Dosage for the Congenital Anomalies Prevention Program Using Excel

	A	B	C	D	E	F
1		Program Component	Unit	Planned Amount	Amount Received	Percentage Received
2						
3	Participant A	Screening	Session	2	2	100%
4		Prenatal counseling	Session	4	4	100%
5		Prenatal vitamins	Weeks	52	43	83%
6		Cooking class	Hours	8	7	88%
7						
8	Participant B	Screening	Session	2	2	100%
9		Prenatal counseling	Session	4	3	75%
10		Prenatal vitamins	Weeks	52	32	62%
11		Cooking class	Hours	8	4	50%
12						
13	Participant C	Screening	Session	2	1	50%
14		Prenatal counseling	Session	4	2	50%
15		Prenatal vitamins	Weeks	52	20	38%
16		Cooking class	Hours	8	1	13%
17						
18					Average Amount	Avg Percent
19	Program Average	Screening	Session	2	1.7	83%
20		Prenatal counseling	Session	4	3.0	75%
21		Prenatal vitamins	Weeks	52	31.7	61%
22		Cooking class	Hours	8	4.0	50%
23		Weighted average dosage				84.2%
24						
25	Formulas for Calculating Program Average, Based on 3 Participants	Screening	Session	2	$(e3+e8+e13)/3$	$(f3+f8+f13)/3$
26		Prenatal counseling	Session	4	$(e4+e9+e14)/3$	$(f4+f9+f14)/3$
27		Prenatal vitamins	Weeks	52	$(e5+e10+e15)/3$	$(f5+f10+f15)/3$
28		Cooking class	Hours	8	$(e6+e11+e16)/3$	$(f6+f11+f16)/3$
29		Weighted average dosage (assuming all participants have the same planned amount)				$((F19*D19) + (F20*D20) + (F21*D21) + (F22*D22)) / D19 + D20 + D21 + D22) / 100$
30						

satisfaction as an outcome. The rationale for thinking of satisfaction as a process output stems from the definition of *satisfaction*: the degree to which participants receive what they expect to receive and the extent to which their expectations are met with regard to how they are treated (Parasuraman, Zeithaml, & Berry, 1985). Similarly, the Expectancy Confirmation Model describes patient satisfaction as the difference between patient experience and patient expectation (Lo, 2014; Oliver, 2014). Thus, program satisfaction can be affected by the degree which the participants experience beneficial effects (VanWormer,

Martinez, Cosentino, & Pronk, 2010). What is received, what is expected, and how one is treated are all elements of the service utilization plan rather than parts of the effect theory of how the intervention leads to health changes.

Client satisfaction may not predict program outcomes for individuals. For example, the associations between satisfaction with providers and use of preventive services, such as flu vaccinations and colorectal cancer screening, was actually explained by the patient's characteristics (Jerant, 2014). Nonetheless, satisfaction with health services provides some insight as a

process indicator of the experience of program participants.

Most funding agencies are interested in knowing the satisfaction level of participants in the programs they are funding. Most participants, clients, and patients report being somewhere in the range “satisfied to very satisfied” in regard to the services they receive, perhaps because it is difficult to measure satisfaction in a way that does not lead to a ceiling effect. A *ceiling effect* occurs when the measurement tool is constructed so that respondents do not have an opportunity to distinguish among levels at the high end of the scale. It becomes apparent when the item has a high mean value and very low standard deviation. In other words, the distribution becomes highly skewed to one side. For example, if the satisfaction scale is a 5-point Likert-type scale, the difference between 4 (“somewhat satisfied”) and 5 (“very satisfied”) will lead to 5 being chosen more often than 4. One remedy for avoiding the ceiling effect is to use a scale of 1 to 10, which then allows respondents to distinguish among 7, 8, 9, and 10 as levels of satisfaction.

Other difficulties exist in developing a measure of satisfaction. One challenge is to create a measure that is culturally sensitive and appropriate to subgroups of participants. Fongwa, Hays, Gutierrez, and Stewart (2006) specifically created a measure for African Americans but caution that the measure needs testing with other ethnic groups.

Another challenge is that self-report responses on a questionnaire may not be the same as responses in an interview. Marcinowicz, Chlabicz, and Grebowski (2007) found a discrepancy in this regard, with more negative responses being offered by patients when interviewed compared to their questionnaire responses.

Yet another difficulty stems from the conceptual definition of satisfaction, which entails a match between expectations and experience. Developing a measure that captures both expectations and actual experience will be longer, more complex to complete, and more challenging to analyze. For these reasons, the use of an existing satisfaction measure is highly recommended.

The last difficulty in measuring satisfaction concerns the scope of what is important in terms of satisfaction. For example, inpatient satisfaction questionnaires generally include items about parking and food service alongside items about the courtesy of staff members (Gesell, 2001; Mostyn, Race, Seibert, & Johnson, 2000). If such items are included, their relevance to the organizational and service utilization plans should be explicit.

Manary, Boulding, Staelin, & Glickman (2013) suggest taking the following into consideration when developing measures of satisfaction: whether and which specific visit/program participation measures correlate with which outcome measures, the nature of the interactions between recipient and provider, promptness of data collection after receiving the service or program, statistically adjusting for potential moderators and mediators, and the difficulty in a defining “patient satisfaction” (Manary et al., 2013).

Program Logistics Workflow

Interaction inevitably occurs between the procedures used for managing waiting participants and the amount of work done by program staff members. Measures of workflow are one indicator of the amount of work done by the program staff members and the queuing of participants. Examples of workflow measures include minutes that participants wait to be seen, number of days between signing up for the program and beginning the program, number of days between being referred to the program and being accepted into the health program, and amount of time required for program staff members to complete a specific task. Of course, the amount of work done by program staff members is influenced by the volume of program participants and the rate at which they participate in the program. For direct services health programs, the volume and queuing greatly affects the workflow of program staff. For population-based programs, the level of

cooperation from others, such as media representatives, may affect the workflow of program staff, such as in delivery of a mass media campaign.

Data related to both volume of participants and workflow come from a variety of sources. Observations of program staff members, participant records, appointment logs, class sign-in sheets, and billing statements are common sources of these data. For some programs, planners may decide to develop specific data collection forms. If this step is taken, the program staff members should be involved in the process because using an empowerment approach to planning increases the likelihood that optimal measures will be developed and used by the staff members.

Materials Produced

Both the quantity and the quality of the materials produced for the health program need to

be considered in the process evaluation. Having data about the materials provides insights into the work done by staff members and the extent to which the intervention was delivered in the manner planned. One managerial insight that has emerged from studies tracking the production of materials is that resources are sometimes directed more toward materials than toward implementation of the intervention. Data about production of materials can be difficult to obtain and will be very program specific.

► Across the Pyramid

Across the public health pyramid, process monitoring and evaluation focus on the inputs and outputs of the organizational and service utilization plans—albeit tailored to the specific program, of course (TABLE 9-5). For programs at

TABLE 9-5 Examples of Process Evaluation Measures Across the Public Health Pyramid

	Direct Services	Enabling Services	Population Services	Infrastructure
Organizational Plan Input	Provider credentials, location	Provider credentials, physical resources (e.g., cars)	Provider credentials, managerial resources	Personnel qualifications, managerial resources, fiscal resources
Organizational Plan Output	Protocols and procedures for service delivery, data about individual participants	Protocols and procedures for service delivery, data about participants	Protocols and procedures for service delivery	Budget variance, fiscal accountability, data, management information systems
Service Utilization Plan Input	Wait times, characteristics of participants	Wait times, characteristics of participants	Characteristics of the population	Characteristics of the workforce
Service Utilization Plan Output	Measures of coverage	Measures of coverage	Measures of coverage	Materials produced, number of participants

the direct services level, data measure units of service, such as number of individuals served and number of contact hours with individuals. Such measures are consistent with the nature of health programs designed for the direct services level of the pyramid. Many measures of patient satisfaction relate to the direct services. For instance, the Visit-Specific Questionnaire assesses perceptions of the convenience and technical competence of a specific encounter (Ware & Hays, 1988), whereas the Client Satisfaction Questionnaire measures overall satisfaction with a specific healthcare provider (Atkinson & Zibin, 1996). The Consumer Assessment of Healthcare Providers and Systems® includes sets of items relating to perceived access to care and communication among healthcare providers.

At the enabling services level, process monitoring and evaluation indicators are likely to be similar to those used at the direct services level, but they are modified to reflect the specific program and the use of different sources of data.

At the population-based services level of the public health pyramid, as at the direct services and enabling services levels, program process evaluations and monitoring efforts ought to address inputs and outputs of the organizational and service utilization plans. This is true if the program is implemented at the community level. For example, Glick, Prelip, Myerson, and Eilers (2008) focused on documenting the extent to which a fetal alcohol syndrome prevention campaign reached the intended audience. The units of service measured at this level of the pyramid could include number of individuals served, number of agencies involved, or number of households reached.

At the infrastructure level, process monitoring and process evaluation focus on the program infrastructure. If the health program is designed for one of the other levels of the pyramid, then the infrastructure becomes the source of the inputs and outputs of the organizational and service utilization plans. Of course, a program may be designed and intended to change the infrastructure itself. For example, Pearson and colleagues (2005) conducted a process evaluation of the implementation of the chronic care model across

42 organizations, focusing on the changes made at the sites to implement the program. Infrastructure changes may not always lead to the intended outcomes. An evaluation of the Substance Abuse and Mental Health Services Administration's incentive grant program to nine states for infrastructure changes concluded that infrastructure changes did not result in better service changes and consumer outcomes (Leff, Chchocki, Chow, & Lupton, 2014). Infrastructure units of service could comprise the number of employees involved, and outputs might consist of the number of policy or procedure updates and job satisfaction, especially if the employees are considered inputs into the program.

► Discussion Questions and Activities

1. Involvement of community coalitions and consortia in the implementation of health programs has become widespread. What would be possible and appropriate measures or indicators of having implemented community coalitions or consortia as part of the program delivery?
2. What would you suggest as methods and techniques to avoid the failure of interventions? Justify your ideas in terms of the various ways that interventions can fail.
3. To obtain accurate measures of coverage, which information systems and data collection methods need to be in place? Which steps can ensure that these elements are put in place in a timely manner?
4. Process monitoring and process evaluation data are useful only when they are interpreted correctly and subsequently used to make program changes. Outline a plan, with actions and stakeholders, for increasing the likelihood that the process data will contribute to accurate and responsible program delivery.
5. Using the information in Exhibit 9-2, calculate the measures of coverage for

the Congenital Anomalies Prevention Program, specifically the percentage overcoverage, percentage undercoverage, and coverage efficiency. Of the various programs for which coverage data are available, which programs seem to be most efficient?

- Using the information in Exhibit 9-2, calculate the dosage for the Adult Immunization Program for 150 people who needed and received the full vaccination.

► Internet Resources

Workbook for Designing a Program Evaluation

The CDC has developed a workbook for program managers for a range of evaluation topics (<http://www.cdc.gov/obesity/downloads/cdc-evaluation-workbook-508.pdf>) and on writing process evaluation questions (<https://www.cdc.gov/healthyyouth/evaluation/pdf/brief4.pdf>)

United Nations Children's Fund (UNICEF)

UNICEF has created a one-page tool that includes key questions to ask as part of a process evaluation: www.unicef.org/lifeskills/index_10489.html#Process%20indicators%20for%20the%20programme.

Children, Youth, and Families at Risk (CYFAR)

CYFAR, through a cooperative agreement with the University of Minnesota, offers instructional videos on program evaluations as well as process and outcome evaluations at <https://cyfar.org/resource-core-competencies/program-evaluation>.

World Health Organization (WHO) Process Evaluation Handbook

This handbook presents a lot of content in a simple and straightforward manner, with lots of examples and links to tools: http://whqlibdoc.who.int/hq/2000/WHO_MSD_MSB_00.2e.pdf.

References

- Argyris, C. (1992). *On organizational learning*. Cambridge, MA: Blackwell Publishers.
- Atkinson, M. J., & Zibin, S. (1996). Evaluative review of quality of life instruments, Chapter 5 in *Quality of life measurement among persons with chronic mental illness: A critique of measures and methods*. Retrieved from http://www.phac-aspc.gc.ca/mh-sm/mentalhealth/pubs/quality_of_life/
- Breitenstein, S. M., Gross, D., Garvey, C., Hill, C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions. *Research in Nursing & Health*, 33(2), 164–173. doi:10.1002/nur.20373.
- Chalmers, M. L., Housemann, R. A., Wiggs, I., Newcomb-Hagood, L., Malone, B., & Brownson, R. C. (2003). Process evaluation of a monitoring log system for community coalition activities: five-year results and lessons learned. *American Journal of Health Promotion*, 17(3), 190–196.
- Countdown Coverage Writing Group. (2008; on behalf of the Countdown to Core). Countdown to 2015 for maternal, newborn, and child survival: The 2008 report on tracking coverage of interventions. *The Lancet*, 371(9620), 1247–1258. doi:[http://dx.doi.org/10.1016/S0140-6736\(08\)60559-0](http://dx.doi.org/10.1016/S0140-6736(08)60559-0)
- Dobson, K. S., & Singer, A. R. (2005). Definitional and practical issues in the assessment of treatment integrity. *Clinical Psychology: Science and Practice*, 12(4), 384–387. doi:10.1093/clipsy.bpi046
- Fongwa, M. N., Hays, R. D., Gutierrez, P. R., & Stewart, A. L. (2006). Psychometric characteristics of a patient satisfaction instrument tailored to the concerns of African Americans. *Ethnicity & Disease*, 16, 948–955.
- Geneau, R., Legowski, B., & Stachenko, S. (2009). An intersectoral network for chronic disease prevention: The case of the Alberta Healthy Living Network. *Chronic Diseases in Canada*, 29(4), 153–161.
- Gesell, S. B. (2001). A measure of satisfaction for the assisted-living industry. *Journal of Healthcare Quality: Promoting Excellence in Healthcare*, 23, 16–25.
- Glick, D., Prelip, M., Myerson, A., & Eilers, K. (2008). Fetal alcohol syndrome prevention using community-based narrowcasting campaigns. *Health Promotion Practice*, 9, 93–103. doi:10.1177/1524839907309044
- Gottfredson, D., Kumpfer, K., Polizzi-Fox, D., Wilson, D., Puryear, V., Beatty, P., Vilmenay M. (2006). The Strengthening Washington D.C. Families Project: A randomized effectiveness trial of family-based prevention. *Prevention Science*, 7(1), 57–74. doi:10.1007/s11121-005-0017-y
- Jerant, A. (2014). Satisfaction with health care providers and preventive care adherence: A national study. *Medical Care*, 52(1), 78.
- Leff, S., Cichocki, B., Chow, C. M., & Lupton, C. (2014). Infrastructure change is not enough: An evaluation of SAMHSA's mental health transformation state incentive grants. *Psychiatric Services*, 65(7), 947–950. <http://dx.doi.org/10.1176/appi.ps.201300324>

- Lo, J. Y. (2014). Measuring patient satisfaction in cancer care. Doctoral dissertation, The University of Wisconsin-Madison. Madison, WI.
- Macinko, J., de Souza, M., Guanais, F. C., & Simoes, C. (2007). Going to scale with community-based primary care: An analysis of the family health program and infant mortality in Brazil, 1999–2004. *Social Science and Medicine*, *65*, 2070–2080. doi:10.1016/j.socscimed.2007.06.028
- Manary, M. P., Boulding, W., Staelin, R., & Glickman, S. W. (2013). The patient experience and health outcomes. *New England Journal of Medicine*, *368*(3), 201–203.
- Marcinowicz, L., Chlabicz, S., & Grebowski, R. (2007). Open-ended questions in surveys of patients' satisfaction with family doctors. *Journal of Health Services Research and Policy*, *12*, 86–89. doi:10.1258/135581907780279639
- Mihalic, S. F., Fagan, A. A., & Argamaso, S. (2008). Implementing the Life Skills Training drug prevention program: Factors related to implementation fidelity. *Implementation Science*, *3*, 5. doi:10.1186/1748-5908-3-5
- Minnesota Department of Health. (2015). *Hospital Nurse Staffing and Patient Outcomes: A Report to the Minnesota Legislature*. St. Paul, Minnesota, MN: Division of Health Policy/Health Economics Program.
- Mostyn, M. M., Race, K. E., Seibert, J. H., & Johnson, M. (2000). Quality assessment in nursing home facilities: Measuring customer satisfaction. *American Journal of Medical Quality*, *15*, 54–61. doi:10.1177/106286060001500203
- Oliver, R. L. (2014). *Satisfaction: A behavioral perspective on the consumer* (2nd ed.). New York, New York: Routledge.
- Osuch, E. A., Vingilis, E., Summerhurst, C., Forster, C. I., Ross, E. E., & Wrath, A. J. (2015). Process evaluation of an early-intervention program for mood and anxiety disorders among older adolescents and young adults. *Psychiatric Services*, *66*, 1113–1117. doi:10.1176/appi.ps.201400256
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, *49*, 41–50. doi:10.2307/1251430
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Pearson, M. L., Wu, S., Schaefer, J., Bonomi, A. E., Shortell, S. M., & Mendel, P. (2005). Assessing the implementation of the chronic care model in quality improvement collaboratives. *Health Services Research*, *40*, 978–996. doi:10.1111/j.1475-6773.2005.00397.x
- Rosenheck, R. A., & Seibyl, C. L. (2005). A longitudinal perspective on monitoring outcomes of an innovative program. *Psychiatric Services*, *56*, 301–307. http://dx.doi.org/10.1176/appi.ps.56.3.301
- Simbar, M., Dibazari, Z. A., Saeidi, J. A., & Majd, H. A. (2005). Assessment of quality of care in postpartum wards of Shaheed Beheshti Medical Science University Hospitals, 2004. *International Journal of Health Care Quality Assurance*, *18*, 333–342. http://dx.doi.org/10.1108/09526860510612180
- Sobo, E. J., Seid, M., & Gelhard, L. R. (2006). Parent-identified barriers to pediatric health care: A process-oriented model. *Health Services Research*, *4*, 148–171. doi:10.1111/j.1475-6773.2005.00455.x
- Turner, C., Yorkston, E., Hart, K., Drew, L., & McClure, R. (2006). Simplifying data collection for process evaluation of community coalition activities: An electronic web-based application. *Health Promotion Journal of Australia*, *17*(1), 48–53. http://dx.doi.org/10.1071/HE06048
- VanWormer, J. J., Martinez, A. M., Cosentino, D., & Pronk, N. P. (2010). Satisfaction with a weight loss program: What matters? *American Journal of Health Promotion*, *24*(4), 238–245. doi:10.4278/ajhp.080613-QUAN-92
- Ware, J. E., & Hays, R. D. (1988). Methods for measuring patient satisfaction with specific medical encounters. *Medical Care*, *26*(4), 393–402.

CHAPTER 10

Program Quality and Fidelity: Managerial and Contextual Considerations

This chapter reviews the current trends and approaches to managing healthcare organizations that collectively form the context within which a health program must function. The trends and approaches described here are important for the influences they exert on healthcare organizations. The assumption is that health programs are based in a wide range of types of healthcare organizations, such as federal agencies, state and local health departments, non-profit and for-profit health systems, community-based nonprofit organizations, church-affiliated organizations, and international relief and assistance organizations. Commonalities exist across these organizations, with many being subject to the same legal constraints and obligations. Most healthcare organizations experience similar pressures to ensure and document the quality of health care provided, and all face a growing need for electronic information systems. The content of this chapter is intended to illuminate that broader

healthcare environment as a contextual influence that managers and planners of health programs must navigate to achieve high program quality and intervention fidelity.

Of primary importance is accountability, and thus the chapter begins with a discussion of this topic. It is followed by a review and synthesis of the currently pervasive approaches used by healthcare organizations to achieve and maintain healthcare quality. Most health programs need to fit within the broader quality system used by the healthcare organization. Processes dealing with ensuring a desired level of quality can lead to change efforts intended to remedy a quality problem or enhance the existing quality of the program. Whether on a small or large scale, managers and directors of health programs need to draw on their knowledge of and skills at facilitating group change. For this reason, the chapter includes a review of key group process concepts that are related to making changes happen.

► The Accountability Context

Accountability and responsibility are cornerstones for program implementation. *Accountability* means being answerable for actions taken and the subsequent success or failure of the program. *Responsibility* means being charged with ensuring that things are done, and done within the specified parameters. Program managers are generally both accountable for the program and responsible for seeing that the program is carried out. Accountability with regard to program implementation fits within the twin realms of program accountability and professional accountability.

Program Accountability

Program managers are accountable for the program in six areas, as summarized in **TABLE 10-1**. Each area of accountability requires some thought, planning, and oversight. In other words, through careful attention to the organizational plan and the services utilization plan, each type of accountability can be achieved.

Two types of accountability relate to the organizational plan: efficiency and legal. *Efficiency accountability* means that the program is delivered with efficient use of the resources. *Legal accountability* means that staff member act in accordance with local, state, and federal laws and within their professional licensure limits.

Two types of accountability relate to the services utilization plan: coverage and service delivery. *Coverage accountability* relates to the program reaching the intended recipients; it is documented with the calculation procedures described elsewhere in the text. *Service delivery accountability* comprises the extent to which the intervention is provided as planned. It is indicated by not only the number of units of service provided but also by the number of times that the program intervention protocol was not followed or by the number of changes made to the intervention.

One type of accountability relates to the effect theory: impact accountability. *Impact*

accountability is concerned with the program having intended outcomes on the intended audience and recipients. The indicators for impact accountability are highly tailored to reflect the effect theory of the program.

Professional Accountability

Professional accountability refers to an individual from a health profession being bound by the corresponding professional norms and codes, including the moral and ethical codes related to serving the public interest. Members of the multidisciplinary team may be licensed and required by law to practice within those boundaries. These professional ethics and norms generally provide a broader, more encompassing, and not as well-codified set of rules to guide professional behavior. Professional accountability becomes important to programs in a variety of ways that can influence the program planning or implementation, with its concerns typically falling into the realm of either personal professional accountability or generic public health professional accountability.

Personal professional accountability relates to the extent to which an individual is performing according to professional standards and norms. This mandate creates the need for managerial support and possible supervision to ensure that professionals are held accountable for meeting their professional standards. Failure to perform according to professional standards should be rare among the program staff members. If it does occur, however, a program manager should seek help from the human resources department to resolve the issue.

The more complex concept of generic public health professional accountability is concerned with the extent to which the program addresses social justice and disparity problems. Given that program budgets are always tight, with correspondingly specific program eligibility criteria usually established, some individuals in need of the program and who would participate may potentially not be accepted into the program. This situation has the potential to create an ethical dilemma that puts professional accountability

TABLE 10-1 Types of Program Accountability, with Definitions and Examples of Process Evaluation Indicators (Source: Adapted from Rossi, Lipsey, and Freeman 2004.)

Accountability Type	Definition: The Extent to Which	Examples of Indicators
Organizational Plan Related		
Efficiency	Resources are utilized without waste or redundancy	Dollars spent on the program, cost per client served, cost per unit of outcome
Fiscal	Resources are managed according to the budget	Existence of receipts and bills paid, number of errors found during annual audit, percentage variance from budget
Legal	Legal, regulatory, and ethical standards are met	Number of malpractice suits, number of investigations, number of personnel with current licensure
Service Utilization Plan Related		
Coverage	The target population is reached	Coverage efficiency, coverage bias (differing participation across subgroups)
Service delivery	The intervention is provided as planned	Number of units of service provided, number of breaches of intervention protocol, number of modifications to intervention
Effect Theory Related		
Impact	Participants change or are changed because of the intervention	Very program-specific health and behavior indicators

at odds with the program limitations. In other words, public health professionals who have social justice as an element of their professional accountability may view the program as being at odds with their professional accountability. The takeaway message is that wise program planners and managers never lose sight of the relevance of professional accountability in their program and take that issue into consideration during the planning stages.

► Performance and Quality: Navigating the Interface

Health programs exist within the complex and constantly evolving healthcare system, which is affected by fads and managerial fashions as much as by new scientific evidence. One fad

that has become standard fare is attention to quality through a variety of now well-established procedures and approaches.

The focus on ways to improve the quality of health care and health services began with Donabedian's work (Donabedian, 1966). Donabedian (1966, 1980) was the first in health care to suggest taking a systems approach by investigating structure, processes, and outcomes. His approach focused attention on organizational processes involved in providing care, which have no standardized reporting, in contrast to the standardized reporting of diagnoses. Over time, however, well-developed approaches to studying processes became accepted as a means of improving organizational efficiency and effectiveness. The more widely known and used approaches are summarized here, and their relevance to managing health programs is discussed in some depth.

The quality improvement approaches discussed in this section are distinct from the older yet still very important field of quality assurance. *Quality assurance* entails using the minimum acceptable requirements for processes and standards for outputs as the criteria for taking corrective action. For example, laboratory tests must be conducted within strict parameters that ensure accuracy of testing. The accuracy of laboratory processes and tests is assessed regularly against those standards and requirements. Quality assurance can be an important element of ensuring that program interventions are delivered as planned and according to the program standards. In other words, quality assurance teams typically seek to identify errors and bring processes into compliance. The corrective managerial actions generated from quality assurance programs tend to stress following procedures rather than identifying larger, more systemwide program process improvements. This relatively narrow focus explains why quality assurance does not foster or result in overall improvement but rather serves to maintain a minimum standard.

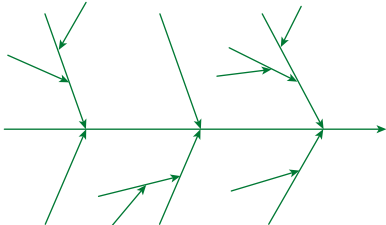
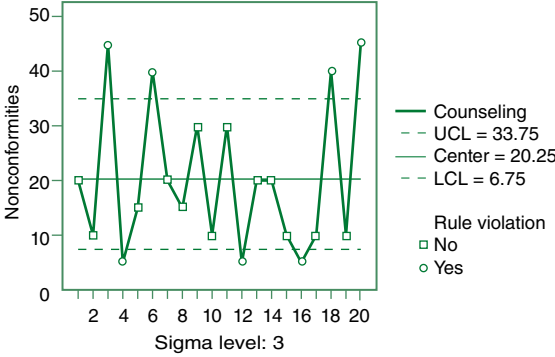
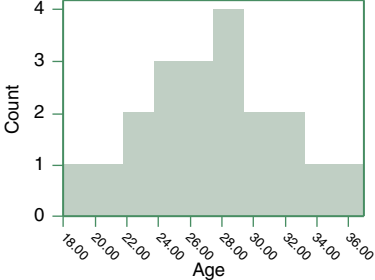
Quality Improvement Approaches

Continuous quality improvement (CQI) (Juran, 1989) was adopted by healthcare organizations as a tool to reduce costs while improving the quality of services. By the 1990s, CQI had become a popular way to enhance organizational effectiveness and was commonplace in healthcare organizations (Shortell et al., 2000). CQI is based on the premise that problems are best addressed through attention to the system as a whole and that employees are the best source of possible solutions. For the most part, these approaches focus on examining organizational processes using statistical and other scientific tools. The explosion of attention to quality in health care has led to the development of other packaged approaches, such as Six Sigma and Lean methodology. These proprietary systems are similar because they focus on making organization-wide changes to improve both the quality of services and the overall performance of the organization. Health program staff members may become involved in or may become affected by the actions of those in the organization who are applying these approaches.

In addition to systems theory, CQI has origins in engineering and operations management. Over time, CQI has evolved into more generic ongoing processes for assessing the inputs into key organizational processes that influence the use of resources and influence patient outcomes. As ongoing organizational processes, quality improvement efforts are conducted by standing quality improvement committees that include employees who are directly involved in the processes being addressed.

Quality Improvement Tools

The tools used by improvement committees rely on statistical analyses and graphic displays of the statistical information. Seven basic tools are used to control the processes statistically (**FIGURE 10-1**).

Tools	What the Tool Does	Visual Example
Cause-and-effect diagram, Ishikawa or Fishbone chart	Identifies many possible causes for an effect or problem, sorts contributing causes into sequenced and useful categories	
Check sheet	Form for collecting and analyzing data	<p>Check <u>Data to Be Reviewed and Analyzed</u></p> <ul style="list-style-type: none"> <input type="checkbox"/> Name recorded <input type="checkbox"/> Age recorded <input type="checkbox"/> Visit at recommended interval <input type="checkbox"/> Attended health promotion class
Control charts	Graph shows values relative to upper and lower control limits set at 3 sigma (standard deviations) for one variable either across individuals as shown or across time (not shown)	<p>Control Chart: Counseling</p> 
Histogram	Shows frequency distributions for one variable (used age of 20 women in the NTD prevention program)	

(continues)

FIGURE 10-1 List of Quality Improvement Tools with Graphic Examples

Tools	What the Tool Does	Visual Example
<p>Pareto chart</p>	<p>Bar graph to help identify the few “problem” individuals or variables that create the majority of the non-conformity to the process (used 0 = no screening, 1 = yes screened to understand who exceeds the recommended 20 minutes, numbers above bars represent numbers of persons)</p>	<p>The top Pareto chart shows counts for counseling minutes: 10 (1 person), 20 (1 person), 5 (1 person), 40 (2 persons), and 30 (1 person). The bottom Pareto chart shows counts for counseling minutes: 10 (4 persons), 20 (3 persons), 5 (3 persons), 40 (1 person), 15 (2 persons), and 30 (1 person). Both charts have a secondary y-axis for 'Percent screen' ranging from 0% to 200%.</p>
<p>Scatter diagram</p>	<p>Graph shows relationship between pairs of numerical data with one variable on each axis (used age and number of minutes of counseling, dots represent individuals)</p>	<p>The scatter diagram plots Age (y-axis, 20.00 to 35.00) against Counseling minutes (x-axis, 10.00 to 40.00). Data points are scattered, showing a weak positive correlation.</p>
<p>Flowchart</p>	<p>Shows separate process steps in sequential order, with connections between processes and end points</p>	<pre> graph TD Start([]) --> Contact[Contact woman] Contact --> Decline{Decline} Decline --> Document[Document] Contact --> Agree{Agree to come} Agree --> Create[Create class and menu] Agree --> Count[Count women coming] Agree --> Give[Give directions] Create --> Shop[Shop for food] Count --> Shop Give --> Arrive{Women arrive} Arrive --> Cook[Do cooking class] Cook --> Eat[Eat healthy food] </pre>

FIGURE 10-1 List of Quality Improvement Tools with Graphic Examples (continued)

These tools are easy to use and require minimal statistical knowledge, which probably accounts for their wide application. The creation of the displays and the data displayed both contribute to insights for process improvement.

Program evaluation review technique (PERT) charts diagram the sequence of events against a specific time line, thereby showing when tasks need to be accomplished. *Fishbone diagrams*, or cause-and-effect diagrams, represent sequential events and major factors that produce a given outcome and are often used to diagnose problems. *Control charts* show an average, with upper and lower confidence limits and standard deviations. They indicate whether a variable is within the acceptable parameters, and they result in a heavy focus on setting and staying within control limits and parameters for a select set of outcome indicators. *Histograms* are simple bar graphs showing the frequency of a value for one variable. *Pareto charts* are a bit more complicated—they use a bar graph to identify the major source of a problem. *Scatter diagrams* show the relationship between two variables by using the data from each individual. They make it easy to see the direction of the relationship. *Flowcharts* diagram the sequence of activities from start to outcome.

Since their initial application in health care, CQI methodologies have been updated by popular, albeit similar, approaches to process improvement. For example, Six Sigma is a process to reduce variation in clinical and business processes (Lazarus & Neely, 2003). Lean Six Sigma, the integration of Lean manufacturing techniques with Six Sigma, adds a focus on minimizing waste (Koning, Verver, Heuvel, Bisgaard, & Does, 2006; Pepper & Spedding, 2010; Ahmed, Manaf, & Islam, 2013). Lean Six Sigma, when used, for example, in medical settings, can reduce the time that patients wait for services, reduce the number of charges not billed to insurance, and reduce the length of time to make medical diagnoses. All of the quality improvement techniques use a data-driven approach to analyze process variations as the basis for taking corrective actions to control

the degree of variation in the process. These variations are attributed to actions that deviate from the specified protocol product.

Another widely adopted approach to informing continuous organizational improvement is the Balanced Scorecard. The Balanced Scorecard was developed by Kaplan and Norton (1992, 1996) as a tool for businesses. It integrates financial performance measures with measures of customer satisfaction, internal processes, and organizational learning. Although initially developed for for-profit businesses, the Balanced Scorecard approach has since been adapted for use in the nonprofit sector (Urrutia & Eriksen, 2005), as well as the British National Health Service (Radnor & Lovell, 2003) and in other countries (e.g., Peters et al., 2007). By showing trends in organizational learning and internal processes as well as the customer satisfaction and profit that should ensue, the Balanced Scorecard seeks to make healthcare organizations more market-oriented and customer-focused, promote effective communication among units within an organization, improve accountability throughout the organization, and provide managers continual feedback for making necessary adjustments (Inamdar, Kaplan, & Reynolds, 2002).

Relevance to Health Programs

Evaluations can be affected by the presence of CQI and related methods in several ways. Mark and Pines (1995) suggest that the evaluations are influenced by whether an organization is engaged in such a process improvement methodology because employees will already be sensitized to the use of data, will have had an introduction to data analysis methods, and will be accustomed to participating in analytic and change activities. Staff members, because of their participation in process improvement teams, may expect to be involved in the development of a program and its evaluation. For all these reasons, involving program staff members in evaluations may be slightly easier in organizations using CQI and related approaches to data-based process improvement.

It also may be easier to develop program theory in organizations that use process improvement methodologies because staff members will already have training and knowledge of techniques that can be very useful in program planning, especially PERT charts, fishbone diagrams, and control charts. These skills help personnel articulate and then construct diagrams of underlying processes, especially when program planners are developing the process theory. They also make the CQI way of thinking and methods useful in designing and conducting process monitoring evaluations. In particular, familiarity with CQI methods facilitates the identification of problems related to the implementation of the service utilization plan and deficiencies in the organizational plan.

Even organizations with strong process improvement processes need program evaluations, however, especially outcome or impact evaluations. When they were originally implemented, the quality improvement approaches emphasized a strict focus on processes as they

affected outputs. More recently, improvement approaches have begun encompassing consideration of outcomes. Evaluators and program planners working in healthcare organizations will inevitably experience the continual introduction of new approaches to improve the processes and outcomes of those organizations. Because such methodologies direct attention toward solving problems, program evaluators need to be sensitive to how current process improvement approaches might influence program development and implementation, and hence the evaluation. Organizational process improvement approaches differ from evaluation with regard to the underlying philosophy, the purposes, the personnel who carry out the activity, and the methods used (TABLE 10-2).

Performance Measurement

As with most terms in evaluation, there is no consensus on the definition of the term *performance measures*, but the following generic definition

TABLE 10-2 Comparison of Improvement Methodologies and Program Process Evaluation

	Process Improvement Methodologies	Program Process Evaluation
Philosophy	Organizations can be more effective if they use staff expertise to improve services and products	Programs need to be justified in terms of their effect on participants
Purpose	Systems analysis and improvement focus on identified problem areas from the point of view of customer needs	Evaluators determine whether a program was provided as planned and if it made a difference to the participants (customers)
Approach	Team-based approach to identifying and analyzing the problem	Evaluator-driven approach to data collection and analysis
Who does it	Employees from any or all departments, mid-level managers, top-level executives	Evaluators and program managers, with or without the participation of employees or stakeholders
Methods	Engineering approaches to systems analysis	Scientific research methods

gives a sense of what they are. *Performance measures* are indicators of process, output, or outcomes that have been developed for use as standardized indicators by health programs, initiatives, practitioners, or organizations.

The passage of the Government Performance Results Act of 1993 placed a new and stronger emphasis on performance measures. The act requires that each federal governmental agency “express the performance goals for a particular program activity in an objective, quantifiable, and measurable form” (Office of Management and Budget, n.d.). This mandate has led federal agencies that fund health programs, whether as research projects or national infrastructure building, to develop and implement performance measures to which they hold their grantees accountable. The trend toward performance measures continued with the No Child Left Behind Act of 2001 and with the Common Core Standards for kindergarten and grades 1 through 12 performance in mathematics and the English language arts (Porter, McMaken, Hwand, & Yang, 2011). As Berry and Eddy (2008) point out, the emphasis on standardized testing and use of evidence in education is affecting the practice of evaluation.

The criteria for good performance measures include the following characteristics (Krumholz et al., 2006). Performance measures must be useful in improving intended audience and recipient outcomes by being evidence-based. They must also be interpretable by practitioners and actionable by improvement committees. Measurement implementation to obtain the performance measures data must be feasible.

TABLE 10-3 shows the key elements generally used in performance measures. Performance measures, including those approved by the federal Office of Management and Budget (OMB), will have these highly specific characteristics. In addition, performance measures vary across accrediting bodies and the various performance measurement systems (**TABLE 10-4**).

The assumption is that the necessity of reporting on performance measures forces program managers to pay attention to what

is measured and subsequently to undertake improvement efforts for performance measures not achieved or reached. In reality, the evidence in support of this premise is equivocal. For example, Bradley and colleagues (2006) found only a small correlation between performance on a set of performance measures related to acute myocardial infarct care and patient outcomes. The method of data collection, whether self-reported or provider assisted, may influence outcomes (Gelkopf, Pagorek-Eshel, Trauer, & Roe, 2015). Nonetheless, even the marginal or small improvements in outcomes that result from attention to performance measurement may contribute cumulatively or over time to a clinically significant improvement in health outcomes.

Relevance to Health Programs

The list of performance measure characteristics described earlier reveals that performance measures are fairly similar to program objectives in terms of both intent and content. Thus, when developing process and outcome objectives, it may be important to align the objectives with the performance measures or to have objectives that are the performance measures on which the program needs to report. Achieving such cohesion, of course, requires that program planners, managers, and evaluators communicate and share full information on the performance measures required by funding agencies. In addition, it is incumbent upon the evaluator to explain the connection between performance measures and the program evaluation, as well as to help reduce redundancies in what is being tracked and measured.

Informatics and Information Technology

Informatics and information technology are vital to developing, managing, and evaluating a health program. The appropriate information technology for the program should have been carefully considered and chosen during the

TABLE 10-3 Definitions of Terms Used in Performance Measurement

Performance Measurement Element	Definition
Measure type	Broad health status that the performance measure is intended to describe
Measure	Statement in measurable terms of the desired health status or behavior as it relates to the measurement type
Numerator	Definition by which to assign individuals into the numerator to quantify the measure
Denominator	Definition by which to assign individuals into the denominator to quantify the measure
Rationale for measure	Brief explanation of the relationship of the measure to the measurement type; includes reference to the appropriate <i>Healthy People 2020</i> objective
Limitations of measure	Statement of which key factors might contribute to the potential failure of program, which would be captured in the measure
Use of measure	Suggestions for how the measure might be used in program development or policy making
Data resources	List of relevant sources of data for estimating the measure
Limitations of data	Brief statement of factors that may contribute to inaccurate, invalid, or unreliable data from the data resource listed

planning phase. Having the technology is only of value if the data collected are first turned into information and then knowledge. With this in mind, the topic is addressed here as it relates to program quality and intervention fidelity.

One potential advantage of having a health information system in place is that it contains and makes available very current data. Most health information systems perform nearly real-time data collection. If program managers or evaluators have access to this information, then they can obtain very current data on the implementation of the program interventions.

Timely access to process data can be used on a regular basis to monitor for any deviations from the intervention as well as for the program's closeness to the targets specified in the process objectives. Naturally, an organization's ability to maximize its access to this type of very useful information depends on program planners having included the key variables in the data being collected and on having negotiated for access to or reports from the database.

Another potential advantage can be derived from implementation of an effective health information system. It is possible that data needed

TABLE 10-4 Partial List of Existing Performance Measurement Systems Used by Healthcare Organizations, with Their Websites

Name (Abbreviation)	Sponsoring Organization(s)	Website
Capacity Assessment for Title V (CAST-5)	Association of Maternal and Child Health Programs and Johns Hopkins University	www.amchp.org/program/sandtopics/CAST-5/Pages/default.aspx
Consumer Assessment of Health Plans (CAHPS)	Agency for Healthcare Research and Quality	http://www.ahrq.gov/cahps/index.html
Healthcare Effectiveness Data and Information Set (HEDIS)	National Commission on Quality Assurance	www.ncqa.org/tabid/59/Default.aspx
The Joint Commission	The Joint Commission	https://www.jointcommission.org/performance_measurement.aspx
National Public Health Performance Standards (NPHPS)	CDC and a variety of public health organizations	http://www.cdc.gov/nphps/index.html

for process monitoring or effects evaluation are already being collected. This is especially likely if the health program is implemented by a health system or plan. The vast majority of health systems and plans already collect key indicators of care and service quality through the Healthcare Effectiveness Data and Information Set (HEDIS) and performance measures for The Joint Commission. Most healthcare organizations have established electronic information systems that are designed to gather data on the performance indicators for these two systems. Knowing whether the program organization collects these data can help the program planner and manager.

For example, consider the Bove County program to increase adult immunizations. One of the HEDIS measures is the percentage of adults older than age 65 who have received influenza immunizations. Thus, if one process objective addresses physician ordering of immunizations, then it would be possible to collaborate with the health systems in Bove County to receive quarterly reports on this HEDIS measure.

► Creating Change for Quality and Fidelity

The purpose of collecting, analyzing, and reviewing process data is to create a feedback loop within the program that both leads to improvements that strengthen the program implementation and maintains the optimally functioning elements of the program. The feedback loop consists of first interpreting the process data, then formulating and implementing corrective managerial or programmatic actions. In other words, changes are made to ensure program quality and fidelity.

Interpreting Implementation Data

Before any actions are taken to modify the program processes, data from the process evaluation need be converted into information. Implementation

data collected about the program processes become information when the data are given meaning.

Giving the data meaning begins by comparing the data from the process monitoring with the targets stated in the process objectives. Ideally, each Time frame, Amount of Action by which Participant/Program Staff (TAAPS) objective dealing with the organizational plan and service utilization plan will be compared with actual program accomplishments as documented in the process evaluation. Assessing the efficiency index is one way to determine the extent to which process objectives were met. An alternative is to use a simple tally of the number of process objectives met, exceeded, or not met. Both approaches summarize the process data in a way that allows the program evaluator, manager, staff members, and stakeholders to make sense of the program's accomplishments and shortfalls. Part of the process of making sense of these data is to determine the sources of gaps between target values and actual performance.

Attention to coverage is a critical area for interpretation because of its importance in determining the cause of program failure should the effect evaluation reveal no programmatic effect on the participants. If the program is delivered to individuals who are not in the intended group and hence would not benefit from the program, then the intervention will appear to have no effect. Understanding the reasons for under- or overcoverage is tricky. The data will reveal whether a programwide problem exists; if one does, a qualitative investigation could then be conducted to understand what occurred. Such an inquiry into the perceptions of participants (especially those not in need of the program who are participants) and of program staff members would provide additional data as a basis for interpretation. Also, the interpretation may suggest that the extent to which the objectives were met was a function of the objective targets not being realistic, in terms of being either too high or too low.

For health programs that are replications, an alternative approach is to use the information from the other programs as a benchmark.

Benchmarking is the process by which the level of attainment by an exemplary program (the benchmark) is used as a point of comparison to current level of achievement. The difference or gap between the benchmark and the current achievement indicates the degree of improvement needed. Benchmarks are easily understood by program staff members and stakeholders, and they can be useful if the source of the benchmark is highly regarded. However, benchmarking provides no insights into what needs to be changed in order to match the level of the benchmark. Thus, it is best used in conjunction with other quality improvement tools and methods. Program staff members should be included in discussions about the degree of congruence found between the objectives and the program's actual accomplishments. Their insights can provide plausible explanations for gaps and reveal the need to collect other information related to the cause of the gap between desired and actual program implementation.

Overcoverage is equally likely to result from poor program administration as from the program inadvertently addressing an unrecognized need. When overcoverage occurs, recruitment, screening, or enrollment procedures may be faulty or inadequate. These administrative issues can be addressed with staff training and revisions of the procedures. Overcoverage resulting from an unrecognized need is more difficult to assess and may require further study.

Consider an immunization clinic for school-age children. If many children are brought for immunizations when those children are already fully immunized, then overcoverage exists. In fact, the unrecognized need in this case may be parental confusion regarding school immunizations or lack of a centralized immunization tracking system.

Another example of overcoverage would be more than the expected number of senior citizens attending a free lunch program. The unrecognized need may be for socialization opportunities for the seniors of the community, or perhaps the free ride to and from the luncheon site is the only way for the seniors to get to a nearby mall. In

this case, overcoverage can be thought of as a form of expressed needs. Thus, the reason for overcoverage, if it is not administrative, may lead to new programs.

Undercoverage is also a possibility. In such a case, barriers to access, availability, acceptability, and affordability are the most likely causes of the problem. Undercoverage also results from inadequate program marketing. One other possible cause of undercoverage stems from a faulty community assessment that overestimated the number in need of the program.

Another consideration is the extent to which program components are provided. This is especially important for complex programs that incorporate interdependent program components. Using the Layetteville program for violence prevention as an example, program components include education, parental support, and policing. If any one of these components is lacking, the other components might fail to show any effects because the components were designed to have synergistic effects. The process data that reveal a missing component challenge program planners and managers to interpret that information in terms of feasibility, the degree to which the objectives were realistic, and events in the context of the program.

Maintaining Program Process Quality and Fidelity

The analysis and interpretation of the implementation evaluation data should lead to identification of actions that can be taken to maintain or improve the program processes. Of course, process data are valuable only if they are used in a timely manner to make alterations in the program implementation. In this regard, implementation monitoring is similar to the various improvement approaches and methodologies already discussed. Once the areas of deficit or weakness or the unachieved objectives have been identified, action is required. Findings from process evaluations may lead program managers to take one of two possible avenues of action.

Implementation-Focused Actions

One set of actions comprises managerial actions focused on the day-to-day implementation processes and activities related to the organizational plan or the services utilization plan. Actions may address, for example, the coverage aspect of the service utilization plan. Suppose undercoverage results from a small program that attempts to tackle a big problem. In this instance, documentation of the undercoverage, along with documentation of program success in achieving the desired effects, can be used to seek funds to expand the program. If serious undercoverage is found in the process evaluation, then the social marketing plan that was implemented should be reviewed and perhaps revised based on the available process evaluation data about the participants. In addition, attention may need to be directed toward reviewing the cultural acceptability of the program.

If the program has been ongoing for several years and the indices of efficiency are declining, then it may be time to fine-tune the program (Rossi & Freeman, 1993). This might entail updating the needs assessment to have more current and accurate numbers for the population in need, revising program objectives to reflect current practice and situations, altering the marketing strategy, or adjusting elements of the organizational plan. The specific fine-tuning would depend on the process data and the interpretations made in regard to those data.

A report with a description of how the process data were collected and an outline of the connections between the data and the program objectives is usually a product or output of the process evaluation. The report should address organizational plan objectives and achievements, as well as service utilization plan objectives and achievements. The report, or appropriate portions of it, can be shared with program staff members, funding agencies, and other stakeholders. A periodic process evaluation that is shared with program staff members can be used to identify factors that contribute to the success or poor

performance of the program. Findings from the process evaluation can be shared in ways that maintain individual confidentiality but allow staff members to see their productivity in comparison to that of other staff members. Reasons for higher and lower productivity that are amenable to managerial intervention might be identified. In addition, sharing the process evaluation data with staff members provides an opportunity for personnel to express their concerns, point out challenges in complying with intervention specifications, and demonstrate their overall morale as it relates to the project.

Process Theory–Focused Actions

The other set of actions that can be taken in response to an evaluation focuses on revising or modifying aspects of the process theory, excluding the interventions themselves. The interventions, which are part of the effect theory, are excluded from revision until sufficient effect data are available to determine whether the types and dosage of interventions are actually contributing to the desired health outcomes. However, the quality and fidelity of delivery or provision of the interventions are appropriately the focus of actions.

This fine distinction can be seen in the following example. Suppose that adult immunization clinics are scheduled for early Monday afternoons, but the attendance records indicate poor attendance. Some informally collected qualitative data indicate that the seniors go for free lunches at that time. Then, in interpreting the low percentage of reach and managing the program to achieve its target objectives, it would be reasonable to change the dates for the clinic and increase the social marketing efforts, without changing the immunizations provided or the eligibility criteria. This example shows the extent to which much of the revision to the process theory is just common sense. Although a change in the day of the immunization clinic is a simple solution, it may not be simple to implement because of other constraints and organizational factors. In some cases, a simple solution may require considerable effort to make the change.

Managing Group Processes for Quality and Fidelity

Identifying the aspects of the process implementation that are in need of modification, refinement, or change is useful only to the extent that the program staff members can then be motivated to make the changes. Managerial capabilities—an organizational plan input—become important at this juncture. Two considerations are the natural individual and group negative reactions to being evaluated, and the corresponding skills needed to address such reactions.

The natural inclination toward being evaluated, especially on work performance, is to become defensive. When the program manager and evaluators bring to the program staff members a set of recommendations for change, the staff members may interpret this action as a poor performance evaluation. Also, the program staff members' reaction is likely to be defensive. The approach used to gain support from the program staff for the proposed modifications can determine the success or failure of making the needed modifications, so managers need to utilize skills related to managing groups (specifically, the program staff members) as part of this approach.

Group process skills include communicating, motivating, identifying a purpose and direction setting, setting group norms, understanding stages of group formation, cooperation building, and providing guidance but not control. This partial list of group skills hints at the potential complexity of steering a group toward a desired set of actions and processes. The program manager may use a wide range of strategies to ensure the quality and fidelity of the program implementation. Drawing on network theory, he or she may choose to win over a central and well-connected staff member as a way to influence the broader net of staff members. Alternatively, the program manager may draw upon the theory of group formation (forming, storming, norming, performing; (Tuckman & Jensen, 1977) if a new program has brought together a new team to implement the health program. If the program staff members

are highly motivated and professional, the program manager may instead choose to rely on the self-organization principle of complexity theory and allow the staff members to evolve in ways that enhance the quality and fidelity of the program. Whichever approach is employed, the program manager should consciously choose an action path that fits with the program and the degree of change needed.

When and What Not to Change

For all the talk about making changes to ensure the quality of the program and fidelity of the intervention, sometimes it is not advisable to make changes. The obvious time not to make changes is when activities and objectives are on track, on target, and on time. For some programs, circumstances may make it imprudent to implement changes, such as when critical stakeholders are resistant, insufficient funds are available to complete a change, or the process data are ambiguous and open to widely divergent interpretations.

If a program is based on a tried-and-true theory or is a replication of an existing successful program, then it would be best not to modify the actual intervention until sufficient outcome data are available to determine the program's effectiveness. A change to the intervention should be postponed until sufficient outcome data substantiate that the intervention is causing adverse, undesirable, or dangerous side effects or that it is not at all effective. Making a change to the actual intervention, rather than the processes, requires revising the effect theory as well as any corresponding portions of the process theory.

► Formative Evaluations

A *formative evaluation* is an assessment during the initial stages of the implementation process and preliminary outcomes. It is performed if a program is new or experimental and is under some type of political scrutiny. The term *formative* suggests that the evaluation is conducted as

the program is forming or in its early stages of implementation.

Because they are conducted before the full implementation of a program, formative evaluations can be thought of as diagnostic of early problems with the program process theory. For such evaluations to be useful, program managers need to have at least some preliminary data on whether the intervention is the “right” intervention. This may require collecting some very rudimentary but salient health outcome data. Formative evaluations also frequently entail the use of multiple methods, employing both quantitative methods and qualitative approaches. The qualitative approaches, such as focus groups and interviews, can be helpful in understanding problems from the perspective of either the program staff members or the program participants. The other requirement for a useful formative evaluation is that it is done in a timely manner, and with strong feedback and reporting to the program manager. This approach ensures that corrective modifications can be made early in the program implementation.

► Across the Pyramid

At the direct services level of the public health pyramid, accountability and attention to quality are most readily seen in the vast array of accrediting bodies and certifications for both individual practitioners and healthcare organizations that provide direct services. Depending on the health program, the quality assurance and fidelity issues are likely to be related to individual providers or to the program staff members in general. The source of program infidelity influences the action plans and the interpersonal communication needed to achieve high program quality.

At the enabling services level, accrediting bodies for enabling services and programs are also likely to hold the program managers accountable for the quality of the program. The challenges facing these program managers and evaluators will be similar to the challenges faced by their counterparts at the direct services level. The nature of enabling services, such as being

more community-based and group-focused, may make it more challenging to initiate corrective actions should process monitoring reveal poor program fidelity.

At the population services level, it can be difficult to monitor the implementation in a timely manner, which in turn makes it difficult to have a quality improvement process. This challenge does not negate the importance of maintaining the quality and fidelity of population-level interventions. The other issue at the population services level is the fact that programs may be decentralized. For example, immunization clinics are run by local health departments and large healthcare systems. But if the program is essentially a state health program, with state financial support, then the state immunization program manager will be accountable for the quality of the decentralized system.

At the infrastructure level, the management of the health program is an infrastructure element. Naturally, managers of programs targeted at the health workforce must also be held accountable for the quality of the implementation. Most of the quality improvement approaches mentioned earlier are intended to lead to changes in the infrastructure of personnel deployment, resource distribution, and process modification. The Balanced Scorecard might be particularly relevant as an infrastructure tool because it addresses the organization's ability to learn (Kaplan and Norton, 1992, 1996).

► Discussion Questions

1. Imagine that you need to explain to a community of stakeholders the difference between a quality improvement approach and a process evaluation approach. What would you summarize as being unique to each approach, and what are the similarities between the two approaches?
2. Construct a hypothetical flowchart for a program to increase the rates of immunizations for adults, beginning with recruitment and proceeding through complete immunization.

3. Look at the scatter diagram, Pareto chart, and control chart in Figure 10-1, which were based on data from 20 women in the Congenital Anomalies Prevention Program in Layetteville. For each one, write one statement that describes the key information conveyed in that chart or diagram. Based on your three statements, what would you recommend to the program manager in terms of quality improvement?
4. Of the different types of accountability, which do you view as most important, most difficult to achieve, and most valuable to stakeholders? Justify and explain your choices.

► Internet Resources

Many organizations focus specifically on quality improvement in health care.

National Committee on Quality Assurance (NCQA)

NCQA is the organization that has established the HEDIS measures. Information about HEDIS is available at <http://www.ncqa.org/hedis-quality-measurement/hedis-measures>.

American Society for Quality

The American Society for Quality (<http://asq.org/index.aspx>) has a variety of resources that are accessible from its website.

Total Quality Management Glossary

A list of terms used in CQI/TQM is given at this website: www.skymark.com/resources/qualglos.asp.

National Quality Forum

An interesting report on the relationship between choosing good measures and achieving quality of care is available at www.qualityforum.org/Measuring_Performance/ABCs_of_Measurement.aspx.

Agency for Healthcare Research and Quality (AHRQ)

AHRQ has published an excellent online book about performance measurement specifically for child health programs at www.ahrq.gov/chtoolbx/index.htm. Although this book is geared toward child groups, the content is also applicable to other populations.

Public Health Foundation

The Public Health Foundation serves as a key source of quality improvement (QI) tools for public health agencies. See its handbook on QI at www.phf.org/resourcestools/Pages/PublicHealthQIHandbook.aspx.

Ohio Department of Health

The Ohio Department of Health is collecting six measures endorsed by The Joint Commission, the Centers for Medicare and Medicaid Services, and the National Quality Forum regarding heart attacks, heart failure, and pneumonia. It is also collecting five patient safety indicators created by the Agency for Healthcare Research and Quality regarding surgical procedures. These measures can be found at www.odh.ohio.gov/healthStats/hlthserv/hospitaldata/hospperf.aspx. The Ohio example shows how much data are already being collected that may be useful for monitoring program implementation and effects.

Government Accountability Office (GAO)

The GAO produced a fascinating report following the Hurricane Katrina disaster that focused on the alignment of CDC and Health Resources and Services Administration (HRSA) performance measures regarding preparedness. Find it at www.gao.gov/new.items/d07485r.pdf.

University of Washington, Turning Point Program

The Turning Point Program at the University of Washington has published a book on performance measurement for public health organizations.

Find it online at www.turningpointprogram.org/Pages/pdfs/perform_manage/pmc_guide.pdf.

References

- Ahmed, S., Manaf, N. H., & Islam, R. (2013). Effects of Lean Six Sigma application in healthcare services: A literature review. *Reviews on Environmental Health*, 28(4), 189–194. doi:10.1515/reveh-2013-0015
- Berry, T., & Eddy, R. M. (2008). Consequences of No Child Left Behind for educational evaluation. *New Directions for Evaluation*, 117, 21–36.
- Bradley, E. H., Herrin, J., Elbel, B., McNamara, R. L., Magid, D. J., Nallamothu, B. K., . . . Krumholz H. M. (2006). Hospital quality for acute myocardial infarction: Correlation among process measures with relationship with short-term mortality. *Journal of the American Medical Association*, 296, 72–78. doi:10.1001/jama.296.1.72
- Donabedian, A. (1966). Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly*, 44(3, Part 2), 166–206.
- Donabedian, A. (1980). Explorations in quality assessment and monitoring. *Volume 1: The definition of quality and approaches to its assessment*. Ann Arbor, MI: Health Administration Press.
- Gelkopf, M., Pagorek-Eshel, S., Trauer, T., & Roe, D. (2015). Routine outcome measurement in mental health service consumers: Who should provide support for the self-assessments? *Evaluation and Program Planning*, 50, 43–46. Doi: <http://dx.doi.org/10.1016/j.evalproplan.2015.02.001>
- Inamdar, N., Kaplan, R. S., & Reynolds, K. (2002). Applying the balanced scorecard in healthcare provider organizations/ Practitioner's response. *Journal of Healthcare Management*, 47(3), 179.
- Juran, J. M. (1989). *Juran on leadership for quality: An executive handbook*. New York, NY: Free Press.
- Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard: Measures that drive performance. *Harvard Business Review*, 70, 71–79.
- Kaplan, R. S., & Norton, D. P. (1996). *The Balanced Scorecard: Translating strategy into action*. Boston, MA: Harvard Business School Press.
- Koning, H., Verver, J. P., Heuvel, J., Bisgaard, S., & Does, R. J. (2006). Lean Six Sigma in healthcare. *Journal for Healthcare Quality*, 28(2), 4–11. doi:10.1111/j.1945-1474.2006.tb00596.x
- Krumholz, H. M., Anderson, J. L., Brooks, N. H., Fesmire, F. M., Lambrew, C. T., Landrum, M. B., . . . Writing Committee to Develop Performance Measures on ST-Elevation and Non-ST-Elevation Myocardial Infarction. (2006). ACC/AHA clinical performance measures for adults with ST-elevation and non-ST-elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association Task Force

- on Performance Measures (Writing Committee to Develop Performance Measures). *Journal of the American College of Cardiology*, 47, 236–265. doi:10.1016/j.jacc.2008.10.012
- Lazarus, I. R., & Neely, C. (2003). Six Sigma: Raising the bar. *Managed Healthcare Executive*, 13, 31–33.
- Mark, M. M., & Pines, E. (1995). Implications of continuous quality improvement for program evaluation and evaluators. *Evaluation Practice*, 16, 131–139. doi:10.1016/0886-1633(95)90022-5
- Office of Management and Budget. (n.d.). Government Performance Act of 1993. Retrieved from <http://govinfo.library.unt.edu/npr/library/misc/s20.html>
- Pepper, M. P. J., & Spedding, T. A. (2010). The evolution of lean Six Sigma. *International Journal of Quality & Reliability Management*, 27(2), 138–155. doi:<http://dx.doi.org/10.1108/02656711011014276>
- Peters, D. H., Noor, A. A., Singh, L. P., Kakar, F. K., Hansen, P. M., & Burnham, G. (2007). A balanced scorecard for health services in Afghanistan. *Bulletin of the World Health Organization*, 85(2), 146–151. doi:10.2471/BLT.06.033746
- Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common Core Standards: The new US intended curriculum. *Educational Researcher*, 40(3), 103–116. doi:10.3102/0013189X11405038
- Radnor, Z., & Lovell, B. (2003). Success factors for implementation of the Balanced Scorecard in the public sector. *Journal of Corporate Real Estate*, 6(1), 99–108. <http://dx.doi.org/10.1108/09536860310465618>
- Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach* (5th ed.). Thousand Oaks, CA: Sage.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Shortell, S. M., Jones, R. H., Rademaker, A. W., Gilles, R. R., Dranove, D. S., Hughes, E. F. X., . . . Huang C. F. (2000). Assessing the impact of total quality management and organizational culture on multiple outcomes of patient care for coronary artery bypass graft surgery patients. *Medical Care*, 38(2), 207–217.
- Tuckman, B. W., & Jensen, M. A. C. (1977, December). Stages of small-group development revisited. *Group & Organization Studies*, 2(4), 419–427.
- Urrutia, I., & Eriksen, S. D. (2005). Application of the Balanced Scorecard in Spanish private health-care management. *Measuring Business Excellence*, 9(4), 16–26. doi:<http://dx.doi.org/10.1108/13683040510634808>

SECTION V

Outcome and Impact Evaluation of Health Programs

CHAPTER 11

Planning the Intervention Effect Evaluations

In the daily work of implementing a program, evaluation of intervention effects can seem like a luxury. The reality is that conducting an evaluation whose purpose is to identify whether the intervention had an effect requires considerable forethought regarding a broad range of issues, each of which has the potential to detract seriously from the credibility of the evaluation.

The intervention effect evaluation deserves the same degree of attention during program planning as does development of the program interventions; ideally, it should be designed concurrently with the program. All too often, attention is focused on developing the evaluation only after the goals and objectives are finalized and the program is up and running. Well-articulated program outcome goals and outcome objectives facilitate development of the evaluation, but insights about the program process can be gained from developing an evaluation plan. As highlighted in the planning and evaluation cycle (**FIGURE 11-1**), the planning and decisions about the effect evaluation should occur as the program is being developed.

The contents of this chapter address the broad areas of data collection and evaluation

rigor within the context of the program theory and feasibility considerations. The information presented on designs and sampling is not intended to duplicate the extensive treatment of research methods and statistics provided in research textbooks. Instead, basic research content is presented as the background for the problems commonly encountered in conducting a health program evaluation, and practical suggestions are provided for minimizing those problems. Because the focus here is on practical solutions to real problems, the suggestions offered in this chapter may differ from those usually found in research and statistics textbooks. Nonetheless, good research methods and statistics textbooks are invaluable resources and references that should be on the bookshelf of every program evaluator.

Planning the evaluation begins with selecting the evaluation questions and then developing the details of the evaluation implementation plan, similar to the details of the program organization plan. Aspects of the evaluation plan related to data collection—namely, levels of measurement and levels of analysis, as well as techniques to collect data—are discussed next. These elements of evaluations are closely aligned with research

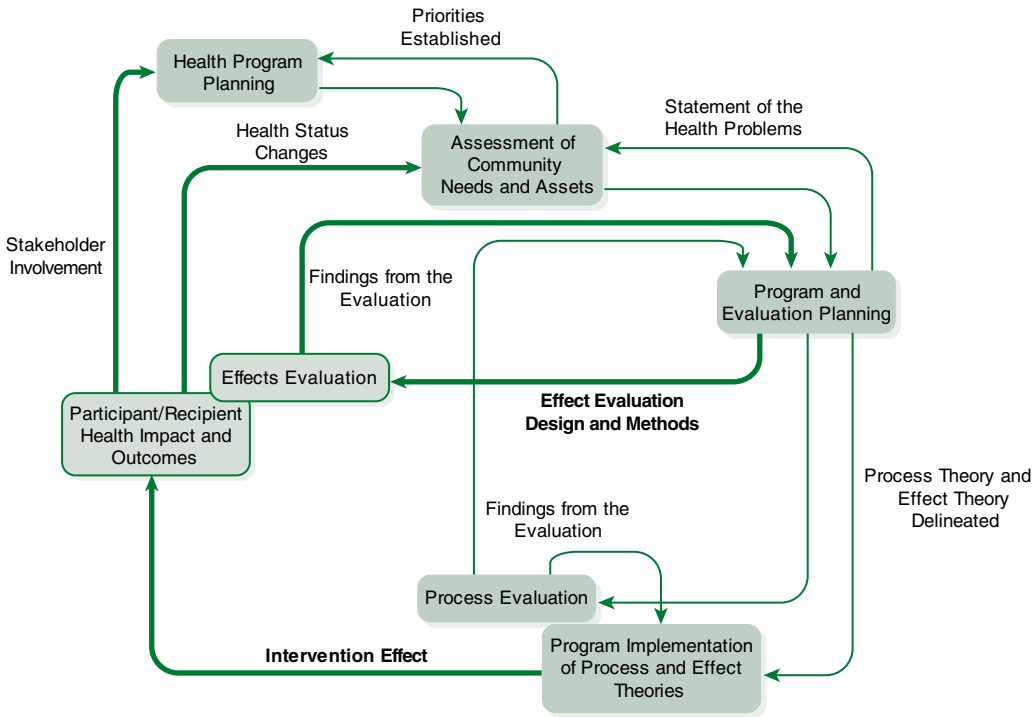


FIGURE 11-1 Planning and Evaluation Cycle, with Effect Evaluation Highlights

methodology, and achieving scientific rigor is the first yardstick used when planning the intervention effect evaluation.

► Developing the Evaluation Questions

The first step in planning the evaluation is deciding which questions the effect evaluation must be able to answer. Development of the evaluation questions actually began with development of the logic model, the effect theory, and the outcome TREW (Timeframe, what portion of Recipients experience what Extent of Which type of change) objectives. Those planning tools also form the basis for decisions about the focus and purpose of the intervention evaluation. The effect theory draws attention to the specific aspects of the

health problem that are being addressed by the program. Many aspects of the health problem and possible health outcomes of the program could potentially be addressed by an evaluation, and novice evaluators and enthusiastic program supporters will be tempted to include as much as possible in the evaluation. Succumbing to this temptation will lead to higher evaluation costs, produce an overwhelming amount of data to analyze and interpret, and distract from the essence of the program. Thus, staying focused on the outcome objectives minimizes the chance that the evaluation will become a fishing expedition. In other words, designing an evaluation can quickly lead to the development of a creative “wish list” of statements such as, “if only we knew X about the recipients and Y about their health.” The TREW objectives serve as a sounding board against which to determine whether the “if only we knew” statement has

relevance and importance to understanding whether the program was effective.

The obvious reason for doing an evaluation is to determine the effect of the program on recipients. Patton (2008) has argued that the usefulness of evaluation information should be a major reason for doing the evaluation. Nevertheless, evaluations may be conducted for other reasons, such as meeting the requirements of funding agencies. Responding to funding agency requirements can be an opportunity to engage program stakeholders and elicit their interests with regard to why an evaluation should be performed. Information that stakeholders want from an evaluation, once made explicit, then can be incorporated into the evaluation.

A key aspect of the question about doing an evaluation in the first place is determining who cares whether the evaluation is done and what it might find. There might also be a desire to prove that the program was the source of some beneficial change. Evaluations that attempt to answer causal questions are the most difficult type to perform, but causality testing provides the richest information for understanding the validity of the effect theory.

Characteristics of the Right Question

Evaluations should be useful, feasible, ethical, and accurate (Patton, 2008). To be useful, the data collected must be relevant. Relevant data may not be easily accessible, and their successful collection may require negotiation and reliance on the effect theory, as the following true story reveals. A community agency wanted to know if its program was having an effect on the substance abuse rates among school-age children. The stakeholders did not believe that data collected several years prior to the program's implementation from across the state were relevant to their community. Yet data could not be collected from children in grades 6 through 8 regarding their use of illegal drugs (an intermediate program outcome) because the school board refused to allow the evaluators into the schools. Thus, the

evaluation question of whether the substance abuse prevention intervention changed the behavior of these children could not be answered. Eventually, the program staff members restated the question to focus on whether children who had received the substance abuse prevention program had learned about the negative health effects of illegal substances (a direct program outcome). The school board was willing to allow data on this question to be collected from children.

Another characteristic of the right evaluation question is that more than one answer is possible. While this characteristic may seem counterintuitive, allowing for the possibility of multiple answers shows less bias on the part of the evaluator for arriving at the desired answer. Evaluations that are flexible and inclusive may yield multiple answers that may reveal subtle differences among participants or program components that were not anticipated. Compare the evaluation question of, "Did the program make a difference to participants?" with the evaluation question of, "Which types of changes did participants experience?" The second question makes it possible to identify not only changes that were anticipated based on the effect theory but also other changes that may not have been anticipated.

A third characteristic is that the right evaluation question produces information wanted and or needed by decision makers and stakeholders. Ultimately, the right question produces information that decision makers *can* use, regardless of whether it actually is used in decision making. The test of usefulness of the information generated by the evaluation helps avoid the fishing expedition problem and, more important, could be a point at which developing the evaluation provides feedback relevant to the intervention design.

As these three characteristics suggest, having a clear purpose for the evaluation and knowing what is needed as an end product of the evaluation are the critical first steps in developing the evaluation question. The nature of the effect evaluation is influenced by the skill and sophistication of those persons doing the

evaluation as well as the purpose of evaluation. The key factor in stating the intervention effect evaluation question is the degree to which the evaluation must document or explain health changes in program participants.

Outcome Documentation, Outcome Assessment, and Outcome Evaluation

Evaluation of the effect of the intervention can range from the simple to the highly complex. At minimum, it should document the effect of the program in terms of reaching the stated outcome and impact objectives. An *outcome documentation evaluation* asks the question, “To what extent were the outcome objectives met?” To answer this question, an outcome documentation evaluation uses data collection methods that are very closely related to the objectives. In this way, the TREW objectives that flowed from the effect theory become the cornerstone of an outcome documentation evaluation.

The next level of complexity is an *outcome assessment evaluation*, which seeks to answer the question, “To what extent is any noticeable change or difference in participants related to having received the program interventions?” An outcome assessment goes beyond merely documenting that the objectives were met by quantifying the extent to which the interventions seem related to changes observed or measured among program recipients. With this type of effect evaluation, the data collection may need to be more complex and better able to detect smaller and more specific changes in program participants. Note that the outcome assessment addresses the existence of a relationship between those persons who received the program and the presence of a change, but it does not attempt to determine whether the change was caused by the program. This subtle linguistic difference, which is often not recognized by stakeholders, is actually an enormous difference from the point of view of the evaluation design.

The most complex and difficult question to answer is, “Were the changes or differences due to participants having received the program and nothing else?” To answer this question, an *outcome evaluation* is needed. Because this type of effect evaluation seeks to attribute changes in program participants to the interventions and nothing else, the data collection and sample selection must be able to detect changes due to the program and other potentially influential factors that are not part of the program. This highly rigorous requirement makes an outcome evaluation the most like basic research (especially clinical trials) into the causes of health problems and the efficacy of interventions.

Thinking of the three levels of program effects evaluation (**TABLE 11-1**) as outcome documentation, outcome assessment, and outcome evaluation helps delineate the level of complexity needed in data collection, the degree of scientific rigor required, and the design of the evaluation.

Evaluation and Research

The distinction between evaluation and research can be ambiguous and is often blurred in the minds of stakeholders. Nonetheless, fundamental differences do exist (**TABLE 11-2**), particularly with regard to purpose and audiences for the final report. Much less distinction is made with regard to methods and designs—both draw heavily from methodologies used in behavioral and health sciences.

The differences between research and evaluation are important to appreciate for two reasons. First, communicating the differences to stakeholders and program staff members helps establish realistic expectations about implementing the evaluation and about the findings of the evaluation. As a consequence, it will be easier to gain their cooperation and feedback on the feasibility of the evaluation. Second, understanding the differences can allay anxieties about spending undue amounts of time doing research, which takes time away from providing the program, which in turn is the primary concern of program staff members.

TABLE 11-1 Three Levels of Intervention Effect Evaluations

	Outcome Documentation	Outcome Assessment	Outcome Evaluation
Purpose	Show that outcome and impact TREW objectives were met	Determine whether participants in the program experienced any change/benefit	Determine whether participating in the program caused a change or benefit for recipients
Relationship to program effect theory	Confirms reaching benchmarks set in the objectives that were based on program effect theory	Supports program effect theory	Verifies program effect theory
Level of rigor required	Minimal	Moderate	Maximum
Data collection	Data type and collection timing based on TREW objectives being measured	Data type based on program effect theory; timing based on feasibility	Data type based on program effect theory; baseline or pre-intervention data as well as post-intervention data required

TABLE 11-2 Differences Between Evaluation and Research

Characteristic	Research	Evaluation
Goal or purpose	Generation of new knowledge for prediction	Social accountability and program or policy decision making
Questions addressed	Scientist's own questions	Questions derived from program goals and impact objectives
Nature of problem addressed	Areas where knowledge is lacking	Outcomes and impacts related to program
Guiding theory	Theory used as basis for hypothesis testing	Theory underlying the program interventions, theory of evaluation

(continues)

TABLE 11-2 Differences Between Evaluation and Research (*continued*)

Characteristic	Research	Evaluation
Appropriate techniques	Sampling, statistics, hypothesis testing, and so on	Whichever research techniques fit with the problem
Setting	Anywhere that is appropriate to the question	Any setting where evaluators can access the program recipients and nonrecipient controls
Dissemination	Scientific journals	Internal and externally viewed program reports, scientific journals
Allegiance	Scientific community	Funding source, policy preference, scientific community

Research, in a pure sense, is done for the purpose of generating knowledge, whereas program *evaluation* is done for the purpose of understanding the extent to which the intervention was effective. These need not be mutually exclusive purposes. That is, a good program evaluation can advance knowledge, just as knowledge from research can be used in program development. *Evaluation research* is performed for the purpose of generating knowledge about the effectiveness of a program and, as such, represents the blending of research and evaluation.

While these three terms are often used interchangeably or ambiguously, it is easiest to think of evaluation research as research done by professional evaluators, following standards for evaluation and using research methods and designs. Evaluation research is most often an outcome assessment or an outcome evaluation. In this regard, it tends to be more complex, to be costly, and to require more evaluation skill than most program staff members have. This discussion is not meant to imply that simpler outcome documentation is not valuable. The value always lies in the evaluation addressing the right question for the program.

Rigor in Evaluation

Rigor is important in evaluation, as in research, because there is a need to be confident that the findings and results are as true a representation as possible of what happened. The playing cards are often stacked against finding any difference from a program because of programmatic reasons, such as having a weak or ineffective intervention, and because of evaluation research methods reasons, such as having measures with low validity or reliability. Rigor results from minimizing the natural flaws associated with doing evaluation, which might otherwise diminish the evaluators' ability to identify the amount of effect of the program. The net effects are those that are attributable only to the program, whereas the total change includes effects from the intervention as well as effects that are artifacts of the evaluation design (**FIGURE 11-2**), such as history, maturation of the participants, or societal changes. The purpose of the effect evaluation is to identify the net effects, so rigor is used to minimize the inclusion of nonintervention effects and design effects.

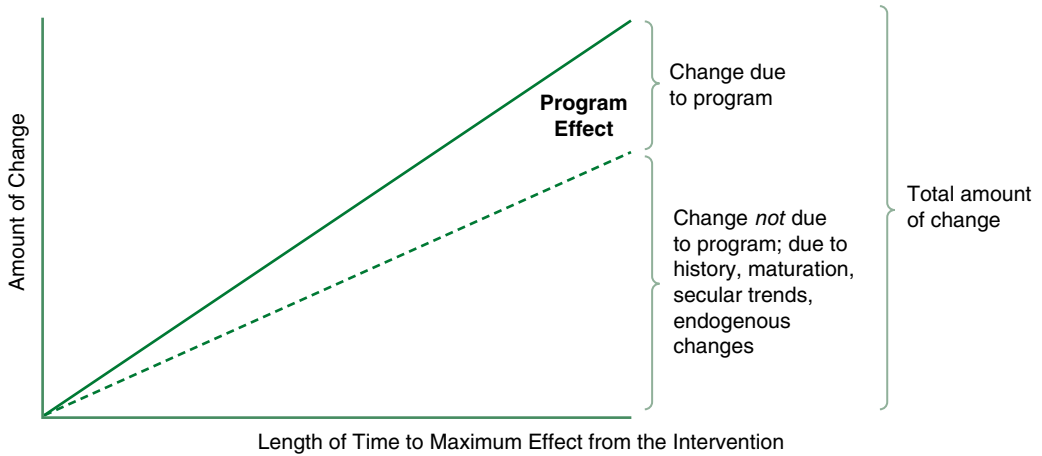


FIGURE 11-2 Diagram of Net Effects to Which Measures Need to Be Sensitive

► Variables from the Program Effect Theory

Based on the effect theory, the health program planner should have developed outcome and impact objectives that focus on what the program should achieve. Just as process objectives are useful in developing the program monitoring evaluation, so the effect theory and the outcome and impact objectives serve as cornerstones that guide decisions regarding what to measure in the intervention effect evaluation.

The effect theory is composed of three theories. The theory of cause/determinants explains the relationships among the existing factors, the main causal factors of the health problem, the moderating and mediating factors, and the health outcome. The theory of intervention specifies how the programmatic interventions change the main causal/determinant factors of the health problem, as well as the moderating and mediating factors. The theory of outcome explains how the immediate health outcomes become the longer-term health impacts. The effect evaluation uses these theories as the basis for deciding what to measure, specifically with regard to the main causal factors of the health

problem and the health outcomes. At minimum, the causal factors and health outcomes specified in the theories need to be measured.

Program evaluators use the nomenclature convention used in research and statistics. The outcome and impact variables are designated as y , the dependent variable, and the variables that precede the impact are designated as x , independent variables. Strictly speaking, any antecedent existing, causal, moderating, or mediating factor is an independent variable, as is the intervention. This nomenclature is shown in **FIGURE 11-3**. Labeling the variables as x and y at this stage of planning will help later, during data analysis.

Outcome and Impact Dependent Variables

The heart of any evaluation is *what*—that is, which health or behavioral characteristic—is assessed for a change or outcome. The “what” question should be answered directly from the “what” aspect of the outcome objective, which in turn comes from the “risk of x ” portion of the community health statement. For example, for the Bowe County diagnosis regarding African Americans older than age 40 being at high risk

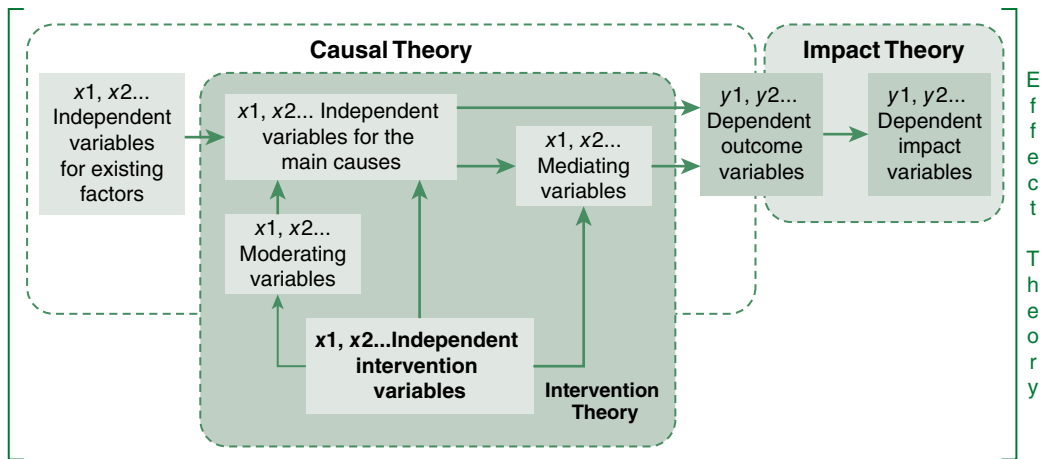


FIGURE 11-3 Using the Effect Theory to Identify Effect Evaluation Variables

for developing type 2 diabetes and its associated comorbidities, the evaluation measures the outcome in the form of diabetes incidence. The choice in this situation revolves around which laboratory test or indicator of a new diagnosis of diabetes to use—not whether to measure incidence or knowledge about diabetes.

The outcome variable, called the *dependent variable* (y), is the same as was specified in the outcome objectives. Similarly, the long-term health impact variable is also a dependent variable, as specified in the outcome objectives. The difference between these two dependent variables is a matter of the evaluation question. In the newborn health example, the outcome as a dependent variable would be used for questions such as, “Was the rate of neural tube defects lower among program participants?” In contrast, the impact variable, the rate of birth defects, would be used for a question such as, “Was having a state folic acid promotion program related to a decrease in birth defects?”

Typically, health program evaluations assess the effect of programs across various domains of health, such as physical health, knowledge, environmental health, lifestyle behaviors, intellectual and cognitive processes, emotional and mental health, social health, spiritual health, and financial resources. These domains of health and

well-being can also be related to the existing, causal, moderating, and mediating factors of the health problem of concern. During development of the program effect theory and logic model, the health program planners should have determined the role of those domains on the health problem.

In planning the evaluation, it will be crucial to refer back to those discussions and decisions as a means of deciding whether the health domain is the dependent variable or an independent variable in the intervention evaluation. At this juncture, during the discussion of what exactly is the dependent variable, program staff members and stakeholders might gain new insights into their program and want to modify the effect or intervention theory. This discussion, if it occurs before the program is implemented, can then become one of the feedback loops to program development or improvement. In any event, the outcome objectives provide direct guidance and should specify which indicators, and subsequently which measures, to use in the evaluation. If the outcome objectives fail to be helpful, this is another point at which developing the evaluation can provide refinement to the program development effort.

If the program planners were enthusiastic, energetic, and perhaps overly ambitious, the list

of outcome TREW objectives may have become extensive. The program evaluators then face a choice between measuring everything on the laundry list or selecting key outcome objectives for measurement. Attempting to measure many outcomes is another form of going on a fishing expedition. Remember, there is a 5% probability that something will be statistically significant (i.e., $p < .05$) just by chance alone. Therefore, evaluators would be wise to work with program planners and stakeholders to narrow the field from all possible outcomes to only those outcomes about which having information is critical. Also, just because an outcome can be evaluated does not translate into it being worth evaluating. For example, there is no benefit to evaluating the antibody status of the adults who attended an immunization clinic and had received the recommended adult vaccines.

For programs with multiple intervention components and multiple health outcomes, at least one health outcome per program component should be measured. For this reason, it may be difficult to construct a short list of dependent variables, especially if different outcomes are expected for each component. The theory of the intervention related to that program component helps identify the variables that are key or common across program components or are central to achieving overall program success. Those dependent outcome or impact variables need to be measured.

Who is included in the evaluation is derived from the “among [target population]” portion of the community health statement and the “recipient” in the outcome TREW objectives. The delineation of “recipient” leads to consideration of the sample and of the appropriate sampling frame given the program’s target population. If the health program has the potential to reach and affect an audience beyond known recipients, such as might happen with a mass-media educational campaign, then the issue of who constitutes “recipients” to assess becomes more complex. This decision also leads to consideration of the use of a comparison group whose members did not receive the program.

Causal Factors as Independent Variables

Evaluators may be tempted to include in the evaluation measures of various existing factors and causal/determinant factors to the health problem. Because the list of these factors can be extensive, decisions need to be made as to which factors are crucial in explaining the health outcomes of the program. For outcome documentation, these factors need not be measured; instead, only the outcomes as stated in the TREW objectives are measured. For outcome assessment and especially for outcome evaluation, however, key antecedent factors and causal factors are also important to measure. The program effect theory can help identify the specific variables in each category that should be included in the evaluation.

The causal/determinate factors are considered a set of *independent variables* because they are independent relative to the health outcome that is the true dependent variable. Whether these independent variables will be measured before and after the program, or in both participant and control groups, is a design decision that is made based on available resources and the level of scientific rigor needed in the evaluation. Causal/determinant factors, as key independent variables, are the first variables to be considered for measurement. Having data on the presence and strength of the causal factors of the health problem can provide extremely useful information that helps program planners and managers understand how the intervention effect was manifested. For example, descriptive data on causal factors can reveal patterns of intervention effects among different groups of participants and identify whether the interventions were mismatched to the causal factors found among the actual program participants.

Antecedent, Moderating, and Mediating Factors as Variables

Other factors can influence the effectiveness of interventions, such as moderating and mediating variables. Recall that a *moderating variable* affects

the strength or direction of the relationship between two other variables, whereas a *mediating variable*, sometimes called an intervening variable, is necessary for the relationship between two other variables to exist (Donaldson, 2001). The difference between these two types of variables is based on their role in the relationship between the independent and dependent variables (Donaldson, 2001). Distinguishing whether a factor functions as a moderating or mediating variable can inform the temporal sequence of data collection if it is feasible to collect mediators after the main independent variables, and definitely affects data analysis and interpretations. Critical to the outcome evaluation is the assumption that the intervention theory is correct—in other words, that the interventions actually lead to the desired change in the causes of the health problem, given the antecedent, moderating, and mediating factors.

Inclusion of moderating and mediating variables facilitates understanding which factors influence the effectiveness of an intervention. For example, Foshee and colleagues (2005)

found that prior experience with violence moderated the violence prevention program effects, whereas changes in community norms mediated the effectiveness of the program's interventions. Perhaps not surprisingly, Burke, Beilin, Cutt, Mansour, and Mori (2007) found that gender moderated the effect of an exercise program such that the program had a large effect among women. These authors also found that self-efficacy was a mediator of changes in diet. In one study (Dour et al., 2014), perceived social support had a mediating effect on the degree to which an intervention changed depression symptoms and had a lesser mediating effect on anxiety as an outcome. These studies demonstrate how having information on moderating and mediating variables provides more specific information about program effectiveness.

FIGURE 11-4 depicts the effect theory for reducing congenital anomalies in Bowe County. It shows that the factors identified through the needs assessment become variables in the effect evaluation.

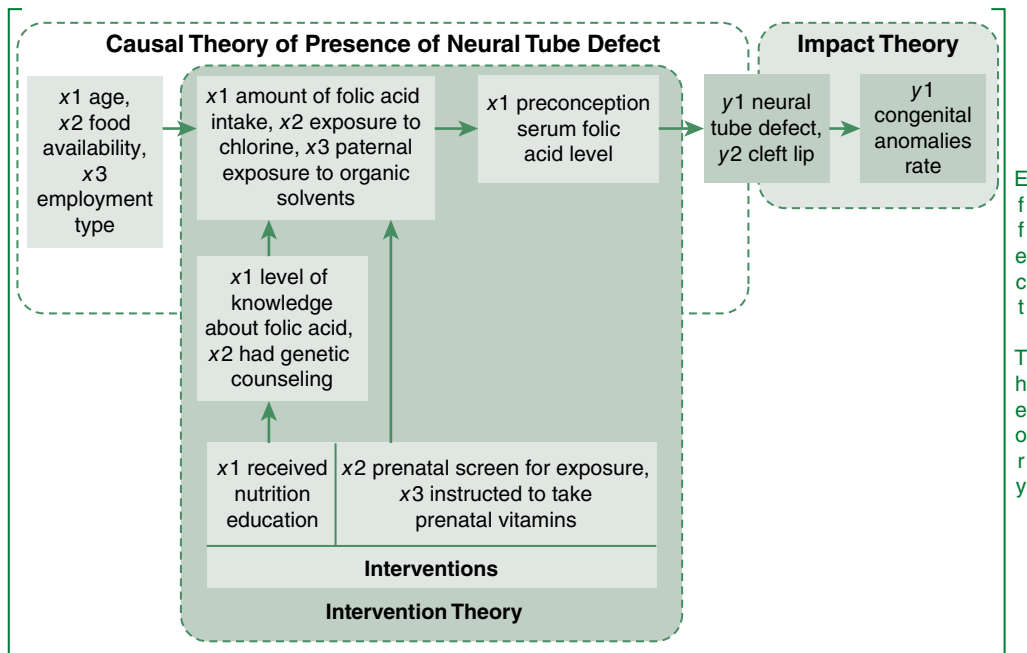


FIGURE 11-4 Effect Theory of Reducing Congenital Anomalies Showing Variables

► Measurement Considerations

Units of Observation

The unit that is observed or measured in the evaluation must match the level at which the program is targeted and delivered. For example, the violence prevention program interventions in Layetteville are targeted at schools and neighborhoods as well as at high-risk adolescents. Consequently, the effect evaluation should take place at the individual, school, and neighborhood levels. The evaluation seeking to determine the degree of change within the neighborhood might use a questionnaire that asks community members about their neighborhood. With the congenital anomalies prevention program, the intervention component is aimed at families. If the questionnaire is given to only one individual within the family and asks questions about that individual's perceptions or behaviors, then there is a mismatch between the level targeted by the intervention (family) and the unit on which evaluation data are

collected (individual). The same rationale applies to evaluating community- or neighborhood-level interventions (Best, Brown, Cameron, Smith, & MacDonald, 1989). As these examples demonstrate, the unit and level of observation are important in terms of ensuring that the effect evaluation data are collected in a way that maximizes the potential to find the intervention effects. A levels mismatch decreases that possibility.

Types of Variables (Levels of Measurement)

Another factor to consider is which type of variable to use in the evaluation, meaning how complex is the information gathered by each variable. Variables are classified according to the level of complexity of information, and each level has its own advantages and disadvantages (TABLE 11-3).

Nominal variables are the simplest in that the information indicates only yes/no, absent/present, or a name. Nominal variables that give only yes or no types of information are also

TABLE 11-3 Advantages and Disadvantages of Using Each Type of Variable

Type	Examples	Advantage	Disadvantage
Nominal categorical	ZIP code, race, yes/no	Easy to understand	Limited information from the data
Ordinal, rank	Social class, Likert-type scale, "top 10" list (worst to best)	Gives considerable information, can collapse into nominal categories	Sometimes statistically treated as a nominal variable, ranking can be a difficult task for respondents
Interval, continuous	Temperature, IQ, distances, dollars, inches, dates of birth	Gives most information; can collapse into nominal or ordinal categories; used as a continuous variable	Can be difficult to construct valid and reliable interval variables

called dichotomous; variables that have simple categories, such as job title or medical diagnosis, are called categorical.

Ordinal variables provide slightly more information by indicating an order, a sequence, or a rank. The most common ordinal variables are generated from a Likert-type scale, such as good, fair, and poor. Giving a list of items a rank can be a difficult task for respondents and creates challenges for data entry and analysis.

The most complex, rich, and complete information is provided by interval variables, in which the intervals between the values are equal on an absolute scale. Generally, interval variables are continuous, with no practical starting or ending points to the values, like the number of DNA base pairs. As a consequence, they are

also called continuous variables. However, some interval variables do have limits, such as the lab values for HbA1c, a measure of blood sugar, or the number of hospital admissions for gunshot wounds. These items are considered discrete interval variables. In terms of statistical tests, whether a variable is an interval variable is less important than whether zero can be a starting point. In health care, the issue of having a starting point, or zero value, is often quite relevant and can affect the subsequent statistical analysis. For example, the number of days hospitalized is an interval variable with a valid zero value as the starting point for the variable.

For each domain of health and well-being, variables can be constructed at each of the three levels (TABLE 11-4). Careful attention to the level

TABLE 11-4 Examples of Nominal, Ordinal, and Continuous Variables for Different Health Domains

Dependent Outcome Variable	Nominal, Categorical	Ordinal	Interval Continuous
Physiological Health Domain			
Childhood immunization	Yes/no up-to-date	None required, one immunization required, more than one immunization required	Rubella titer
Physical abuse	Yes/no have experienced physical abuse; type of abuse	Level of abuse is same, more, or less than last month	Rate of physical abuse in a county; number of times abused in past 6 months
Workplace injury	International Classification of Diseases (ICD)-10 cm for injury	Level of severity of injury	Number of disability days per year in construction industry
Knowledge Domain			
Understand how alcohol affects judgment	Agree/disagree	Three most common ways alcohol affects judgment	Score on test of knowledge about effects of alcohol

Dependent Outcome Variable	Nominal, Categorical	Ordinal	Interval Continuous
Behavioral Domain			
Breastfeeding	Yes/no breastfed	Category for how long breastfed: 2 weeks, 3–6 weeks, more than 6 weeks	Number of days breastfed
Smoking	Yes/no smoke	Light, moderate, or heavy smoker	Nicotine levels
Readiness for change	Yes/no likely to change next week	Stage of change in terms of readiness	How likely (on 100-point scale) will change in next week
Cognitive Process Domain			
Legally competent	Yes/no; Diagnostic and Statistical Manual Mental Disorders (DSM-5), code for mental illness	Level of mental incapacity	Score on standardized competency questionnaire
Mental Health Domain			
Depression	Above/below clinical cutoff on score for depression	Rank of depression on list of prevalence of mental health problems	Score on standardized depression scale, such as Center for Epidemiologic Studies Depression Scale (CES-D) or Beck Depression Inventory
Social Health Domain			
Social support network	Yes/no have friends	Degree of support in network (low, medium, high)	Number of people in social support network
Resources Domain			
Housing situation	Homeless or not	Housing autonomy (own, rent monthly, rent weekly, homeless)	Number of days living at current residence

of the measurement is crucial for several reasons. One is that the measurement level relates directly to the statistical tests that can be used with that level of measurement. Basically, a higher level of measurement (interval) enables use of a greater variety of statistics. Another reason is that the level of measurement influences whether clinically meaningful information is contained in the variable. In addition, the level of measurement must be sufficiently sensitive so that a distinction can be found among participants with regard to program effects. For instance, screening instruments used to identify potential health needs may not be sufficiently sensitive to detect the effects of programs developed to address those needs (Rossi, Lipsey, & Howard, 2004).

Timing

The timing of the evaluation is guided by the “time frame” portion of the TREW outcome objectives. The time-frame determination should be based on the theory of the intervention, which indicates the duration of the intervention effects and the amount of time needed to achieve a noticeable effect. The time-frame consideration thus provides guidance about three aspects of the intervention effects: the earliest time at which an intervention effect should be found, the time at which the maximum intervention effect is expected to occur, and the rate at which change in the health domain is expected to occur. All of this information is used in choosing methods, specifically with regard to when to collect data. An additional consideration with regard to the time frame is whether baseline or pre-intervention data are needed as part of the evaluation plan. If baseline data are needed, then decisions center on how far in advance of receiving the program it is reasonable and feasible to collect data.

If a program involves cycles, as with a series of classes, then the effect evaluation should occur after at least one group has received the full intervention. In this case, the process TAPPS objectives should be reviewed, along with the findings from the process evaluation, because, at minimum, the outcome evaluation

should be done after the program is sufficiently well implemented. Maticka-Tyndale, Wildish, and Gichuru (2007) created a time-line table that shows both intervention activities and evaluation activities. **TABLE 11-5** provides a generic example of such a time line. A time line of intervention activities and data collection, although not standard practice, is extremely valuable, especially for complex, multicomponent interventions.

Sensitivity of Measures

An interplay exists between the nature of the factor being measured and the measure itself. Of concern is that the measure be sensitive enough to variations and fluctuations in the factor that even minor changes can be detected. *Measure sensitivity* is the extent to which the difference among or within individuals can be detected. Fluctuations can occur because of individual variations, changes over time, or the influence of other factors. Given these possibilities, evaluators need to choose the measure used to collect data on the factor carefully. For many of the measures used in health care and health programs, publications of evaluations or research on the measure or studies that used the measure exist. Such reports provide information about which measures will be sensitive enough to meet the needs of the evaluation.

An example helps to illustrate this point. If one outcome objective of the health program is to reduce anemia in pregnant women, then some blood samples are used to determine anemia. Different tests can be used, depending on the program’s focus. If the program is focused on overall rates of anemia, measuring hematocrit is sufficient. If the program is focused on a specific cause of anemia, such as folic acid deficiency, a more sensitive measure of anemia is needed.

Sensitivity concerns have prompted the development of almost innumerable measures, some of which have been compiled in books, such as Cohen, Underwood, and Gottlieb’s (2000) compilation of social support measures or Bowling’s (2005) review of quality-of-life measurement scales. These and similar compendia are helpful

TABLE 11-5 Example Time Line Showing the Sequence of Intervention and Evaluation Activities

Date	Intervention Activity	Evaluation Activity
Month 1	Pilot intervention with small group	Conduct focus group to refine intervention acceptability and element of services utilization plan
Month 2	Recruit into program, screen for eligibility	Randomly assign to program or wait list, collect data for baseline and comparison Participants: $n = 150$ Nonparticipant controls (from wait list): $n = 150$
Month 3	Provide intervention (i.e., group sessions, counseling, media messages) to first group of participants	Analyze baseline pre-intervention data
Month 4	Recruit participants into program, screen for eligibility	Collect post-intervention data Participants (time 1) who completed program: $n = 125$ New nonparticipant controls (from wait list): $n = 130$
Month 5	Repeat intervention (i.e., group sessions, counseling, media messages) for those on the wait list	Analyze data
Month 6		Collect post-intervention data Current program participants (time 1): $n = 120$ Previous program participants (time 2): $n = 95$ Current nonparticipant controls: $n = 110$ Analyze data

in selecting a measure that matches the program outcome objectives. Naturally, literature searches also yield current research on instruments and tests that have validity and reliability.

► Threats to Data Quality

No matter how good the planning or the amount of effort put into generating the best possible

data, some problems are guaranteed to exist that will affect the overall quality of the data available for analysis.

Missing Data

A very common problem is missing data. Data can be missing for entire surveys or records or for single items within either. It is important to anticipate possible reasons for the missing

data. For example, survey data could be missing because of a sensitive question, a page skipped, no response category to capture a possible valid response, or an answer that is not known to the participant. Similarly, data can be missing if medical information is to be abstracted from medical records, but the data were not recorded by the provider, the provider did not gather the data, or laboratory reports are missing. Part of planning the evaluation is to imagine the worst-case scenarios in terms of data collection and to have plans to prevent those situations from occurring. Once the data have been collected, it is highly unlikely that any missing data can be retrieved. Being proactive about training data collectors, surveyors, and data abstractors and making the measurement tools as fail-safe as possible are the best steps to minimize the problem posed by missing data.

After the data have been collected, evaluators should review the data for missing items before undertaking the analysis of those data. If the missing data follow a haphazard rather than a systematic pattern, then the occasionally missed item will not dramatically affect the analysis and subsequent findings. Conversely, if there is a systematic pattern to the missing data, then decisions need to be made regarding the analysis plan. Systematic skips or extensive missing data may bias results of analyses involving that item or variable.

Reliability Concerns

One potential threat to data quality is shortcomings in the reliability of the data. *Reliability* refers to the extent to which the data are free of errors or multiple indicators of a common phenomenon coincide. Several sources of errors can diminish the reliability of the data. For example, the instrument, tool, measure, or test used to collect the data may be flawed. Another factor that can influence reliability is the age appropriateness of the questionnaire, especially when children provide the data (Varni, Limbers, & Burwinkle, 2007). The reliability of the instrument can be statistically assessed as the

Cronbach's alpha coefficient (Vogt & Johnson, 2011), which can be calculated using either statistical packages or spreadsheet programs (e.g., Black, 1999). The extent to which the instrument consistently measures what was intended can be affected by the wording of the questions.

Another source of data error is the person providing the data—in other words, the test taker. An individual may vary from day to day in terms of his or her responses, mood, knowledge, or physiological parameters. To assess the likelihood of this type of data error, the reliability of the instrument, regardless of such influences, can be statistically tested using the test–retest approach in which the same group of individuals is measured twice. Higher correlations between those two scores indicate that the instrument was constructed so that data collected are not affected by minor fluctuations within individuals.

An instrument can also be affected by differences across members of the evaluation team. In such cases, the degree to which the individuals agree is assessed with *the interrater agreement*, sometimes referred to as interobserver agreement. It is particularly important to establish high interrater agreement, meaning between individuals, when abstracting data from medical or other records. Interpretations of what is being abstracted and what needs to be abstracted can vary across individuals if they are not given specific training on the data abstraction task. Having a high degree of interrater agreement increases the confidence that the data collected across abstractors is highly comparable. The simplest approach to calculating interrater agreement is a percentage agreement, where the denominator is the number of items to be coded and the numerator is the number of items for which both raters had “correct” agreement. Alternatively, a statistical approach is to calculate the kappa statistic (Cohen, 1960), a measure of agreement that takes into account not only the observed agreements but also the expected agreements. Kappa values can be calculated using statistical software packages.

Another aspect of data reliability is the quality of the data entry into the statistical software. Data

need to be entered reliably. The problem of low data entry reliability can be minimized greatly by training those doing the data entry and by performing data entry verification. To estimate the data entry accuracy, evaluators may select a portion of the data to be double-entered by a second data entry person, and then keep track of the percentage of data entered for which there was a discrepancy. At least a 95% accuracy rate is needed. On one research project that involved a large survey, female inmates from the state prison entered survey data into a database. Being skeptical, the researchers performed data verification on 10% of the surveys. The data entry accuracy rate was found to be 99.8%—much better than the minimum 95% accuracy rate.

Validity of Measures

The *validity* of a measure is the degree to which the tool captures what it purports to measure—in other words, the extent to which the tool measures what it is intended to measure. A measure is valid if it truly measures the concept. Note that the validity of measures is different from the validity resulting from the study design, which is concerned with being able to say the results generally apply to the population (generalizability of findings).

Establishing the validity of measure through advanced statistical analyses is more typically a research activity rather than an evaluation activity. However, validity can be established more informally as face validity. Face validity, in the simplest procedure, involves asking a panel of experts whether the questions appear or seem to be measuring the concept. If the panel agrees, then face validity has been established. Because face validity can be relatively straightforward, stakeholders and program staff members can be involved and help establish this type of validity.

Time and resources used to establish face validity can be saved if existing instruments and questionnaires are selected. Several books are available that specifically review the validity and reliability of health-related instruments (Bowling, 2005) as well as websites (see Internet

Resources at the end of this chapter). Other sources for locating existing instruments include the published literature. Researchers who have developed instruments for specific constructs often publish information about those instruments in scientific journals. Another source is electronic databases of instruments, which are typically available through major universities. The Health and Psychological Instruments database, for example, contains published articles using health and psychological instruments, facilitating the search for an instrument applicable to a specific variable.

► Contextual Considerations in Planning the Evaluation

Several broad considerations can influence the effect evaluation. Each of these considerations comes into play to varying degrees during the evaluation.

Evaluation Standards

Drawing on standards developed by the Joint Committee on Standards for Educational Evaluation, Patton (2008) outlines four criteria for a good evaluation. The first criterion is that the evaluation meets the practical needs of its users. The second criterion is that the evaluation be feasible, both diplomatically and financially. The third criterion is that stakeholders can trust the evaluation as being scientifically and ethically conducted and can trust the evaluator as a person to do what is ethical and scientifically sound. The fourth criterion is that the evaluation is described with enough information for others to make their own determinations about its merit. Each evaluator has the responsibility of designing outcome evaluations that meet these criteria to the greatest extent possible.

Ethics

Another consideration when conducting an effect evaluation is the ethics inherent in conducting outcome assessments. Ethical considerations can affect not only the criteria for eligibility as a program participant but also eligibility for evaluation participants and procedures in the evaluation. From the perspective of some health agencies, program eligibility—not just program evaluation participation—of those in need may be both a moral and an ethical issue. In other words, the ethics of program eligibility needs to be addressed during the program development stage, and those decisions affect the subsequent ethics of who is included in the evaluation.

The federal Office for Human Research Protections' (OHRP) spotlight on health programs receiving federal funds has highlighted the need for evaluators to participate in efforts to comply with federal regulations. One area of great concern is obtaining informed consent, which is discussed elsewhere in the text. Consultants from a local university can be helpful in navigating this process. In larger healthcare systems, the internal research review panel can be consulted for assistance in this area.

These caveats and considerations can create circumstances in which an ideal outcome assessment is not possible, despite efforts to make the outcome assessment scientifically rigorous. Recognizing this constraint early in the process and working with it as a limitation is a strength. The solution then becomes to focus on creating not a “perfect” assessment but rather a “good enough” assessment; “good enough” outcome assessments use the best methodological design given the realities at hand.

One reason why a “good enough” outcome assessment is acceptable is that intervention effect evaluations differ from research in important ways (Table 11-2). The differences between research and evaluation can get blurry for some health program outcome assessments, but recognizing that they exist can help health program evaluators develop a strong, “good enough” intervention

evaluation. These differences may not be well understood by stakeholders, so education about how the evaluation is different from research can help stakeholders have greater trust and confidence in the evaluation. A “good enough” evaluation needs to have the same characteristics as the criteria discussed earlier for all evaluations—namely, it must use a feasible design that produces credible, believable, and potentially useful findings.

Stakeholders

One consideration is the stakeholders of the program, and the need to take into account who wants to know what from the outcome assessment. Once the key outcome evaluation questions have been posed, it is important to ask stakeholders what they want to know about the program. Involvement of the stakeholders in the development of the effect evaluation can produce some real payoffs. Not only will they become invested in the findings, but they will be more likely to believe the findings. Because neither programs nor stakeholders are static, some outcome objectives might have been developed by people who are no longer involved with the program and hence may not be valued as highly by current stakeholders or still be relevant given the current circumstances. If the ultimate goal of the outcome evaluation is its use in future program or policy decisions, then stakeholder involvement is imperative. Also, involvement of stakeholders can result in an improved outcome assessment. Another benefit of stakeholder involvement is that expectations can be addressed; that is, if stakeholders' expectations are unrealistic given resource limits or methodological reasons, those expectations need to be acknowledged, discussed, and made more realistic.

At the end of the evaluation, the intended users of the evaluation need to be able to judge the utility of the design and recognize both the strengths and the weaknesses of the evaluation. The intended users may have different criteria for judging the quality of the evaluation, so they may

need to learn about the methods used to carry out the evaluation. A debate about the possible findings before the evaluation is complete can help to uncover what would and would not be acceptable findings from the perspective of the stakeholders.

The creation of a long list of variables to be measured may be met with resistance and skepticism by stakeholders, particularly those who advocate for the protection of the participants. This reality reinforces the need for stakeholder involvement in the development of the effect theory and its evaluation. Nevertheless, in some circumstances it will be important to go beyond the immediate evaluation objectives in data collection. For example, the evaluation can offer an opportunity to collect information deemed

necessary for future program refinement or to update aspects of the community health assessment. The data collected for these purposes are best kept separate, with separate analyses and reporting of them.

For each element of an evaluation, both scientific and programmatic considerations apply (TABLE 11-6). These considerations have the potential to influence the ultimate design and implementation of the effect evaluations. Reviewing these differences and considerations with stakeholders can help establish realistic expectations and identify points on which consensus is needed. Whatever choices are being faced, ultimately the planners need to be flexible so that the evaluation can adapt to the realities that are encountered (Resnicow et al., 2001).

TABLE 11-6 Summary of Evaluation Elements

Elements of Effect Evaluation	Science Considerations	Program Considerations
What to evaluate	Outcome and impact variables most likely to demonstrate the strength of the evidence for the effect theory	Highest-priority outcome and impact objectives, variables that meet funding agency requirements
Who to evaluate	Representativeness of sample and comparability to nonparticipants, ethics of participant assignment to the program or not	Accessibility of program participants, availability of easily accessed target audience members
When to evaluate	Onset and duration of effect	Convenience and accessibility of program to participants
Why evaluate	Scientific contributions and generation of knowledge	Program promotion, refinement of program, funding agency requirements
How to evaluate	Maximize rigor through choice of measures and design and analysis	Minimize intrusion of evaluation into program through seamless integration of evaluation with program implementation

► Across the Pyramid

At the direct services level of the public health pyramid, the evaluation of the outcome of direct services is the most straightforward, albeit not necessarily the easiest, type of evaluation. It is straightforward in terms of the effect on individuals who receive the service. At this level, it is individuals who complete questionnaires and about whom data are collected via secondary sources. For this reason, the means used to construct the questionnaire and collect secondary data are major considerations.

At the enabling services level, the same issues arise as at the direct services level of the public health pyramid—namely, the need to construct sensitive and valid measures and the reliability of the means used to gather secondary data. In addition, evaluators face the issue of how to identify program participants, given that enabling services are likely to be embedded within other services and programs. At the enabling services level, the unit of observation is more likely to become a point of consideration.

At the population services level of the public health pyramid, the major issue deals with aggregation of data and selection of the unit of observation. The inclination is to say that the population is the unit, but in actuality, the unit is most often an aggregation of individual-level data. True population measures would include the gross national product (GNP), population density, biodiversity, and community social capital. Data collection cannot be collected from an entire population, except by census. It is possible to use data from existing national surveys to develop a rate, which is then assumed to apply throughout the population. However, this assumption may be false at local levels.

Often health programs at the population services level address public knowledge, public opinion, or community outcomes. For these outcomes, methods such as surveys, focus groups, and document reviews could be appropriate tools, depending on the specific evaluation question (Stead, Hastings, & Eadie, 2002), as well as social media.

At the infrastructure level, the evaluation itself is an infrastructure process. But if the program is intended to affect the infrastructure, outcomes are measured as related to the infrastructure. If the skills of public health employees (capacity building) are the program intervention focus, then individual-level data are needed to measure changes in their skills. If the intervention outcome of an infrastructure change is at another level of the pyramid—say, physical education to change a specific direct services practice—then the impact is measured as it relates to the other pyramid level. If the program is intended to have an effect at the infrastructure level, however, then the challenge is to construct measures that are at the infrastructure level. The National Association of County and City Health Officials (2016) profiles local health department infrastructure based on its national survey and has resources on its website for community health assessment–related infrastructure measurement. *Healthy People 2020* lists a range of potential ways to measure public health infrastructure, as does *Partners in Information Access for the Public Health Workforce* (2016). However, these global assessments of the public health infrastructure may not be useful unless the programmatic intervention has sweeping policy impacts that would be measurable throughout the infrastructure.

► Discussion Questions and Activities

1. Develop a working plan for how you would involve stakeholders in the development of an intervention evaluation plan. How does your plan address timing and budget constraints? In what ways does your plan include attention to interactions and group processes? Which strategies will be used to address the four criteria for having a good evaluation question?
2. Select one of the five health programs implemented in Bowe County and

Layetteville (i.e., vaccination, birth defects, pregnancy prevention, gun-related violence prevention, or Type 2 diabetes). Construct a levels of measurement table for at least one outcome (dependent) variable in three different health or well-being domains.

3. What are the major problems in using secondary data sources, such as birth or death certificate data? What, if anything, can the evaluator do about these problems?
4. How do the steps in instrument development apply to making surveys cross-culturally appropriate? What effects might you expect in the data, and subsequently in the interpretation of those data, if surveys are not culturally appropriate?

► Internet Resources

Centers for Disease Control and Prevention (CDC)

The CDC has a useful site called “Measurement Properties: Validity, Reliability, and Responsiveness” (www.cdc.gov/hrqol/measurement.htm) that lists studies and research related to the validity and reliability of health status and quality-of-life measures.

Research Methods Knowledge Base

William M. K. Trochim’s website has valuable information related to research and evaluation, including the levels of measurement (www.socialresearchmethods.net/kb/measlevl.php).

Research Methods

Webster University summarizes key concepts regarding research methods on validity, types of research studies, and populations and samples (<http://faculty.webster.edu/woolfm/statmethods.html>).

References

- Best, J. A., Brown, K. S., Cameron, R., Smith, E. A., & MacDonald, M. (1989). Conceptualizing outcomes for health promotion programs. In M. T. Braverman (Ed.), *Evaluating health promotion programs* (pp. 19–32). San Francisco, CA: Jossey-Bass. doi:10.1002/ev.1523
- Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement and statistics*. London, England: Sage.
- Bowling, A. (2005). *Measuring health: A review of quality of life measurement scales* (3rd ed.). Philadelphia, PA: Open University Press.
- Burke, V., Beilin, L. J., Cutt, H. E., Mansour, J., & Mori, T. A. (2007). Moderators and mediators of behaviour change in a lifestyle program for treated hypertensives: A randomized controlled trial (ADAPT). *Health Education Research*, 23, 202–217. doi:10.1093/her/cym047
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cohen, S., Underwood, L. G., & Gottlieb, B. H. (2000). *Social support measurement and intervention: A guide for health social scientists*. Oxford, UK: Oxford University Press.
- Donaldson, S. I. (2001). Mediator and moderator analysis in program development. In S. Sussman (Ed.), *Handbook of program development for health behavior and practice* (pp. 470–496). Newbury Park, CA: Sage.
- Dour, H. J., Wiley, J. F., Roy-Byrne, P., Stein, M. B., Sullivan, G., Sherbourne, C. D., Bystritsky, A., Rose, R. D., & Craske, M. G. (2014). Perceived social support mediates anxiety and depressive symptom changes following primary care intervention. *Depression and Anxiety*, 31(5), 436–442. doi:10.1002/da.22216
- Foshee, V. A., Bauman, K. E., Ennett, S. T., Suchindran, C., Benefield, T., & Linder, G. F. (2005). Assessing the effects of the dating violence prevention program “Safe Dates” using random coefficient regression modeling. *Prevention Science*, 6, 245–258.
- Healthy People 2010. US Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Accessed on 3.26.2017 at <https://www.healthypeople.gov/>.
- Maticka-Tyndale, E., Wildish, J., & Gichuru, M. (2007). Quasi-experimental evaluation of a national primary school HIV intervention in Kenya. *Evaluation and Program Planning*, 30, 172–186. doi: <http://dx.doi.org/10.1016/j.evalprogplan.2007.01.006>
- Partners in Information Access for the Public Health Workforce. (2016). Retrieved from https://phppartners.org/health_stats.html#County and Local Health Data
- Patton, M. (2008). *Utilization focused evaluation* (4th ed.). Newbury Park, CA: Sage.

- Resnicow, K., Braithwaite, R., Dilorio, C., Vaughan, R., Cohen, M. I., & Uhl, G. (2001). Preventing substance abuse in high risk youth: Evaluation challenges and solutions. *Journal of Primary Prevention, 21*, 399–415. doi:10.1023/A:1007029910822
- Rossi, P. H., Lipsey, M. W., & Howard, E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Stead, M., Hastings, G., & Eadie, D. (2002). The challenge of evaluating complex interventions: A framework for evaluating media advocacy. *Health Education Research, 17*, 351–364. doi:10.1093/her/17.3.351
- Varni, J. W., Limbers, C.A., & Burwinkle, T. M. (2007). How young can children reliably and validly self-report their health-related quality of life? An analysis of 8,591 children across age subgroups with the PedsQL 4.00 Generic Score Scales. *BioMed Central, 5*, 1. doi:10.1186/1477-7525-5-1
- Vogt, W. P., & Johnson, R. B. (2011). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences*. Thousand Oaks, CA: Sage.

CHAPTER 12

Choosing Designs for Effect Evaluations

L. Michele Issel, PhD, RN, and Arden Handler, DrPH

Of the many factors contributing to the scientific rigor of an effect evaluation, design is key. In a health program evaluation, evaluators, compared to researchers, may not have the same degree of control over timing, quality and amount of data collection, or amount of exposure to the intervention. Such constraints and realities contribute to the differences between evaluation and pure research. Much of the evaluation literature discusses design from a research perspective and is geared toward maximizing the scientific rigor of the evaluation. While this approach is needed and appropriate if sufficient funding is available, most small health programs do not have such resources. Most agencies are more focused on and have expertise related to providing or delivering health programs. Personnel in such programs can understandably become confused or overwhelmed when faced with the task of designing an effect evaluation. This chapter discusses designs from the perspective of evaluators and program managers who must address issues posed by program realities.

Several assumptions underlie the discussion in this chapter. The first assumption is that, regardless of the program or agency, at least one design exists that is both scientifically valid and realistically feasible for the program. A second assumption is that program personnel would choose the best such option if it was clear and easy to identify. Thus, designs are described here with respect to their feasibility for both small, local, direct services programs for individuals as well as for large-scale, full-coverage programs for populations. Most designs can be adapted for either individual-level or population-level interventions if the same decision criteria are met. For simplicity, the explanation about a design is framed around the usual and customary use of that design.

A third key assumption is that there is a programmatic need to demonstrate that the recipients or population have changed more than might happen by chance. This assumption means that only collecting data on whether the outcome objectives were met is not sufficient

to meet the needs of stakeholders, including funders. Choosing the right design becomes critical in showing that changes did not occur by chance or that one intervention is more effective than another. The fourth assumption is that the intervention was delivered as planned to the intended audience. This assumption is critical in terms of then assuming that changes should be related to the intervention. Finally, for some evaluations, a qualitative or mixed methods design yields more accurate and complete information. However, the focus of this chapter is on quantitative designs.

► Evaluation Design Caveats

The *design* of the evaluation is the grand scheme that delineates when and from whom data are collected. Many types of designs are drawn from health and social sciences. The *methods* indicate the way in which the data are collected

as part of the evaluation and typically consist of strategies such as surveys, interviews, or observations.

Evaluation designs can be very simple or extraordinarily complex. Most research textbooks cover designs and methods in considerable depth and from a perspective in which the researcher has a great deal of control over factors affecting the choice of a design. However, the choice of a design for an effect evaluation is complicated by the realities and limitations that are inherent in health programs.

Effect evaluations can be conducted from the perspective of a number of different disciplines, each of which has its own terminology for describing designs (TABLE 12-1). The social science of psychology typically focuses on individuals and uses the terms *experimental* and *quasi-experimental* to describe the major classification of designs. Health education mostly uses social sciences terminology for designs. Epidemiology and sociology are among the other disciplines from which to draw on when developing an evaluation plan, particularly for

TABLE 12-1 Contribution of Various Disciplines to Health Program Evaluation

Discipline	Typical Outcome or Impact Question Terminology	Typical Design
Psychology	Are outcomes for individuals in the program (recipients) different from those in the comparison or control group (nonrecipients)?	Experimental, quasi-experimental
Epidemiology	Are cases (individuals who have the outcome characteristic) less likely to have had exposure to the program than controls (individuals without the outcome characteristic)?	Observational
Health services research	Does differential utilization of services (participation or not) by enrollees (intended audience) and non-enrollees (nonintended audience) lead to differential outcomes?	Experimental, quasi-experimental, clinical trial
Sociology	How do the values and power structure of the policy context affect program options?	Observational

full-coverage or population-based programs with intended audiences that are populations. Another discipline that is relevant to the evaluation of health programs is health services research, which uses terminology drawn heavily from economics and medicine. Designs from all these disciplines are available for use in program evaluation.

The terminology used throughout this chapter represents a blend of terms used in social sciences and epidemiology. Application of epidemiological designs to the evaluation of health programs has increased over the last two decades (Handler, 2002; Rosenberg & Handler, 1998). However, confusion often exists about when designs from epidemiology, originally developed for observational purposes, should and can be used in program evaluation. As such, particular attention is given to understanding and applying epidemiology designs. In addition, this chapter introduces the terms *pretest* and *posttest*, which have been adopted from social sciences. When we apply these terms to health programs, two points are worth stressing.

First, *pre* and *post* refer to the sequence of when data are collected; *pre* is any time before receiving the program intervention, and *post* is any time after receiving the program intervention. The precise timing of data collection depends, of course, on logistical issues. More important, however, it depends on the expected timing of the first program effects and the duration of those effects as articulated in the program effect theory. The timing of the pretest and posttest data collection also should have been specified in the “by when” element of the outcome objectives.

Second, *test* is a convenient, shorthand term that refers to the measures being used to quantify the program effect. In other words, it is the measurement of the specific outcomes and impacts. When data are collected at the population or aggregate level rather than at the individual level, the pretest equivalent is *baseline data*; the latter term is more commonly used in epidemiology, public policy, and health services studies of populations.

Considerations in Choosing a Design

The aim in choosing a design is to come as close as possible to a design that can demonstrate an effect actually caused by the intervention, program, or policy. An ideal design has three salient characteristics: (1) a comparison control, or unexposed group that is typically as similar as possible to the experimental or exposed group—however, in some population-level designs, the comparison group of choice may be a dissimilar group not expected to experience a change in the outcomes; (2) measurement of the outcome variables occurs before and after the intervention for unbounded health outcomes, that is, those outcomes that can occur before and after program exposure; and (3) as described below, minimal threats to internal and external validity exist. Designs that meet these three criteria allow the evaluator to make statements that attribute differences found between the two groups to the intervention—and only to the intervention. As discussed below, some designs incorporate more than one control or comparison group. The following eight considerations become factors in the choice of a design.

Causality

The first decision in choosing a design is about whether it is important to determine if a cause-and-effect relationship (also known as a causal relationship) exists between receiving the health program or policy interventions and the health outcomes. There is a direct relationship between the ability to show causality and the costs and complexity of the design (Bamberger, Rugh, & Mabry, 2011). Specifically, as the ability to show causality increases, so do the costs and complexity of the design (**FIGURE 12-1**). Thus, these two factors become the first considerations in choosing a design.

While seeking to identify a causal relationship with increased certainty leads to a more costly design, such a design may be appropriate if the intervention is new or being adapted to a new

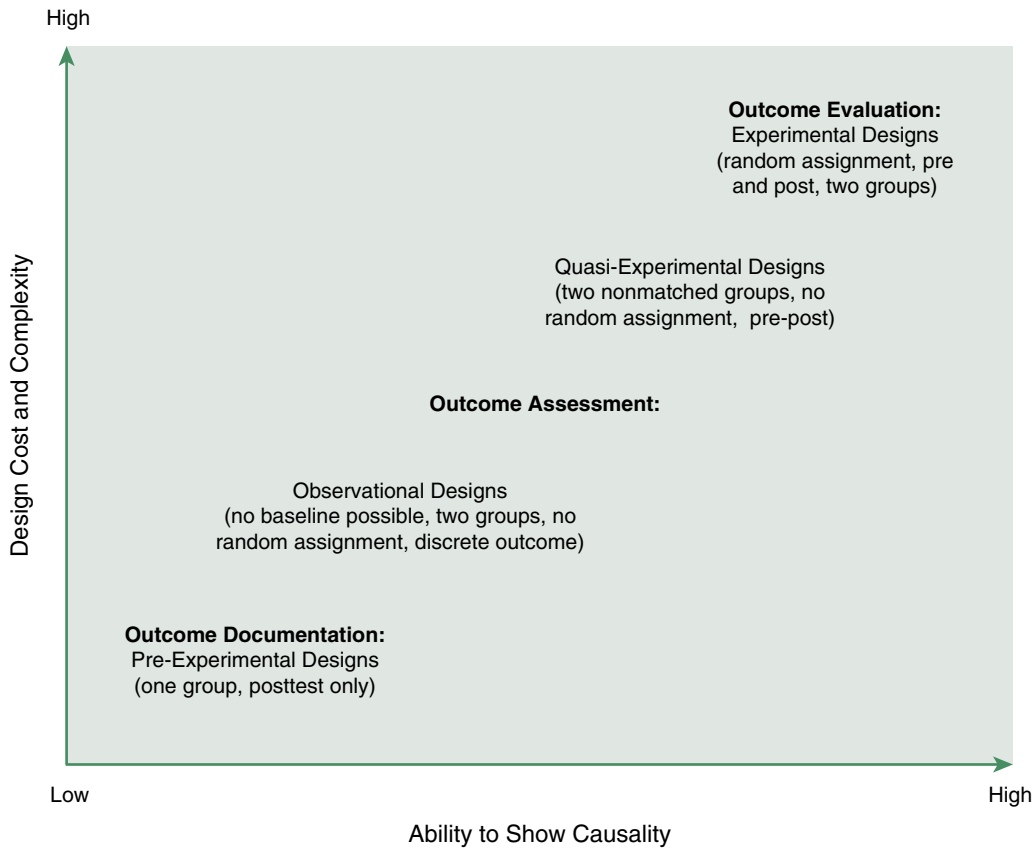


FIGURE 12-1 Relationship Between the Ability to Show Causality and the Costs and Complexity of the Design

intended audience. Some designs are better suited than others to establishing a causal connection. This first decision is tempered by considering the need to balance the design with the skill and resources available to the health program personnel. A creative evaluator or evaluation team may devise a complex, scientifically rigorous design that may provide sound evidence for the effectiveness of the program. Nevertheless, unless resources for personnel, data collection expenses, consultants, and incentives for participants are adequate and available, the design will probably not be implemented as planned. As such, the necessity of adapting the ideal to reality when conducting effect evaluations is ever-present.

Bias

The choice of an evaluation design is also influenced by the need to have a design that is as free of bias as possible given the realities of the evaluation. *Bias* in design refers to the extent to which the design is flawed and therefore more likely to lead to an inaccurate conclusion about the effectiveness of the health program. The flaws are categorized based on whether they affect the ability to generalize the findings to other populations (*external validity*) or whether they affect the ability to say that the intervention made a difference (*internal validity*; Campbell & Stanley, 1963; Shadish & Cook 2002). Understanding the design flaws

that constitute threats to the internal and external validity of each design helps in choosing a design that is as free of bias as possible given the evaluation question and the circumstances. In general, designs with fewer flaws are more complex and costly, and are usually more likely to demonstrate a causal relationship between the health program and outcomes.

Retrospective Versus Prospective Orientation

Another major consideration in the choice of a design is whether data are collected retrospectively or prospectively. *Retrospective designs* entail gathering data from the point of intervention backward in time. In other words, participants will have received or been exposed to a program or policy before any pretest or baseline data are gathered or accessed. *Prospective designs* entail gathering data forward in time, beginning from a point prior to the initiation of the intervention. If the evaluation is planned during the development of the health program, a prospective design is possible and is generally the preferred approach. Retrospective designs typically limit the options for measurement and for the selection of comparison groups, whereas prospective designs require that evaluation resources be obtained prior to the start of the program.

Time Span

The length of time during which the evaluation is conducted—whether looking back or looking forward—needs to be considered. A design can be described as *longitudinal* if it includes multiple pretests, multiple posttests, or both, with these tests spaced out over time. The word *longitudinal* has the connotation of longer time spans, such as weeks, months, years, or decades, over which the pretest and the posttest data are collected, although in reality data might be collected multiple times during one day, a few days, or a week. As the designs are described, it will be easy to imagine which are amenable to modification to become longitudinal. Three

factors that determine whether a longitudinal design is appropriate in a particular situation are, first, the intervention theory; second, how many times data can be collected, over what time period; and third, the evaluation budget. For some evaluations that use or need data from a large number of individuals, longitudinal designs may be feasible. For example, data from social media, health insurance claims, or medical records with repeated measures for uniquely identified individuals make it possible to conduct longitudinal analyses. Designs at the population level might also use large data sets, albeit in aggregate form, such as injury rates for all children ages 6 to 19 calculated as a rate for a decade.

Groups

Yet another consideration in choosing an evaluation design is whether it is possible to identify and distinguish between participants and nonparticipants or between those exposed and not exposed. There is inevitably an interplay between the design choice and the ability to know who did and did not receive the intervention. Choosing an evaluation design is linked to defining who is a participant versus a nonparticipant or who is exposed as a recipient versus not exposed; this applies to both individual-level and population-level programs. The process of defining participants versus nonparticipants can take place prospectively or retrospectively, but it relies on information developed in the service utilization plan for program eligibility, screening, and recruitment. Careful screening and participation recordkeeping as the program proceeds become useful in distinguishing program recipients from nonrecipients. The group whose members received or participated in the program is typically called the *experimental*, *intervention*, or *exposed group*, whereas the group whose members did not receive or participate in the program is called the *control*, *comparison*, or *unexposed group*. The choice of terms is usually based on the evaluation design, the scope of the effect evaluation (individuals or a population), and the discipline of the evaluator. Related to the

ability to identify the groups is knowing whether it is possible and feasible to collect data from those who did not receive the interventions, which also influences the design choice.

Bounded and Unbounded Outcomes

Another consideration in choosing an evaluation design is the type of outcome being evaluated. In the context of choosing a design, two broad classifications of health outcomes are distinguished: unbounded and bounded.

One class of outcomes has the possibility of existing before and after the program; these outcomes are therefore *unbounded* by time or an event. Most health conditions and behaviors fall into this group. Examples of unbounded outcomes include knowledge; attitudes; behavioral intentions; behaviors; and health or physical conditions that can exist before (and after) the program, such as blood pressure, immune status, weight, or depression. Because the health outcome or behavior is unbounded by an event, it may be possible to collect data on the health outcome before the program begins.

The other classification of outcomes is *bounded*, meaning that these outcomes can occur only once (no theoretical possibility of a pretest) for any given individual at a particular time and therefore are bound to an event and time. In epidemiology, these results are called discrete health outcomes. Birth and death are the prime examples of bounded health outcomes, as are low birthweight, pregnancy, or amputation of toes for diabetics. Some health behaviors or experiences also occur only once—for example, initiating breastfeeding after delivery, self-efficacy of self-administered insulin after being started on insulin, or an experience of sexual abuse. For bounded outcomes, the program participants or recipients cannot provide pretest information on these outcome variables because the outcomes occur only after the programmatic intervention has been implemented. For these types of outcomes, there is no naturally occurring pretest or baseline value for individual

program participants. However, when data are considered at the population level, there actually can be a “pretest” or baseline value for these bounded outcomes. For example, the incidence of low birthweight births or rate of adolescent pregnancy from 2010 to 2016 would be the pretest data if the intervention began in 2017. Finally, the variable used to measure bounded outcomes may sometimes be measurable only as a dichotomous variable, which also affects the design choice.

Individual- Versus Population-Level Interventions

Health programs designed for different levels of the public health pyramid require different designs. Some designs are used when the intervention is delivered to individuals, whereas other designs are applicable when the intervention is delivered to the population and when the outcome is also measurable at the population level. For population-level interventions, a design is required that captures the changes in the population rather than just in a group of affected individuals. Often the population-focused intervention takes time to manifest and show an effect, which then requires considering the time span of the evaluation. Some programs focus on changing infrastructure, such as the development of a regional referral network; changes in reimbursement or coverage; or a change in licensure requirements may require the use of designs for populations because the recipients are populations rather than individuals.

Intervention and Observational Designs

Recall that program evaluation has roots in educational evaluation. From that starting point came the classic book on evaluation designs, *Experimental and Quasi-Experimental Designs for Research*. In this work, Campbell and Stanley (1963) identified and described 16 different intervention designs: 3 nonexperimental, 10 quasi-experimental, and only 3 truly experimental.

Designs that use random assignment of potential participants to either receive or not receive the program are called *experimental*, while designs that do not use random assignment but are more robust than nonexperimental designs are called *quasi-experimental* designs.

The assumption in the traditional literature (Shadish & Cook, 2002) is that exposure to the programmatic intervention is manipulated by the health professional, as typically occurs with health programs delivered to individuals and groups. *Manipulation* implies that those providing the program have some degree of control over the choice of who among those eligible receives the program. Designs that rely on such manipulation of the intervention, broadly called *intervention designs*, have different names depending on the discipline. For example, in epidemiology and clinical medicine, intervention designs are typically called clinical trials, while in the social sciences, they are called experimental designs or quasi-experimental designs.

In program evaluation, experimental and quasi-experimental designs at the individual level are typically used when four conditions exist. First, it is theoretically possible to collect information or data on the state of the outcome variable prior to intervention. Second, outcome information is usually collected from or on members of at least two groups. Third, an intervention takes place and is received by members of one of the groups. Fourth, after delivery of the intervention, data are collected from or on the same members of the groups as were collected pre-program. Evaluation questions that ask whether the intervention changed the unbounded outcome at the individual level are best answered through experimental and quasi-experimental designs (Handler, 2002; Rosenberg & Handler, 1998). This classification of designs is widely used and understood, but it does not include all possible designs. What is lacking in this classification scheme are the observational designs that are primarily from epidemiology but are now often used for program evaluation.

Observational designs, which come from the field of analytic epidemiology (Rothman,

2012), were initially developed for situations in which exposures are not manipulated but are simply observed as they naturally occur. These designs are typically used to study what constitutes environmental or lifestyle risks. Historically, observational designs such as case-control and cohort designs were reserved exclusively for examining health risk factors and outcomes. However, as epidemiologists have become more thoroughly integrated into health services research and become actively involved in conducting program evaluations, use of observational designs to examine the relationship between receiving health program interventions and health outcomes has grown. This trend toward a more interdisciplinary and integrated approach in program evaluation is consistent with the continuing evolution of designs and the science behind designs for program evaluation (Barnett & Wallis, 2005; Zuckerman, Lee, Wutoh, Xue, & Stuart, 2006).

When epidemiologic designs are used for program evaluation, they are no longer purely observational. The fact that a program was created and delivered and that some individuals or populations were exposed and others were not amounts to a *de facto* manipulation of the exposure, even if only at the policy level, such as eligibility for, or exposure to, a public program as a percentage of the federal poverty level. Based on this logic, manipulation of the exposure to the program is no longer a useful way to distinguish between observational designs and intervention experimental and quasi-experimental designs when conducting program evaluations (Handler, 2002; Rosenberg & Handler, 1998).

An observational design is best used for program evaluation at the individual level when the health outcome is bounded—that is, discrete (e.g., birth or death)—because there is essentially no theoretical possibility of collecting pretest data on the outcome for the individual (Handler, 2002; Rosenberg & Handler, 1998). For example, baseline data are generally not available for gunshot wounds at the individual level. Hence, when we examine the effect of a violence prevention program, we are essentially

employing a cohort study (see below) in which, for instance, youth who participated in conflict avoidance training are considered the exposed group and those who did not receive this training are considered the unexposed group. Although, in this circumstance, individual-level pretest data are not available (gunshot wounds before program participation), baseline data for bounded outcomes on the population level usually are available (e.g., community rates of gun violence). When this is true, designs for the population level become possible, such as time-series designs (described below) and thus typically fall into the quasi-experimental rather than epidemiologic typology.

Using Designs Derived from Multiple Paradigms: An Example

The following example demonstrates how both experimental/quasi-experimental and observational evaluation designs might be used to measure different components of a health program or intervention. High-risk women attending a county birth defects prevention program receive education about nutrition and environmental hazards that contribute to neural tube defects. The program consists of preconception risk screening, nutrition counseling, distribution of multivitamins prior to and during pregnancy, and cooking classes. Each component is evaluated for its effect on food choices, increased dietary folic acid intake, and the presence of neural tube defects in the infants.

For the nutrition education and cooking class components, the outcome of interest is increased dietary folic acid intake. At the beginning of the first nutrition education session, women provide information on their current food-purchasing and dietary practices. The same women also provide data on their food-purchasing and dietary behaviors at the end of the nutrition education session and when they have completed the cooking classes. The possibility that the same women provide both pre- and post-program data on their behaviors permits the program evaluator

to choose from among a set of experimental or quasi-experimental designs.

The situation is different for evaluating the effect of the program on the occurrence of a neural tube defect. Whether or not women have had prior children, it is not possible to collect pretest data about the prevalence of neural tube defects in the current pregnancy because women in the program have not yet given birth. Consequently, the only evaluation design choices to evaluate this program effect are observational designs. For this intervention, an observational design such as a two-group prospective cohort study might be used because neural tube defect data can be collected after delivery from the birth records of women in the program (exposed) and from the birth records of a group of women not in the program (unexposed). At first pass, the collection of data from two groups after the intervention may seem like a posttest-only nonrandomized design, which is considered a weak, nonexperimental design. However, when examining neural tube defects, a posttest-only design in the form of a cohort study, for example, is a robust design because it is not possible to have pretest data at the individual level on the health outcome of interest (i.e., neural tube defects). Observational designs are used in evaluation for outcomes, such as the occurrence of disease (e.g., asthma, injury) or finite events (birth or death), for which it is not theoretically possible to collect pretest data at the individual level; this contributes to these designs increasingly being used in evaluations of public health programs.

► Choosing the Evaluation Design

Identifying Design Options

From the point of view of health program evaluation, designs can be understood more easily if we think of them as having three levels of ability to attribute effects to the program, yielding three groups of designs: outcome documentation, outcome assessment, and outcome evaluation.

These three levels correspond more readily to the trade-offs and choices that program administrators and evaluators must make in choosing a design. Overall, it is more important for the effects of the theory of intervention mechanisms to be assessed than the theory of outcome to impact because, unless the causal factors, as identified in the theory of intervention mechanisms, are changed by the intervention and the health outcomes are achieved, the longer-range health impacts will not occur. This chapter reviews only the most common designs, noting the types of bias that are likely to occur with each.

Ideally, feasible design options can be identified before the program begins to ensure that the strongest, most rigorous design possible can be implemented for the effect evaluation. Most of the stronger program evaluation designs must be planned and launched *before* recipients receive the program, which means that the evaluation must be ready for implementation well before program implementation. The discussion of designs generally comes after the discussion of program implementation and monitoring. This is somewhat misleading and should *not* be interpreted as indicating when planning for the evaluation should occur.

An underlying assumption is that program managers and evaluators have some means of knowing that an individual or individuals received the program intervention. For population-level programs, such as a seat-belt law or a mass-media campaign, exposure to the program can be assumed for all members of the population, although it is sometimes possible to obtain precise measures of coverage. For programs at the direct services or enabling services level of the public health pyramid, indicators of having received the program can be obtained more directly.

Overview of the Decision Tree

To assist with the choice of an evaluation design, **FIGURE 12-2** presents a flow diagram of the key questions that need to be asked. Each of the designs discussed in this chapter (except the ecological design) is shown in the decision tree. At each branch, a yes or no response to the key question

leads either to a design or to a subsequent choice. At each choice branch, the program evaluator and the program staff members need to confer and agree on which branch to follow.

The designs shown in the decision tree in Figure 12-2 are the ones most likely to be used in the majority of program evaluations. More complex designs do exist, such as the Solomon four-group design. Such complex designs are more likely to be used in research rather than in an applied setting, so they are not discussed or covered in this text. Also, the use of more than one comparison or control group or the collection of post-program data more than once (except for time series) is not reflected in the list of designs; these procedures are considered modifications to the basic designs that do not intrinsically alter their structure or rigor.

As the decision tree in Figure 12-2 shows, the questions of whether a comparison group is possible and whether random assignment is possible are also critical to determining the design. If the answer to both questions is yes, the resulting possible designs tend to be stronger. It is important to note that the first decision question is whether it is possible, either theoretically or practically, to collect any posttest or outcome data from program participants. These data might be obtainable either directly from individual participants or indirectly from secondary sources, such as medical records or vital records. If the answer is no, then it is not possible to determine whether the program had any effect, and an evaluation is not possible.

The second branch of the decision tree focuses on whether the program, and hence the evaluation, is focused on the entire population or aimed at specific individuals within the population. For population-based evaluations, it is highly desirable to have baseline data on members of the population to determine if an intervention effect occurred. If it is not possible to have such data, a strong evaluation design usually cannot be constructed. The discussion of population-level designs is interwoven into the discussion of designs for effect evaluations. In Figure 12-2, many of the designs can be used for either individual-level or

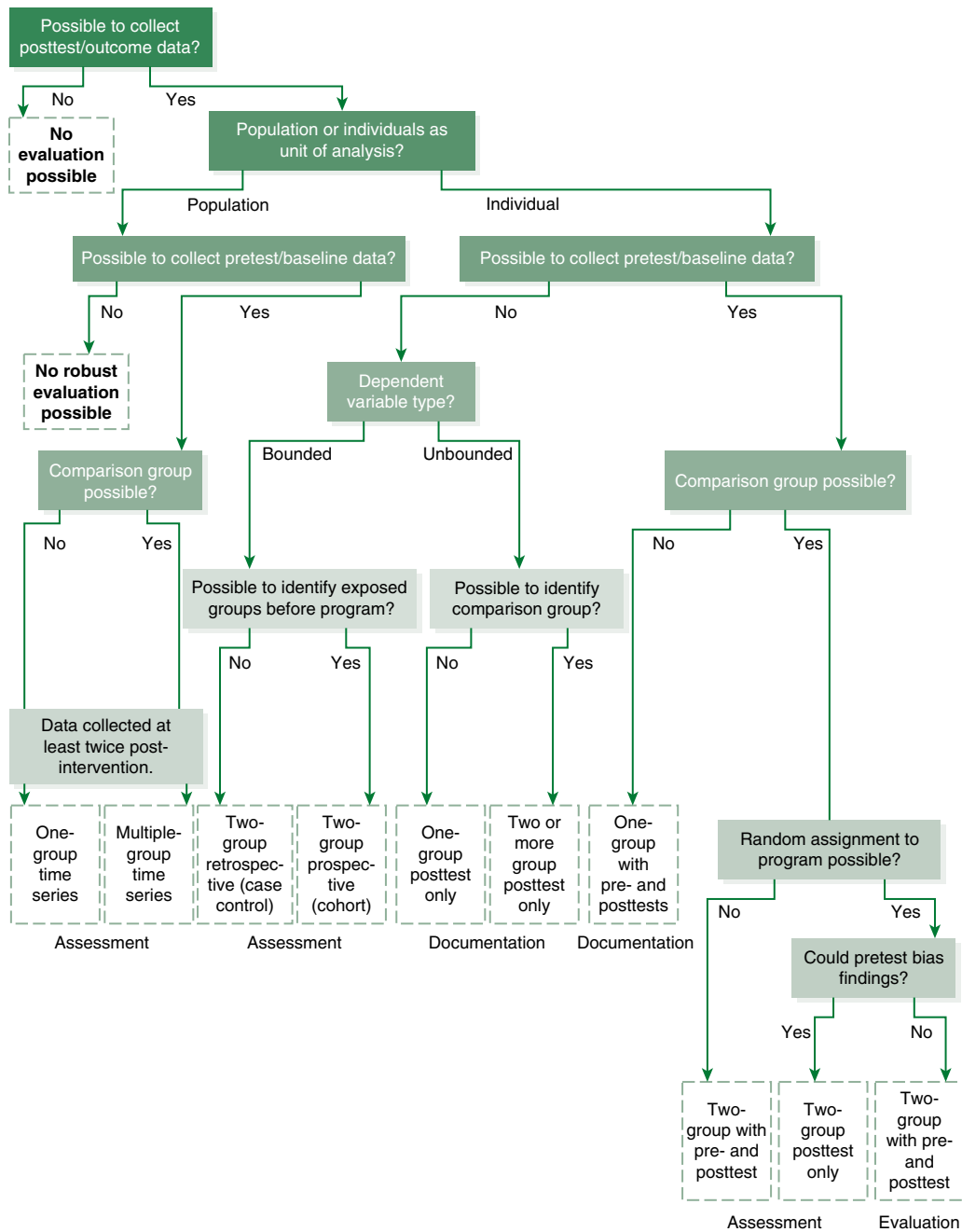


FIGURE 12-2 Decision Tree for Choosing an Evaluation Design, Based on the Design's Typical Use

population-level programs as long as all other prior decision points are the same.

In the following overview of designs, it is clear that most designs can be implemented at either the individual or population level (TABLE 12-2). Using the design at the individual or population level does not change the assumptions that must be met, the timing of data collection, or the

comparability of the groups. A few designs can be used at only one level—a restriction that is a result of the nature of the design.

In the following discussion, if a design is known by more than one name, all alternative names are given. This is the case for designs that are structurally the same but have roots in both epidemiology and the social sciences.

TABLE 12-2 Summary of Main Designs and Their Use for Individual or Population-Level Evaluations

Design	Is Design Possible at the Individual or the Population Level?	
	Individual	Aggregate, Population
Designs drawn from Experimental (with random assignment), Quasi-experimental/Nonexperimental Typology		
One-group posttest only	Yes*	Yes *
One-group pretest/posttest (can also utilize multiple pretests/posttests)	Yes*	Yes
Comparison groups posttest only (with or without random assignment)	Yes*†	Yes*†
Comparison groups pretest/posttest (with or without random assignment)‡	Yes†	Yes†
Epidemiologic Designs		
Two-group, retrospective; case control	Yes	No
Two-group, prospective; cohort	Yes	No
Ecological study (epidemiologic design used for aggregate data)	No	Yes

*Not recommended (although sometimes can be strengthened).

†Stronger design if random assignment is used.

‡Cluster trial if there is random assignment of groups rather than individuals.

Designs for Outcome Documentation

At a minimum, an evaluation should document the effect of the program in terms of reaching the stated outcome and impact objectives. Focusing on documenting an outcome from the point of view of the objectives often results in using designs that are called nonexperimental. Nonexperimental designs are those at the individual level that involve collecting data after program implementation (i.e., post-intervention) compared either to the objective targets for the stated outcome or to pretest data. These designs are frequently used by health program administrators because they are not complicated and are comparatively inexpensive. Their simplicity also contributes to their wide intuitive appeal and sometimes to their unintentional use. It is important to understand that nonexperimental designs in particular are quite weak in terms of being able to attribute changes or differences to the program. Because of their weaknesses, these designs are appropriately used to document the level or degree of reaching the target stated in the outcome objective. Program administrators and evaluators must remember, however, that these designs do not allow for the attribution of any causality to the program. Put simply, these designs can answer the evaluation question of whether any noticeable change or difference occurred, but they cannot identify the source of the change as the intervention.

In the social sciences, posttest-only designs, and particularly those without random assignment, are considered weak nonexperimental designs. They are considered weak because no baseline data exist by which to judge if a change in an outcome has occurred. However, when using epidemiologic designs for bounded outcomes at the individual level (e.g., cohort and case-control studies), concern about the weakness of the posttest only is unwarranted because these designs do not measure a change in a health status outcome, only whether or not a health status outcome occurs in the exposed and unexposed groups (e.g., number of low

birthweight infants in women exposed to prenatal case management versus those not exposed to prenatal case management).

One-Group Posttest Only

An evaluation that collects data on the outcome variable solely from program participants and only after the program takes place is employing a one-group posttest-only design. A posttest-only design involves providing the intervention and then collecting data only for those who received the intervention, which makes it useful primarily for individual-level evaluations. This design is inexpensive and simple to understand, so it is often the decision of choice for smaller programs with limited budgets and for staff members with minimal training in program evaluation. It may be the only option if the program was initiated before the evaluation could be started. It would also be the only design option if evaluators do not have access to members of the intended audience who did not receive the program.

Findings based on the data from this nonexperimental design can be misleading because the design has two major biases, or threats to the internal validity of the evaluation: history and maturation. *History threats* are specific events that happen to participants between the beginning and the end of the program—these events might also explain the findings. *Maturation threats* arise when participants mature physically or emotionally between the beginning and the end of the program, independent of the program. In the one-group posttest-only design, there is no information about participants except at one point in time; thus, evaluators have no way of knowing whether external historical events or internal maturational processes affected the outcomes. As a consequence, this design has minimal usefulness in determining the effect of the program. While not a recommended design, one way to strengthen this design slightly at the individual level is to include multiple or repeated posttest measures. Similarly, at the population level, collecting multiple aggregate measures after the introduction of an intervention or policy

(e.g., a decline in diabetes rates in a community for each of five years after the construction of walkways and parks) might allow an evaluator to claim that there is a suggestion of an effect.

One-Group Pretest/Posttest

An evaluation that collects only outcome data for participants before receiving the program and again after receiving the program is employing a one-group pretest/posttest design. This non-experimental design involves collecting data from program participants any time before and any time after they have received the program. It can be implemented readily by many direct services programs that collect data on a set of measures before starting the program and again at exit from the program.

The one-group pretest/posttest design can also be used with a group of individuals from whom data are collected multiple times over a long-term follow-up. For example, all patients with diabetes at one clinic might be followed for several years, with data being collected on their HbA1c values and weight. During that time, if a weight-loss program intervention is introduced, this design would enable evaluators to identify whether a change occurred among those patients exposed to the program. Generally, if a group of individuals is followed over time and outcome data are repeatedly collected on them, the study design is called a *repeated measures* design.

A major advantage of the one-group pretest/posttest design is that data can be analyzed for indications of the amount of change in program participants. Often the data are collected in a manner that allows for linking the pre- and posttest data for a single individual. For this reason, the simple pretest/posttest design is used mostly when data are collected at the individual level.

The disadvantage of the one-group pretest/posttest design, as with the posttest-only design, is that both history and maturation can affect the data. In addition, three other threats or problems may arise with this design. A *testing effect* occurs when the process of being involved in providing the pretest data in some way affects

the posttest data. For example, a questionnaire about exercise given at baseline may motivate program participants to exercise more so that, on the posttest data, they may report more exercise because of their exposure to the questionnaire rather than because of their exposure to the program or policy change. Another possible threat is *instrumentation*, in which the concern focuses on possible changes or alterations to how or which data are collected for the posttest compared to the pretest. Unless exactly the same information is collected in exactly the same way, the findings can be influenced by the data collection method rather than by the program. Finally, as with the one-group posttest-only design, there is no comparison group, so it is unclear if change would have occurred regardless of the intervention. All of these issues make this a weak design for effect/impact evaluation but adequate for program documentation.

A simple pre- and posttest one-group comparison can be made with population data as well. For example, injury rates before and after the introduction of a bicycle helmet law is a simplistic version of a one-group time-series design. The difference between the before- and after-program scores or levels for the outcome variable is usually easy to calculate and to understand. In addition, because there is no expense involved in finding and gathering data from nonparticipants, this design has a relatively low cost. This design is strengthened by multiple measures before and after the intervention (see “One-Group Time Series” below).

Comparison Group Posttest Only

An evaluation that collects data on or from participants as they complete the program as well as from a group of nonparticipants from the target audience is said to use a posttest-only design with a comparison group. In this third nonexperimental design for program documentation, data are collected once from both program participants and another group that did not receive the program. For instance, Moazami et al. (2014) used a posttest-only design to compare knowledge between

two groups of dental students. Compared to the students assigned to the traditional curriculum, those assigned to a tailored virtual curriculum had higher knowledge scores. However, it is not possible to know if there are initial differences between these two groups. On the other hand, if random assignment is used in a posttest-only comparison group design, this design is stronger and is considered experimental rather than nonexperimental.

The simplicity of the posttest design with a comparison group is that evaluation costs are low, especially if the nonparticipants are recruited from a readily available source, such as a companion program or clinic. Like the one-group posttest-only design, a posttest-only design with a comparison group may be the only option available if the evaluation was started after program implementation or if resources are scarce. Usually, the difference between the scores or levels for the outcome variable for the two groups is easy to calculate. Unfortunately, the difference may be incorrectly interpreted as an actual program effect, whereas it may only suggest a program effect or may actually reflect bias in the design. Although this evaluation design reduces the likelihood of maturation, instrumentation, and testing threats, it is vulnerable to other threats—namely, program attrition and selection bias. *Attrition* is the loss of participants over time due to their dropping out of the program and/or evaluation, moving away, being lost to follow-up, or death. Attrition increases as the time span for data collection increases. *Selection bias* refers to the fact that program participants may differ significantly from those not in the program. Unless more extensive data are collected from everyone in the evaluation, the extent to which the two groups differ in fundamental ways will never be known. A difference in outcome between the groups, if seen, may be attributable to factors other than the intervention. For example, those participating in the intervention group may be either those least at risk, leading to an overestimate of effect, or they may be those most at risk, leading to an underestimate of effect. These kinds

of bias problems are often found in studies of services for mothers with low incomes, such as the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) (Lee, Rozier, Norton, & Vann, 2005).

For more about how selection bias might affect evaluation results, let's suppose that WIC has initiated a prenatal education campaign about the importance of breastfeeding. The program evaluator plans to compare breastfeeding knowledge and attitudes of postpartum women in WIC to those of postpartum women not in WIC. However, the group of women in WIC might differ in fundamental ways from the group of women not in WIC, including factors that influenced their decision to be WIC recipients. Perhaps any difference in the postpartum breastfeeding knowledge and attitudes of the two groups of women reflects those factors that influenced their decision to participate in WIC rather than the prenatal breastfeeding program.

Another example of a comparison group, posttest-only evaluation is informative. Klassen and associates (2002) matched women who attended a mammography program with friends and neighbors. They called the friends and neighbors “controls,” connoting that they were not program participants. Data collection occurred only once, after the program was complete and from both groups, making it a comparison group, posttest-only design. However, the researchers incorrectly called this another type of design, a case-control design (described later in “Two-Group Retrospective [Case-Control]”). This example highlights the potential confusion over correctly describing and labeling designs, and hence the potential challenges that program evaluators face when discussing designs with relevant stakeholders. Klassen and colleagues (2002) also claimed to have matched the program participants with their friends and neighbors. True one-to-one matching is done infrequently because of the cost involved in identifying and recruiting individuals who are completely like the program participants. Nonetheless, these authors' use of a convenient and accessible comparison group highlights the potential for

thinking creatively about how to arrive at the best possible design and sample.

Designs for Outcome Assessment: Establishing Association

If more resources are available for understanding the degree of effect from the program, then more complex and costly designs can be used. In general, the quasi-experimental designs fall into this category, along with some observational designs. The designs for outcome assessment answer the evaluation question of whether any noticeable change or difference seems related or associated with having received the program. All of these designs are considered stronger than the outcome documentation designs because, to varying degrees, they attempt to minimize possible differences between groups that may falsely explain program effects or noneffects.

These designs involve at least post-program/intervention data collection, and typically use of a comparison group. When comparison groups are utilized in quasi-experimental designs, these comparison groups are considered nonequivalent, meaning that the groups being compared are not necessarily statistically similar or matched on a variety of potential confounders that can affect the outcome. Nonequivalence occurs because those receiving the program or intervention are in one way or another a “select” group in terms of having factors that influence being in the program or not. Individuals who choose to receive or have access to the program may be different from those individuals who choose not to receive/do not have access to the program, and without extensive data on both groups, it may not be possible to know how the two groups differ initially. Therefore, claiming that the program was the cause of differences in the outcome variable between the groups remains problematic. However, these designs are stronger than the designs for program documentation because typically there are statistical methods to adjust for differences between groups. In addition, at the population level, it is possible to have a strong quasi-experimental design without a comparison group (see “One-Group Time Series” below).

Two individual-level observational designs from epidemiology can also be used for outcome assessment. Generally, these observational designs—cohort and case control—are appropriate for bounded outcomes at the individual level. Also, one additional observational design, the ecological design, is useful for establishing population-level associations between exposure to an intervention or policy change and population-level measures of the outcome. (See Ecological Design section).

In the following subsections, designs for outcome assessment are described in terms of their use in program evaluation and their commonly used names. For each design, the biases inherent in that design are discussed. Again, the trade-offs involved in choosing one design over another need to be weighed in view of the specific program characteristics and the evaluation needs of the program staff members, stakeholders, and funders.

Two-Group Pretest/Posttest

An evaluation that collects outcome data on or from program participants and nonparticipants, both before the program begins and after the program is complete, uses a two-group pretest/posttest design. This design is more formally called a nonequivalent, two-group, comparison group design. It extends the two-group posttest-only design discussed earlier by adding the collection of data from both program participants and nonparticipants before the program/intervention is introduced (pretest). In other words, pretest and posttest data are collected from both program participants and nonparticipants. This design has intuitive and practical appeal because the statistical comparison of the two groups is relatively simple and straightforward.

In using this design, evaluators must attempt to ensure that the groups are as alike as possible by carefully selecting the nonparticipants (i.e., the comparison group). This design, when used with random assignment, is considered the clearest of evaluation designs (Shadish, Cook, & Campbell, 2002) and will be discussed in detail

below. Without random assignment, however, evaluators do not have any assurance that the groups will be alike—hence the “nonequivalence” label. Selection bias is a major threat in the two-group, pretest/posttest design. If sufficient information about members of both groups has been collected, such as demographic variables and relevant antecedent variables, the statistical analysis can be adjusted to account for the differences between the groups. Another threat to this design is regression to the mean, which refers to the tendency for the scores of different groups to become more alike over time. Thus, the longer the time period between the intervention and the collection of the posttest data, the more likely the two groups are to have no differences on the outcome variable. This risk suggests that the design is best used with outcomes that are expected to occur relatively soon after the intervention, and the posttest data should be collected at that time point. The two-group pretest/posttest design at the individual level is theoretically similar to the population-level multiple time series (discussed below) but typically includes only two data points (pre and post) and is focused at the individual level.

One-Group Time Series

As shown in the decision tree in Figure 12-2, the one-group time-series design is one of two designs that can be used if an entire population is the unit of analysis for the program evaluation. Such a design makes sense if aggregate data have been collected or are available at several time points before the program and at several time points after the program (e.g., injury rates before and after a car-seat bill is passed in the state legislature). This design is also known as a single time-series design because a series of observations is measured on only one population (e.g., injury rates in one state); it is also referred to as an interrupted time-series design because the program interrupts the baseline value of the trend in the outcome variable before the intervention is introduced.

The one-group time-series design might be one of the few options available if the program is delivered to or received by an entire population—as is the case with interventions such as seat-belt laws, water fluoridation, or health insurance for underinsured children provided through state health insurance programs. Because the unit of analysis is at the population rather than individual level, there is no assumption that exactly the same individuals contributed to the data at each time point. A time series, because it does not require following specific individuals, is often used in public health for evaluating the impact of policy change on a population. It is also useful in evaluating programs that are delivered to only one distinct aggregate, such as a school district that will be included in the program and evaluation, and for which the same data have consistently been collected over time (e.g., standardized test scores).

The key consideration in choosing this design is the number of time points before and after the program at which data were collected and are available. For a variety of statistical reasons, the optimal number of time points of observation is five before and five after the intervention (Tukey, 1977). For a program targeting a large population, such as a seat-belt law, obtaining multiple years of highway fatality data before and after the law took effect may not be a problem. In other programs, such as a school-based program, where the data collection tools used with students, such as standardized tests, may be revised every so often, having the same metric for multiple time points, such as every year for 5 years, before and after a program may be problematic. A reasonable rule of thumb, when the ideal is not possible, is to have a minimum of three reasonably spaced time points of data before and after the program was initiated. Fewer than three time points worth of data makes it extremely difficult to assess statistically for a change in the outcome from before to after the program/intervention.

The one-group time-series design has intuitive appeal, is easy to plan, and has a

relatively low cost if the data already exist. An additional attractive feature of this design is that data other than physical health outcomes can be used. Conrey, Grongillo, Dollahite, and Griffin (2003) used the rate of redemption of WIC coupons for produce at farmers' markets in New York as the outcome variable for a program designed to increase the use of farmers' markets by women in WIC. Data on WIC recipients were collected over time, such as on a weekly basis. Their study is also an example of how existing services/process data might also be appropriate outcomes for population-focused health programs.

Campbell and Russo (1999) have pointed out that, with the one-group time-series design, usually only effects from sudden and dramatically effective interventions can be identified and distinguished from the normal background variations that occur over time. This shortcoming occurs because the design has several biases, of which history is the major problem. The other threats—maturation, instrumentation, and selection—tend to be less problematic and have less effect on the conclusions. One major disadvantage of the one-group time-series design is the challenge in interpreting the amount of change in relationship to when the intervention occurred. It is possible that the change did not occur immediately after the program was introduced. The different patterns of the intervention–change relationship may become evident only during the data analysis.

Multiple-Group Time Series

If data have been collected from a potentially affected population and a comparison population at several time points before the program or the intervention, and the same data have been collected at several time points after the program or intervention, a multiple-group time-series design is possible. The addition of at least one comparison population that did not receive the intervention is the most obvious way to improve on the one-group time-series design, as long as

the data are from the same time frames for all groups. The classic example of this design is the comparison of two states, one with a new law (e.g., seat belt, bicycle helmet) and the other without such a law. The annual mortality rates for motor vehicle/bicycle accidents in both states are plotted across several years before and after enactment of the law. The rates in the two states are compared to determine whether the motor vehicle- or bicycle-associated mortality rate in the state with the law declined more after the enactment compared to the mortality rate in the state without the law.

Because the multiple-group time-series design requires collecting data for at least two groups multiple times before and after the program, it is generally used with large aggregates, such as schools, or with populations. Nevertheless, the same logic applies when this design is used to evaluate programs for individuals. Using the earlier example of patients with diabetes who visit a clinic, it is easy to imagine a health system that includes several clinics, but where two clinics implement the intervention. In this case, there are naturally occurring groups as well as repeated measures on the patients, which creates the two-group (or more) repeated measures design. When used at the individual level, this design is an extension of a two-group pretest/posttest design without random assignment and is quite a robust design.

The major advantage of the multiple-group time-series design is that few biases are likely to affect seriously the ability to draw conclusions regarding the program effect. In other words, the multiple-group time series is a very strong design to use with population-focused health programs. A major disadvantage of the design is that the same outcome variable data must have been collected on all populations being compared; otherwise, the data will not be comparable across the groups. Another disadvantage of time-series designs (both single and multiple group) is the need for more complex statistical analysis that takes into account the repeated measurement; beyond eyeballing changes in trends over time,

conducting this analysis typically requires the services of a statistical expert.

Two-Group Retrospective (Case Control)

If it is possible to identify individuals with and without a particular outcome and to review their historical data to determine which individuals received a particular program and which did not, a retrospectively constructed, two-group design is possible. This observational design (derived from epidemiology and also called a case-control design) is used at the individual level. It is appropriate when, for whatever reason, the program evaluation is conducted after the program has started or has concluded, access to the individuals for data collection purposes is limited, and the outcome variable is bounded. In this design, those with the outcome are compared to those without the outcome, with regard to whether they were exposed to or received the program.

The retrospective two-group design is useful because the degree to which the program improved the likelihood of the desired outcome can be calculated statistically by comparing those with the outcome to those without the outcome based on whether they were exposed to the program. The exposed/not exposed and the outcome/not outcome relationships are often represented with a two-row by two-column table, called a 2×2 table. In the retrospective two-group design, knowing who falls into which of the four cells in the table becomes possible only after exposure to the program and the occurrence of the outcome; relevant data may be collected through a review of existing records that contain both program participation data and outcome variable data. Thus, this design is feasible only if evaluators have access to those records and if the information in those records includes data on both exposure to the program and the outcome variable.

A retrospective case-control design is generally used when the outcome is bounded. In other words, this design can be used when it

is not theoretically possible to collect baseline or pretest data on the outcome for the same individuals, as discussed earlier. For outcome variables that occur only once, the retrospective two-group (case control) design is quite robust, meaning that it is a strong design (Handler, 2002; Rosenberg & Handler, 1998). The uniqueness of this design derives from the fact that individuals who have and do not have the outcome of concern are identified post hoc and then compared with regard to whether they received the program. In this sense, this design retrospectively assigns individuals to exposure groups.

The major advantage of the retrospective two-group design is that it can be used any time after the program has been implemented (with bounded outcomes only). However, use of this design is not possible unless either the evaluator or program itself has collected data on program exposure, even if the information was not initially intended for evaluation purposes. Like other evaluation designs that are implemented after the program, the retrospective two-group design may be one of the few choices available if the evaluators were not involved in creating or choosing an evaluation design earlier in the planning cycle.

This design does have certain disadvantages. A major limitation relates to the ability to obtain high-quality data on exposure to the program, including whether the individual received the program and what intervention dosage was received. Although healthcare organizations have been increasing their use of data warehousing, there is no guarantee that the variables needed for the evaluation will be available or that they will have been collected in a consistent and reliable manner. The cost associated with the retrieval of existing data—a task that is necessary in all retrospective designs—will be lower if a comprehensive and stable management information system has been used for recording the data that need to be abstracted for the program evaluation. However, if budget constraints and data retrieval difficulties impose severe limitations on the number of individuals for whom adequate retrospective data are available, then the size

of the groups with and without the outcome of interest who were exposed or non exposed can be a concern. Insufficient sample size can affect the statistical conclusions that can be drawn about the effectiveness of the program, particularly if effects are present but weak.

Another potential problem is that, although the retrospective case-control design is robust, selection bias can be present. Data about why individuals were in the program will likely not be available, making it difficult to know whether any apparent program effect might have been due to a selective preference on the part of participants for accessing/receiving the program. However, observational epidemiologic designs adjust for potential differences between groups that might lead to false conclusions or confound the relationship between exposure to the intervention and a bounded health status outcome (disease yes/no). Confounding occurs when an association between two measures is estimated inaccurately because that association really reflects one or more additional measures that are not included in the analysis. Another major issue with the case-control design is recall bias—that is, inaccurate memory, particularly if the participants are asked about their having received the program. Recall bias can be circumvented if evaluators have access to records documenting participation in the program.

Two-Group Prospective (Cohort)

If the target audience is distinct and clearly defined, and can be followed forward in time as a group, it is called a *cohort*. If a cohort exists, then a prospective design becomes possible. Like the retrospective two-group design, the prospective design is used if there is no theoretical possibility of collecting pretest outcome data on the same individuals because the outcome is bounded (e.g., the occurrence of disease, disability, or death). Prospective cohort designs are widely used in the evaluation of health services when the outcomes of interest are health outcomes that occur only once for an individual. These characteristics render this design most suitable for individual-level evaluation.

Two versions of the prospective cohort design are distinguished; they differ based on whether it is possible to know at the outset of the evaluation who will receive the program. In version I, it is not possible to know beforehand who will participate in or receive the program. The members of the intended population are followed forward in time (prospectively) as a group (a cohort) for a given time period, and some individuals are exposed to the program and some are not. The evaluators determine who received exposure to the program at the end of the evaluation. In version II, it is known before the program begins who will and will not participate in or receive the program. Both groups are followed forward as a cohort throughout the duration of the evaluation. In both versions of the prospective cohort design, at the end of the time period, the outcome is measured to determine its occurrence or nonoccurrence among those exposed to the program intervention.

When evaluating the effect of a home visitation program for pregnant women on the prevention of child abuse, for example, it is possible to use a two-group, prospective cohort design. One group might consist of pregnant women in the program who are followed for a year postpartum; at the end of the year, the evaluators assess the incidence of child abuse. The other group might include pregnant women, perhaps from a waiting list, who are also followed forward in time, with child abuse measured at one year. At one year, the rate of child abuse for each group of new mothers would be compared. This example represents version II of the prospective cohort study design.

This design is appealing because data can be collected on key variables, such as relevant antecedent and moderating factors, from members of the target audience before exposure to the program. In this regard, this design does not need to rely on existing data, as is true with the retrospective two-group design. Also, if version I is used, this design does not require knowing ahead of time who will receive the program. The exposed and unexposed groups are identified at the end of the time period, based on data

collected over the duration of the evaluation. If data collection occurs over a long time period, more individuals might develop the outcome expected to result from the program intervention. This extended follow-up period is an important feature for programs with a potentially long time lag between intervention and evidence of outcome, such as those aimed at changing substance use, long-term weight loss, or repeat pregnancies. The major disadvantages of a prospective design are the need to track individuals for a substantial time period and the need to collect follow-up data on the outcome variable. Maintaining contact with the individuals in the evaluation for a period of time can be costly, as well as frustrating, particularly if those persons do not have stable lives or their whereabouts are not reliable. Loss of participants due to attrition does have consequences for the validity of the design and the statistical conclusions that can be drawn. Although attrition of those in the evaluation can be addressed—and lowered—through a variety of strategies (Dhaliwal et al., 2014), these are often both costly and labor intensive.

Often it is not possible to determine who will be exposed to the program before the program begins. When program records are used to establish exposure, but the analysis proceeds from the exposure forward in time (the prospective element) to the outcome, then the design is best called a retrospective cohort design. An example of this design is an evaluation of five types of providers of prenatal care services: public hospital clinics, health department clinics, community clinics, private physicians' offices, and private hospital clinics (Simpson, Korenbrot, & Greene, 1997). In this study, all women on Medicaid from specific geographic regions were included in the evaluation, which sought to determine the association between prenatal care provider type and pregnancy outcome. The medical records of all women in the study regions who gave birth during the evaluation time period were reviewed. The evaluators noted where the women had received prenatal care and then compared the outcomes of the pregnancies across the types of providers. Despite the fact that the groups were

retrospectively constructed, the design used in this study was prospective because it followed the women forward through their pregnancy and reassessed the bounded outcome variables of birthweight and preterm delivery.

Ecological Design

In some instances, the evaluation seeks to understand if a relationship exists between an intervention and an outcome, but only at the population level. Evaluation of interventions designed and implemented to have effects on aggregates or populations could use several of the designs discussed in this chapter. One of the designs that is explicitly focused at the population level is the ecological study design drawn from epidemiology. The ecological study design considers the differences between groups rather than individuals (Morgenstern, 1995) and is used most often to determine whether there is an association between an exposure and a health consequence at the population level. For example, ecological studies can be used for evaluating the effects of policy, such as policies to reduce injuries through mandating the use of seat belts or requiring cars to have airbags (Ferdinand et al., 2015). In the simplest form of ecological design, the exposed and unexposed are entities with rates above and below a certain threshold, such as counties with prenatal smoking rates above and below a target (e.g., a *Healthy People* indicator for prenatal smoking). Those with and without the outcome might also be aggregates with rates above and below the threshold for a key indicator, such as counties with low birthweight rates above and below a *Healthy People* target. What distinguishes the ecological design is that there is no possibility of identifying individuals, so the comparison occurs only across aggregates.

A key advantage of the ecological design is its use of existing population data, often drawn from various sources. At the same time, this design can be subject to the *ecologic fallacy*—that is, the evaluators may wrongly assume that group characteristics apply to all individuals in

the group. Nonetheless, ecological studies can be helpful in assessing whether there is a relationship between a public health intervention and a health outcome at the population level. In keeping with the aforementioned example, an ecological design would answer the question of whether there is a relationship between the percentage of counties in a state with smoking cessation programs and the percentage of counties that meet the *Healthy People 2020* objective for low birthweight.

Designs for Outcome Evaluation: Establishing Causation

The last set of designs is the most costly and complex, but it significantly increases the ability of evaluators to show that the program was truly the cause of the effect. Designs for outcome evaluation answer the question of whether the change or difference can actually be attributed to having received the program or intervention. Experimental designs involve exposure to a program, an outcome measure, and a comparison group from which change due to the program or intervention can be inferred. The distinguishing feature of all experimental designs is the use of random assignment. Only one of the several experimental designs is discussed here—the two-group pretest/posttest design with random assignment. The other experimental designs are essentially variations on this design, but because they are more complex, they are used less frequently.

Despite their analytic strengths, experimental designs are often not feasible for various reasons, including the numbers of participants needed to reach statistically sound conclusions and the lack of evaluator control over the programs in question. In addition, practitioners are sometimes reluctant to participate in studies with random assignment because they are hesitant to deny participation in the program to any potential participants. These challenges are discussed in more depth below. Experimental designs can be expensive in terms of the number of evaluation participants needed to reach statistically sound

conclusions and in terms of the amount of time required to track those in the evaluation sample.

Random Assignment

Random assignment is the process of determining on a random basis who does and does not receive a health program/intervention. This concept is not to be confused with *random selection*, which refers to the random identification from the intended population of those who will be in the program and/or evaluation. Random assignment is carried out by using either a table of random numbers or the flip of a coin, so to speak, although more sophisticated methods of arriving at random assignment also exist. Random assignment is the basis for designs used in clinical trials (epidemiological or medical terminology) and in experimental designs (social sciences language).

The advantage of using random assignment of participants to either the experimental/intervention group or the control/nonintervention group is that it creates comparison groups that are theoretically equivalent. In other words, by assigning evaluation participants to either the experimental group or the control group on a random basis, the possibility of the two groups being different is no greater than one might expect by chance alone. This consideration is critical because it eliminates design biases that stem from groups not being equivalent. With random assignment, the two groups are as alike as is theoretically possible, so subsequent differences found in outcome variables can logically be attributed to the effect of the intervention.

Practical Issues with Experimental Designs

A reality for many health programs is that ethical concerns often preclude the use of random assignment to receive or not receive the program. For each program, staff members and stakeholders should discuss their comfort with random assignment. In the birth defects prevention program described earlier, it might be

ethically acceptable to assign women randomly to the cooking program component, but it would be ethically unacceptable to assign women randomly to the prenatal vitamins component. In highly vulnerable populations or communities, experimental designs may not be ethically realistic or even acceptable to stakeholders. Stakeholders may object to studies using these designs because not all who need the program are able to receive it due to the random assignment. For example, a community agency may have a moral objection to not offering a health promotion program to all who are at high risk. However, if a new program is compared to the standard program, then health program evaluators might choose an experimental design that allows the unexposed group members either to receive the standard program or to receive the new program after the exposed group receives it. Similarly, if the program has limited slots, participants can be randomly assigned to participate instead of being assigned to a waiting list. These approaches to random assignment help minimize ethical concerns yet maintain rigor in the evaluation.

Other practical issues exist with experimental designs. First, the intended population needs to be sufficiently large to have both an experimental group and a control group. Second, the program interventions must be robust, meaning that the interventions must have a statistical probability of having an effect. Experimental designs are *not* appropriate if any element of the effect theory is poorly substantiated or if the intervention theory predicts only a small change in the outcome variable.

An evaluation of an educational intervention focused on reducing risks for cardiovascular disease in older African Americans (Walker, 2000) may have failed to find any program effect as a result of both of these challenges to an experimental design. In this case, the intervention for the experimental group consisted of spiritual and hypertension-related messages delivered by programmed telephone calls. The control group received only the spiritual messages via the programmed telephone calls. Both groups also received pamphlets and home visits by a health educator. Thus, the only difference between the

experimental and control group interventions was the hypertension-related messages provided to the experimental group. It is likely that receiving such messages would produce only a small change in hypertension management behavior. In addition, there were only 43 people in the experimental group and 40 people in the control group. This would be considered an overall inadequate sample size, especially for attempting to find the small effect from the hypertension-related messages provided to the experimental group. The strong experimental study design utilized here has been undermined by having a weak intervention and insufficient sample size. In other words, the choice of design must be considered in light of the program delivered and the overall resources for the evaluation.

A third issue that arises with the use of experimental designs is that, logistically, the evaluation team and budget must be able to accommodate the large number of evaluation participants needed for random assignment to work as intended. When random assignment is undertaken with too few participants, the two assigned groups may not be theoretically equivalent. For these reasons, random assignment, as an approach to the construction of comparison groups, is rarely, if ever, used in health programs that are not also research projects.

Two-Group Pretest/Posttest, with Random Assignment (Randomized Clinical Trial)

A two-group pretest/posttest design with random assignment must meet these conditions: (a) Individuals from the intended audience are randomly assigned to receive or not to receive the health program/intervention and, (b) data are collected from those who receive the program (experimental or intervention group) and from those who do not (controls or comparison group) both before and after receiving the program. This design is very similar to the two-group pretest/posttest design described earlier, except that random assignment is used. This design, which

is also called a randomized clinical trial, is often considered the gold standard in clinical research. Because it is the most rigorous of all possible designs, innumerable examples of this design exist, especially in the evaluation of medical treatments and pharmaceuticals. Nevertheless, the same design can also be used with nonclinical interventions. For example, West et al. (2014) randomly assigned physicians to a 9-month mindfulness program and then examined outcomes for both the intervention and control groups 3 and 12 months after the program ended.

The two-group pretest/posttest with random assignment design is generally perceived as a design at the individual level. In reality, the same design can be used with aggregates, such as schools, or with populations, such as geographically distinct communities. When applied to groups rather than individuals, the design might be called a *cluster trial*; when entire communities are randomly assigned, it is termed a *community trial*. Both the evaluation question and the unit of analysis are different when groups or populations are randomly assigned to receive the intervention; the evaluation issue to be determined becomes which program or program component works better (St. Pierre & Rossi, 2006).

Biases associated with this design are minimal because of the use or creation of equivalent groups. The major bias is the differential attrition that results if underlying but unknown causes prompt members of either the experimental group or the control group to drop out of the evaluation. This factor then results in a systematic difference between the groups, which affects the ability to attribute causality to the program/intervention. Another problem that may emerge is if the comparison group is inadvertently exposed to the intervention. To avoid problems associated with differential attrition and contamination, evaluators need to follow both groups carefully and determine who is and who is not still in each group at the end of the intervention. To compensate for differential loss between groups, analysis usually proceeds according to “intention to treat.”

► Designs and Failures

Three types of failure can prevent a program evaluation from demonstrating an effect. (FIGURE 12-3). TABLE 12-3 summarizes the strategies for avoiding each type of failure.

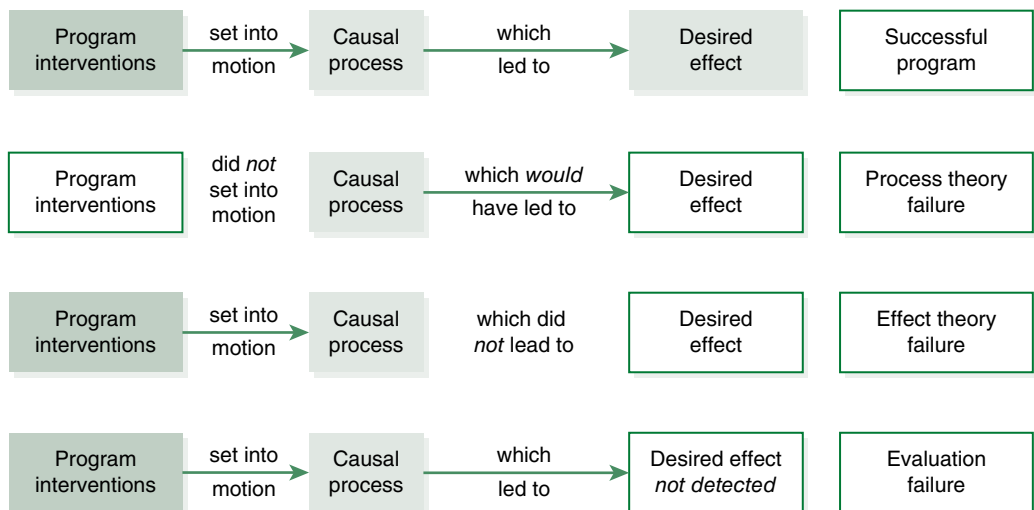


FIGURE 12-3 Three Sources of Program Failure

Modified from Weiss, C. H. Evaluation, Second Edition, © 1998. Pearson Education, Inc., Upper Saddle River, NJ.

TABLE 12-3 Approaches to Minimizing Each of the Three Types of Program Failure

	Process Theory Failure	Intervention Mechanism Theory Failure	Evaluation Design Failure
Definition	The interventions were not sufficiently implemented to (potentially) affect the health problem being targeted	The interventions did not or could not affect the health problem being targeted	The evaluation method, design, or sample was inappropriate, such that the true effects of the program were not detected
Design and methods considerations	Develop logic model and ensure that the intervention supports implementation of all elements	Consider the evidence base for all pathways in the intervention theory; consider timing with regard to finding maximum effect from the program	Tailor design approach and instruments to match the evaluation question and specific impact expected; maximize internal and external validity
Sample considerations	Ideally, include all eligible program participants	Ideally, use random assignment and random selection; if not able to randomize, ensure sufficient numbers to enable adjustment for confounders	Select equivalent intervention and control groups or use random assignment; use an adequate sample size to achieve high power

The first type of failure is process failure, in which the service utilization plan is not sufficiently implemented, thereby resulting in no or inadequate implementation of the intervention. Avoiding process theory failure involves careful program oversight and implementation. The process or monitoring evaluation, rather than the effect evaluation, provides insights into whether a process theory failure occurred.

The second type of failure concerns the effect theory. If the program interventions did not or could not lead to the desired program impacts and outcomes, then the rationale for choosing those interventions may have been flawed. If a high-quality evaluation found no effect and the process evaluation verified that

the intervention was provided as planned, then an effect theory failure may have occurred. However, both the process evaluation and the outcome evaluation must have been conducted and must have provided sufficient scientific data to reach the conclusion that effect theory failure occurred. In particular, the outcome evaluation must have been designed to assess for causality; this would have involved the use of experimental designs. Unless these approaches were used, it is impossible to conclude that only an effect theory failure took place.

The third type of failure is directly related to the evaluation—namely, that it failed to identify correctly an effect that resulted from the interventions or identified an effect that was

actually not attributable to the intervention. If the evaluation was flawed with regard to the sample, the measures used, the design, or the statistical tests employed, then the findings regarding the success of the program are questionable. To prevent an evaluation failure, it is necessary to adhere to the highest level of rigor possible given the programmatic realities (Table 12-3). There are multiple approaches to reducing bias in designs. Some approaches focus on using reliable and valid measures; other approaches focus on increasing the comparability of groups by using techniques such as controlling for individual participant differences through statistical means, whether multiple regression or propensity score matching, which uses all relevant measures to assign an individual's propensity to belong in a given group, and so can be used to control for the potentially confounding effects of those differences (Garrido et al., 2014). If adequate resources are available and it is critical to minimize the risk of evaluation failure, then certain steps can be taken to avoid this outcome. However, these steps are often costly and involve more stringent research methodologies. For example, if measurement is seen as a possible source of an evaluation failure, then additional, complementary, or redundant measures may be used to validate the measures used in the evaluation. To avoid evaluation failure due to inappropriate design choice, evaluators can use a decision flowchart like the one shown in Figure 12-2.

► Across the Pyramid

At the direct services level of the public health pyramid, health programs focus on individuals. Many of the designs discussed in this chapter can be used with programs at this level. It is a fairly straightforward matter to assign individuals randomly to an intervention or nonintervention group. As one might imagine, this approach may present ethical difficulties. Rather than randomly assigning individuals to either the program or control group, participants can become their own controls, as in a pretest and

posttest design, or future participants may serve as controls while they remain on a waiting list. Overall, effect evaluations at the direct services level can resemble or be straightforward research in terms of the sampling technique and design used. In terms of cost, evaluation designs at the individual level can vary greatly in expense depending on the type of data collected and the frequency of data collection.

At the enabling services level, health programs focus on groups of individuals and are provided in a wider range of contexts. Programs at this level are the most challenging to evaluate in terms of their outcomes, for several reasons. One challenge is to identify and recruit a comparison group that has the same characteristics as the program participants. Another challenge relates to the fact that health services programs at this level are probably not suited to experimental designs, although they may be appropriate for quasi-experimental designs. At a minimum, pretest and posttest designs can be used for evaluation purposes. In addition, depending on the program, random assignment might be possible, particularly at the group or community level.

At the population-based level, health programs are provided to entire populations. Although this broad scope of delivery does not preclude the use of experimental designs, it does limit the evaluation options to designs that can reasonably be implemented with populations. Time-series designs are especially useful for evaluating population-level programs. Evaluations that use existing data on populations will be the least costly to conduct, whereas community intervention trials with random assignment will be the most expensive of all designs.

At the infrastructure level, interventions focus on changing the operation of the health-care organization or the public health system. At this level, the outcome evaluation question will likely determine whether the evaluation examines actual changes in the infrastructure or changes to the health status of clients. This distinction in turn influences the type of design that is needed to answer the intervention effect

question. A repeated measures/time-series design would be appropriate if the focus is on a long-term change for which data are available at the individual or population level across many time points, as might be the case for a study of funding changes in economic hard times for health departments. In contrast, if the program focuses on the knowledge of individual workers, many of the designs that use pretest and posttest data and comparison groups at the individual level might be appropriate.

► Discussion Questions

1. Recall the concepts of individual- versus population-level evaluations. Which evaluation designs are more appropriate for population-level evaluations and which are more appropriate for individual-level evaluations? Justify your response.
2. Suppose the county board of supervisors has asked you, as an evaluation consultant, to evaluate the ongoing immunization program for older persons and the yet-to-be-implemented community-based adolescent violence prevention program. Both programs are offered throughout the county. You have been asked to suggest at least three design options for evaluating each program, based on complexity and cost. Which three designs would you propose for each program and why?
3. Which evaluation designs are susceptible to biases due to history and maturation? Explain.
4. Under what conditions would it be accurate to state that a program caused the improvements in a specific health outcome?

► Internet Resources

Numerous resources for study designs are available, but only a few are listed here to show the variety. As you explore websites with content

on research designs, you will notice differences in nomenclature, based on the discipline of the authors.

Garson

A nice list of designs, along with definitions, is available from Garson at: <http://faculty.chass.ncsu.edu/garson/PA765/design.htm>.

Introduction to Research Design and Statistics

This website, from a course developed by Philip Ender at the University of California at Los Angeles, lists the designs using the classic X O R descriptions: www.philender.com/courses/intro/designs.html.

Design Comparisons

The Center for Evidence-Based Medicine summarizes the advantages and disadvantages of the major design types: www.cebm.net/?o=1039. Other resources include the summary at Georgia State University (<http://research.library.gsu.edu/c.php?g=115595&p=755213>) or this one from University of Minnesota (<https://hsl.lib.umn.edu/biomed/help/understanding-research-study-designs>).

References

- Bamberger, M., Rugh, J., & Mabry, L. (2011). *Real world evaluation: Working under budget, time, data and political constraints* (2nd ed.). Thousand Oaks, CA: Sage.
- Barnett, J. J., & Wallis, A. B. (2005). The missing treatment design element: Continuity of treatment when multiple post observations are used in time-series and repeated measures study designs. *American Journal of Evaluation*, 26, 106–123. doi:10.1177/1098214004273253
- Campbell, D. T., & Russo, M. J. (Eds.). (1999). *Social experimentation*. Thousand Oaks, CA: Sage.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.
- Conrey, E. J., Grongillo, E. A., Dollahite, J. S., & Griffin, M. R. (2003). Integrated program enhancements increased utilization of Farmer's Market Nutrition Program. *Journal of Nutrition*, 133, 1841–1844.
- Dhaliwal, J., Nosworthy, N. M., Holt, N. L., Zwaigenbaum, L., Avis, J. L., Rasquinha, A., & Ball, G. D. (2014). Attrition and the management of pediatric obesity: An integrative review. *Childhood Obesity*, 10(6), 461–473.

- Ferdinand, A. O., Menachemi, N., Blackburn, J. L., Sen, B., Nelson, L., & Morrissey, M. (2015). The impact of texting bans on motor vehicle crash-related hospitalizations. *American Journal of Public Health, 105*(5), 859–865. doi:10.2105/AJPH.2014.302537
- Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. *Health Services Research, 49*(5), 1701–1720. doi:10.1111/1475-6773.12182
- Handler, A. (2002). *Lecture notes*. Retrieved from www.uic.edu/sph/mch/evaluation/index.htm.
- Klassen, A. C., Smith, A. L., Meissner, H. I., Zobora, J., Curbow, B., & Mandelbatt, J. (2002). If we gave away mammograms, who would get them? A neighborhood evaluation of a no-cost breast cancer screening program. *Preventive Medicine, 34*(1), 13–21. doi:10.1006/pmed.2001.0956
- Lee, J. Y., Rozier, R. G., Norton, E., & Vann, W. (2005). Addressing selection bias in dental health services research. *Journal of Dental Research, 84*(10), 942–946. doi:10.1177/154405910508401013
- Morgenstern, H. (1995). Ecologic studies in epidemiology: Concepts, principles and methods. *Annual Review of Public Health, 16*, 61–81. doi:10.1146/annurev.pu.16.050195.000425
- Moazami, F., Bahrapour, E., Azar, M. R., Jahedi, F., & Moattari, M. (2014). Comparing two methods of education (virtual versus traditional) on learning of Iranian dental students: a post-test only design study. *BMC Medical Education, 14*(1), 45. doi:10.1186/1472-6920-14-45.
- Rosenberg, D., & Handler, A. (1998). Analytic epidemiology and multivariate methods. In A. Handler, D. Rosenberg, C. Monahan, & J. Kennelly (Eds.), *Analytic methods* (pp. 77–136). Washington, DC: Maternal and Child Health Bureau, Health Resources and Services Administration, Department of Health and Human Services.
- Rothman, K. J. (2012). *Epidemiology: An introduction*. Oxford, UK: Oxford University Press.
- Shadish, W. R., & Cook, T. D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton, Mifflin.
- Simpson, L., Korenbrot, C., & Greene, J. (1997). Outcomes of enhanced prenatal services for Medicaid-eligible women in public and private settings. *Public Health Reports, 112*, 122–132.
- St. Pierre, R. G., & Rossi, P. H. (2006). Randomize groups, not individuals: A strategy for improving early childhood programs. *Evaluation Review, 30*, 656–685.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Walker, C. C. (2000). An educational intervention for hypertension management in older African Americans. *Ethnicity and Disease, 10*, 165–174.
- West, C. P., Dyrbye, L. N., Rabatin, J. T., Call, T. G., Davidson, J. H., Multari, A. . . . to promote physician well-being, job satisfaction, and professionalism: A randomized clinical trial. *JAMA Internal Medicine, 174*(4), 527–533. doi:10.1001/jamainternmed.2013.14387
- Zuckerman, I. H., Lee, E., Wutoh, A. K., Xue, Z., & Stuart, B. (2006). Application of regression-discontinuity analysis in pharmaceutical health services research. *Health Services Research, 41*, 550–561. doi:10.1111/j.1475-6773.2005.00487.x

CHAPTER 13

Sampling Designs and Data Sources for Effect Evaluations

In planning and developing the evaluation of the effects of a program, both sampling and data collection methods will arise as critical decisions. Both ought to be addressed from the point of view of having the most rigorous scientific approach possible, given the various realities of the health program. This chapter reviews both sampling and data collection methods with the intent of presenting the issues and possible solutions that can be used across a variety of programs. For program evaluation purposes, the term *sample* refers to the groups whose members were chosen to be part of the evaluation and who were selected from among those who received the program and from the intended audience who did not receive the program. Although *sampling* is more generally associated with research, the way in which the sample is selected for the effect evaluation can have a major influence on the results of the evaluation. As a consequence, important considerations for developing a sampling plan are covered, and a brief review of calculating response rates is provided here. The content covered will help program managers, evaluators, and stakeholders make the best possible choice of a sampling design, given the type of program,

the resource limitations, and any time constraints. The method for collecting the evaluation data is another key choice in planning an effect evaluation, and the major types of data collection methods are reviewed in this chapter.

► Sampling Realities

Devising a plan for selecting who will actually be included in the evaluation of program effects is often a creative—albeit technical—endeavor that consists of two basic steps (Rossi, Lipsey, & Freeman, 2004). The first step is to identify program participants and the intended population. The second step is to develop a plan to select an unbiased sample from the intended population and from among program participants. When discussing sampling, statisticians use the term *population* to refer to the group from which a sample is selected. This group may or may not be the same as the larger intended population of the program, and it may or may not be the same as the way clinicians use this term.

Unlike sampling for research projects that are under the control of investigators, sampling for program evaluations places several constraints

on evaluators. The foremost limitation is the number of people who can or did participate in the program. If the health program was a small health education class, then including all program participants in the evaluation sample may be feasible. In contrast, if the health program was delivered to the population at large, such as a public awareness campaign or passive protection through policy implementation, then it becomes necessary to select individuals from within the intended population for the evaluation. A corollary constraint to the number of program participants is the size of the intended population, which can vary from a country, if a national health policy is being evaluated, to a small, discrete group, such as adults between the ages of 75 and 80 with glaucoma who live within a small geographic location. Because nonparticipants in the intended audience become controls, the number of potential controls also needs to be taken into consideration.

The second sampling constraint is that it is not always clear who was a member of the intended audience and, more important, who was a participant. Such blurring of lines occurs when there is either unclear program eligibility or inadequate service utilization documentation. Unless the program has clearly delineated criteria for membership in the intended population and for designating a participant, it may not be possible to know who ought to have received the program and who actually received it. This fuzziness can make it difficult or impossible to know who is appropriately classified as a member of the exposed/experimental group versus the unexposed/control group. Hence, there is a need to have developed clear eligibility guidelines and procedures as part of the process theory. Ambiguous group membership may also result if the evaluator has limited access to or ability to identify program participants. This situation could arise if the program has poorly maintained records or if it is provided anonymously on a drop-in basis. A lack of such information makes it difficult to classify who was or is in the exposed/experimental group and unexposed/control group, which in turn has

implications for both design choice and sampling strategy. Just as being able to delineate program participants begins with the process theory, the issue of obtaining information about who the program participants are can be addressed during the development of the process theory.

A third constraint involves how participants are classified by the program, meaning which criteria are used to assign evaluation participants to either the program/experimental group or the control group (i.e., those individuals not receiving the program). In some programs, participation in the program is not a clear-cut, dichotomous variable. This fuzziness is particularly likely with programs that have multiple components that may be provided by multiple providers, or programs that are implemented over an extended time. For example, Manalo and Meezan (2000) were interested in evaluating family support programs that varied with regard to actual content. The typology used to classify the family support programs (the unit of analysis) proved problematic when these authors attempted to evaluate outcomes across programs, which limited their ability to draw conclusions about the effectiveness of different types of programs. Thus, a key part of the sampling strategy is the development of a definition of “participation” in the program. This definition may be the same as the definitions developed for the process evaluation and may be based on a wide variety of criteria, ranging from hours of intervention received to membership in the intended population of a health policy.

The approach chosen for constructing the evaluation sample, along with the design choice, has implications for the ability to draw statistical conclusions. The more carefully the evaluation needs to compare program participants (exposed/experimental) to nonparticipants (unexposed/control), the more carefully sample selection needs to proceed in terms of making the groups as alike as possible. In such cases, random assignment to either the program or not is an important element of the evaluation plan. For example, if the program being evaluated delivers a novel intervention and the evaluation is akin

to evaluation research, then attention to random assignment becomes relevant. In contrast, for most local, ongoing, or smaller health programs, the efforts and resources required to accomplish random assignment are beyond the scope of what is needed or expected by stakeholders.

► Sample Construction

Two broad classifications of sampling approaches are distinguished: probability sampling and nonprobability sampling. Within each of these two approaches are sample types based on increasingly complex methods used to derive the sample. **TABLE 13-1** summarizes the differences between the probability and nonprobability sampling techniques.

If each member of a population has a known chance, or probability, of being chosen to participate in the program, then the sample is said to be a *probability* or *random sample*, depending on the discipline. To have a random sample requires having a list of all possible evaluation participants and then randomly sampling who is included in the effect evaluation. A probability sample, in theory, allows the evaluator to create groups that ought to be the same when statistically compared. This type of sample is recommended

if the evaluation is seeking to demonstrate that the program, and only the program, was responsible for the outcome—in other words, causation. Ideally, probability samples are used with outcome evaluation designs—specifically, experimental or some quasi-experimental designs. A probability sample would be used in an effect evaluation only if the program delivered a novel intervention and therefore was being studied for its effectiveness. Most evaluations of ongoing or smaller programs encounter one of two barriers to using a probability sample: the ethical issues inherent in randomly selecting evaluation participants, and unavailability of a sufficiently large sampling frame for random selection.

The other approach to sampling does not rely on chance to select members of the population for participation in the evaluation, so it is called a *nonprobability* or *nonrandom sample*, depending on the discipline of the evaluator. With this sampling technique, there is no attempt to select individuals randomly for participation in the effect evaluation. Nonprobability samples are used with outcome documentation and outcome assessment designs. None of the nonprobability sample types enables evaluators to say that the samples in the evaluation are representative of the population in need or of participants.

TABLE 13-1 Probability and Nonprobability Samples and Their Usage

	Type of Sample	
	Probability	Nonprobability
Key Characteristics	Each population element has a chance of being selected; known probabilities of selection	Unknown probabilities of selection
Situations When Preferred	When evaluation must demonstrate causation (i.e., experimental or quasi-experimental designs); effect evaluation of a novel program	Outcome documentation and assessment designs; no sampling frame; very small population size; hard-to-reach population

Hard-to-Reach Populations

Hard-to-reach populations refers to individuals who are not easily identifiable or readily accessible for participation in intervention programs or program evaluations (Faugier & Sargeant, 1997). The term *hardly reached* has begun to replace the term *hard-to-reach*, shifting the responsibility for not being reached from the individuals to the program planners and evaluators (Sokol & Fisher, 2016). Typically, *hard-to-reach* individuals are members of rare subgroups within a population and are not readily identifiable in commonly available lists such as agency records. Health programs, for example, often address sensitive issues or health behaviors that are not legal. Recruiting evaluation participants from *hard-to-reach* groups, such as intravenous drug users or individuals experiencing homelessness, may be complicated by both agency privacy protections and personal desires for privacy (Marpsat & Razafindratsima, 2010). For the sake of rigor and a defensible evaluation, however, it may be imperative to include such individuals in the evaluation.

A variety of strategies have been developed as means to include individuals from *hard-to-reach* populations in a sample. One strategy is random-digit dialing (Blair & Czaja, 1982), which involves generating phone numbers based on randomly generating the last four digits of the telephone number. This approach overcomes the problem of accessing individuals with unlisted phone numbers. However, the decreasing proportion of people who have land lines makes this technique less effective, and telephone directories for cell phone numbers do not exist. Alternatively, an effective technique is to recruit via online techniques and websites, with additional advertising for recruitment done through social media that is accessible on cell phones (Wilkerson, Iantaffi, Grey, Bockting, & Rosser, 2014). Another technique is capture–recapture (Xu, Fyfe, Walker, & Cowen, 2014), which involves using two or more lists or observational periods to identify unique individuals who might be eligible for participation.

Another set of techniques has been called a variety of names—multiplicity, referral, or snowball sampling (Rothbart, Fine, & Sudman, 1982)—but is most commonly referred to now as respondent-driven sampling (McCreesh et al., 2012). With this approach, a key informant provides a referral to other potential participants, each of whom then provides additional referrals, and so on. Over time, the list of potential participants snowballs, growing through a multiplicity of referrals. This approach has been used successfully to access very specific subgroups, and when a clear definition of the population of interest exists and when members of that group have a high degree of network connections that make for a concentration of that group.

Another technique for reaching the *hard-to-reach* is venue-based sampling (Muhib, Lin, & Stueve, 2001). This strategy entails going to specific types of locations to find potential participants. It was first developed as a way to reach very-high-risk individuals whose high-risk behaviors were associated with specific locations, such as bath houses, gay bars, or tattoo parlors. Based on this logic, going to alternative high schools to recruit high-risk adolescents into an evaluation of the pregnancy prevention program could be considered using venue-based sampling. A similar technique is time–space sampling (Muhib et al., 2001).

This brief overview of techniques to obtain a sample from *hard-to-reach* populations ought to bring to mind that these same approaches could have been used to recruit the actual program participants. If this is the case, then at the time of the program recruitment, a process ought to have been in place for simultaneously recruiting *hard-to-reach* individuals into the evaluation of the program.

Sample Size

The question of how many evaluation participants—whether individuals, families, or communities—are needed in the evaluation can be a complex issue. Evaluators face a host of practical considerations, such as fiscal limitations, logistics of

data collection and management, and accessibility challenges. Statistical considerations also complicate matters. In a pure research study, the investigator first chooses a level of probability that a significant result might be found by the study. This level is called the *power*, and it is usually chosen to be between 80% and 90%. The number of study participants needed to achieve that level of power is then calculated. When the power is set higher, meaning a higher probability that a significant result will be found if the result truly exists, more subjects are needed. Also, if a large number of moderating, mediating, intervention, or antecedent variables will be included in the evaluation and analyses, then more evaluation participants are needed. The other key element in calculating the sample size is the degree of difference that is expected, called the *effect size*. This difference can be based on the degree of difference between pretest and posttest data or between program participants and nonparticipants.

The statistical process of analyzing the relationships among the power of a study, the type of statistical tests to be done, the number of variables, the effect size, and the sample size is called a *power analysis*. Black (1999) summarizes the issues involved in such an analysis by showing that four factors affect the power of a statistical test: the sample representativeness and size, the quality of the measurement in terms of reliability and design, the choice of the statistical test, and the effectiveness of the intervention (**FIGURE 13-1**). Several software programs can be used to perform the calculations necessary for a power analysis, many of which are available via the Internet (see Internet Resources section). Whether freeware or commercial programs are used, the power analysis requires a clear understanding of the sample, the size of the anticipated effect of the program, and the statistical tests that will be used.

The power analysis might lead the program evaluator to refine, modify, and solidify aspects of the evaluation design that are under the control of the evaluator in an effort to enhance the study's power. For most program evaluations,

however, the number of evaluation participants may already be determined, perhaps because funding is limited or because the program has already been provided. In such circumstances, the power analysis works backward from the given sample size to provide the evaluator and the stakeholders with a sense of how realistic the chances are that the evaluation will find a meaningful difference. With a predetermined sample size, the design of the evaluation takes on greater importance in determining whether the possibility exists of finding a significant difference that might be related to the program. In addition, the power analysis may reveal that the possibility of finding a program effect is so unlikely as to make an outcome documentation or assessment approach might be a better approach for evaluating the program's effectiveness. The evaluator must then make the difficult choice of whether to proceed with the evaluation, search for alternative approaches to the evaluation or data analysis as assessed by the power analysis, or determine whether it is feasible to modify the evaluation in ways that will increase the power.

Calculating Response Rates

During and at the conclusion of the evaluation, the response rate is a key factor that influences the interpretation of the results. *Response rate* is the percentage of individuals who were invited to participate in the evaluation and who actually participated in the evaluation. Response rate is mentioned in conjunction with sampling because of the need to keep in mind that the best possible response rate depends on having a sampling strategy that not only is rigorous but also yields the most possible participants in the evaluation. A basic response rate—say, for a survey—is calculated as the number of usable surveys divided by the number distributed, times 100; this calculation yields a percent response. The same formula can be used with slightly different numerators and denominators if the goal is to provide more detailed information about response rates for different groups in the evaluation or for

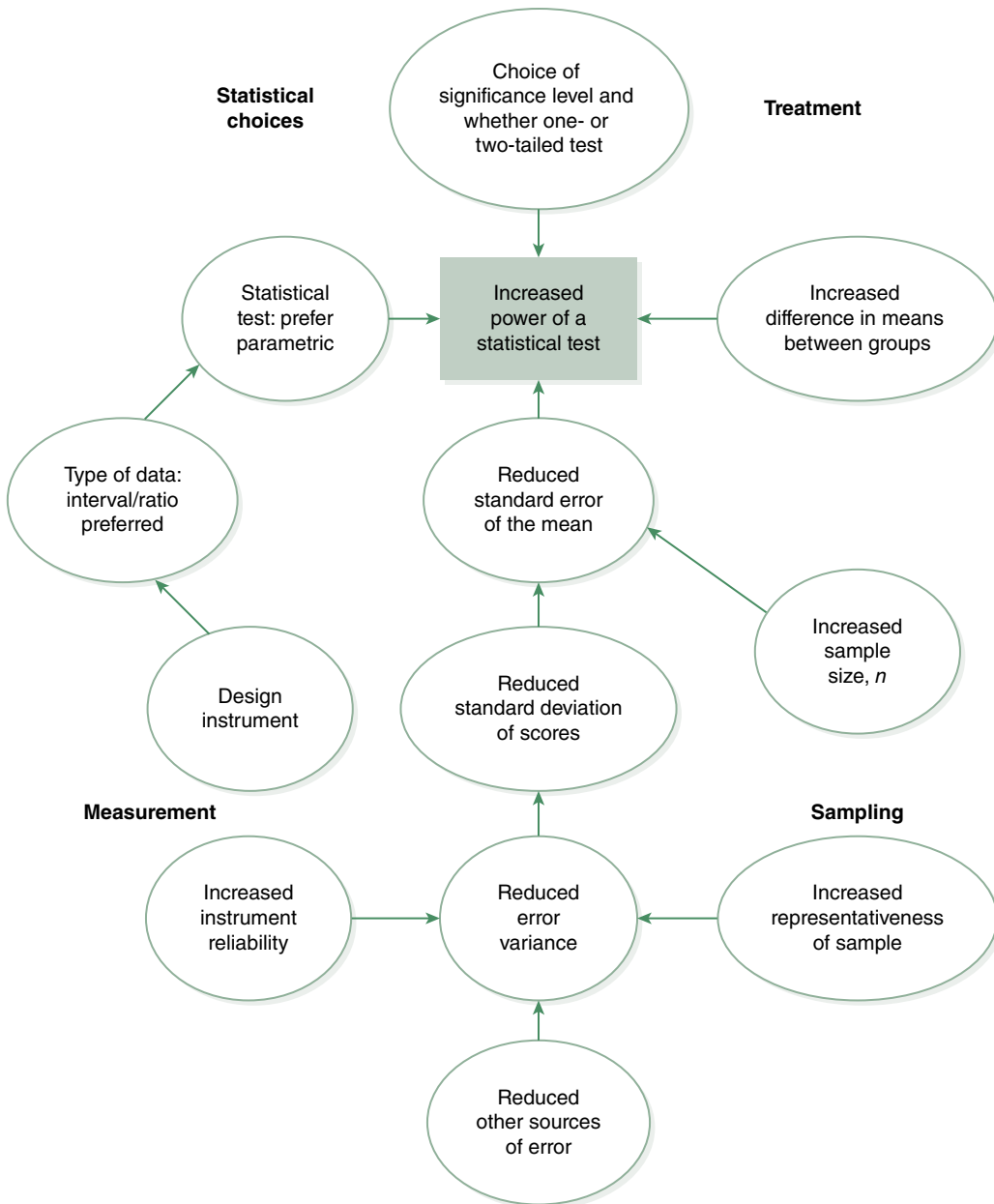


FIGURE 13-1 Probability and Nonprobability Samples and Their Usage

Modified from Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement and statistics*. London: Sage.

different portions of the evaluation. Response rates typically range from a high of 80%, which is achieved with considerable effort and expense, to a low of less than 30%, which is typical for

a mail survey for which minimal incentives for participation are provided.

A plethora of minor variations on that basic formula exist, which then lead to minor differences

in the response rate, depending on three key factors. The first key factor is whether those invited to participate are ultimately eligible to participate in the evaluation. It is easy to imagine situations in which many people are invited to become part of the evaluation but, after going through a brief set of screening questions, are found not to fit the criteria for participating in the evaluation. Subtracting the number found ineligible from the number invited leads to a more realistic denominator of “invited persons who are eligible.” A second key factor that affects the response rate is whether, once invited and found to be eligible, people agree to participate. The third key factor that affects response rates is whether evaluation participants complete their participation in the evaluation. Partially completed participation, such as a half-completed self-administered questionnaire, also affects the calculations of the response rate.

Individuals may either explicitly refuse to participate in an evaluation or not respond to repeated contact attempts. Tracking both the number and reasons for refusing or declining to participate is required for evaluation research, in which the evaluation has approval from the human subjects protection board. To the extent that any of the three key factors might affect the final interpretation of the evaluation findings, it will be important to set up a mechanism to track eligibility, refusal, completion, and invitations.

Nonresponse Bias

Regardless of whether evaluators have a detailed and well-constructed survey design, some evaluation participants will inevitably fail to provide data for any number of reasons. Efforts must be made to minimize these nonresponses to increase the response rate. A low response rate, less than 50%, is important for two reasons. First, those who reply may not be like those who do not reply, biasing the sample and having subsequent consequences on the data and findings. The extent to which any differences between nonparticipants and participants alter the findings is called nonresponse bias. Second, it is costly to try to achieve the desired sample

size by continually attempting to identify and obtain data from “replacements.”

Nonresponse may be due to *attrition*, meaning that program participants or control subjects are no longer part of the evaluation or the program. Attrition occurs for various reasons: no longer fitting the criteria for being in the program or the evaluation, loss to follow-up, death, or declining or refusing to continue. Attrition is a normal part of the evaluation process, and the sample size must be based on the expected loss of 10% to 40% of participants. Attrition rates are particularly crucial for outcome assessment and outcome evaluation designs because they affect the final sample size and can affect the balance of participants and nonparticipants sought through a carefully constructed sampling method.

Some individuals selected for the evaluation will be difficult to reach or to convince to participate in the evaluation. Aggressive recruitment efforts may be needed in such situations. For example, evaluators may need to make multiple attempts to contact these individuals, using different media (telephone, e-mail, and letter) at different times by different voices. The number of attempts made to reach an individual needs to be carefully weighed against the possible appearance of harassing the individual, and the extent to which repeated attempts and refusals will affect the quality of the data eventually obtained. In many situations, evaluators—and particularly external evaluation consultants—have little or no control over who is in the program and little or no control over who is selected for the evaluation. In such instances, additional efforts are needed to address the program staff members as key stakeholders in the evaluation and to train or educate them about recruitment and retention techniques. Actions of program staff members with regard to how the evaluation is presented and supported can dramatically influence the participation and response rates.

Incentives

Incentives, such as money, gift certificates, or small tokens of appreciation, are effective

in increasing participation in research. Many individuals participate in research for humanitarian reasons, such as their desire to contribute to making life better for others and to advance science (Soule et al., 2016). Alternatively, some program participants may be characterized by the attitude “pay me to participate,” which is particularly likely for high-risk individuals who view monetary incentives as additional income (Slomka, McCurdy, Ratliff, Timpson, & Williams, 2007). Thus, it is likely that at least some participants in evaluations of health programs will expect a monetary incentive. An extensive body of literature exists on the use of incentives in research and on the monetary amounts that are most effective in increasing the response rate and comparing various types of incentives. For example, Rodgers and colleagues found that electronic bank cards lead to higher rates of survey completion than cash (Rodgers, Miesel, Weibe, Crits-Christoph, & Rhodes, 2016). This knowledge about incentives needs to be used in developing the data collection procedure, along with common sense and a working understanding of the community standard for incentives for a comparable request of participants’ time and effort.

Designing the monetary incentive must take into account both the total amount and the payment schedule. The amount must not be so great that it would be perceived as coercive nor should it be so small as not to be an incentive. The payment schedule for providing evaluation data needs to be congruent with the frequency with which participants are asked to provide data. The rule of thumb is that incentives ought to be provided at each data collection point.

Consider an evaluation of a 6-week diabetes self-management class for Layingville residents. If class participants are asked to complete a survey before the first class and at the end of the last class, then the incentive could be for the same amount and payments could be delivered at the time the surveys are returned. However, if the participants are asked to undergo physical testing before the class, then the incentive may need to be larger to reflect that more extensive request

or burden on the participant. If the evaluation is longitudinal, for example, with data collection at six months and one year after the classes have ended, then a slightly larger incentive may be needed for those time periods as a way to keep participants’ interest in the evaluation study.

► Sampling for Effect Evaluations

Sampling for Outcome Assessment

Using a nonprobability sample, as is done in outcome assessments, tends to be simple and inexpensive (TABLE 13-2). For many smaller, locally based, or agency-specific health programs, a nonprobability sample will be adequate for the evaluation. This type of sample allows for a statistical comparison of differences between program participants and nonparticipants.

A nonprobability sample can be constructed in several ways (Table 13-2). These types of samples come from the social sciences. A *convenience sample* is constructed by inviting whoever is accessible or available to participate; it is an inexpensive means of obtaining a sufficient number of evaluation participants. In a *purposive sample*, the evaluation participants are typically chosen based on a specific characteristic, thereby ensuring that the program participant and non-participant samples are balanced with regard to that characteristic. Purposive sampling can yield samples that represent the larger population in key attributes believed to affect program outcomes (Yin, 2013).

A *quota sample* also involves selecting participants based on a specific characteristic or strata such that the proportion of evaluation participants having that characteristic or being in each stratum is proportional to those in the population at large. For example, Ramachandran, Wu, He, Jian, and Wang (2016) distributed the number of study participants to reflect the national distribution of women across

TABLE 13-2 Comparison of Main Types of Samples with Regard to Implementation Ease, Degree of Representativeness, and Complexity of Sampling Frame

Type of Sample	Implementation Ease	Representativeness Compared to General Population	Sampling Frame
Nonprobability Samples			
Convenience	Easiest	None ensured but may occur by chance	Willingness to participate in the evaluation
Purposive	Easy	None	Specific characteristics of interest
Quota	Moderately easy, but must track number in each quota category	None ensured but possibly representative of those with characteristics by chance	Specific characteristics of interest
Snowball	Somewhat difficult	None; likely to be biased	Network of initial participants
Probability Samples			
Simple random	Easy-to-use random number chart	High	Entire population
Stratified random	Moderate because must first choose stratification variable and then stratum categories	High	Entire population, but must have information to assign individuals to strata
Systematic	Easy to select each <i>n</i> th from a list	Moderately high, but lower if the listing sequence is not random	List of possible evaluation participants
Random route	Difficult because must define area, construct a random route, and then choose <i>n</i> th house for inclusion	Moderate to poor, depending on the diversity or homogeneity of the residents in the area chosen and the availability of residents	Geographically accessible area
Cluster or nested sampling	Moderate once the cluster has been identified	Moderately high if using random selection of clusters and individuals within clusters	Population with naturally occurring nested clusters

occupational groups. Thus, each occupational group had a quota number of participants. In another example, let's say that age is important in the evaluation and 10% of the population of participants at large is more than 80 years old; then the evaluation sample must meet a 10% quota of 80-year-old participants. A snowball sample is achieved by asking current evaluation participants who have a specific characteristic to identify or nominate other individuals they know who also have the characteristic of interest. As mentioned earlier, this type of sampling is useful for accessing hard-to-reach populations and when a list of names of potential evaluation participants does not exist. As more evaluation participants name others, the snowball of evaluation participants grows.

Overall, these sampling strategies are easy to implement and explain to stakeholders, and therefore they are likely to be used in program evaluations. Each sampling design varies with regard to the ease of implementation, the degree of representativeness of the larger population, and the sampling frame used—in other words, the basis for inclusion in the evaluation sample.

Sampling for Outcome Evaluation

If it is crucial to demonstrate that the health problem was changed (and presumably improved) in program participants, then a probability sample is recommended as part of the program evaluation. Such a sample is also necessary in a needs assessment if the assessment is intended to estimate accurately the rate of a health problem in a population. A *probability sample* is one in which all potential members of the evaluation have a known probability of being selected to participate in the evaluation. The major barrier to obtaining a probability sample is that evaluators may not have control over who receives the program, especially if the outcome evaluation is not designed during program development. Probability samples are used to increase the external validity of the evaluation; however, for the

vast majority of program evaluations, external validity is less of an issue than the biases and threats inherent in the design.

Achieving known probability of selection involves randomly selecting potential participants. Several types of probability samples can be constructed, depending on the specific method used to identify and then select members of the evaluation sample. Random selection of evaluation participants—whether individuals, classrooms, or neighborhoods—from the entire population of possible participants can be done in one of several ways, with each technique resulting in a different type of probability sample. A simple probability sample, for example, is constructed by using a table of random numbers to select individuals from a known list of possible participants. The other types of probability samples involve increasingly more complex selection procedures from increasingly more specific groups within the population.

The various probability sample types are explained in greater detail in various textbooks (e.g., Thompson, 2012). Most local or agency-based program planning efforts and evaluations are not likely to have the resources required to construct and obtain these more complicated samples. The costs associated with the sampling effort increase in proportion to the complexity of the probability sample because of increases in the number and qualifications of personnel needed to find the individual selected, and because of increases in the amount of time required first to establish the sampling procedures and then to carry out those procedures. The sample type chosen may be influenced by the ease of implementing that technique, the degree of representativeness of the intended population, and the complexity of the sampling frame.

► Data Collection Methods

Methods refers to techniques used to collect data, whereas design is the overall plan or strategy

for when and from whom data are collected. Methods generally fall into one of two categories: those for collecting primary data (i.e., the generation of new data) and those for collecting secondary data (i.e., the use of existing data). The evaluation method needs to be consistent with the purpose of the evaluation and the specific evaluation question. The following discussion of methods and data sources focuses on collection of both primary and secondary quantitative data. Methods to collect qualitative data are sufficiently different and are not covered in this chapter. The most common form of primary data collected is through the use of surveys and questionnaires. For each health and well-being domain, various sources of data can be used to generate information (TABLE 13-3).

Surveys and Questionnaires

A *survey* is a method that specifies how and from whom data are collected, whereas a *questionnaire* is a tool for data collection. Typically surveys use questionnaires to collect data. For example, the U.S. Census is a survey that uses a questionnaire to collect data on all persons who reside in the United States. In most cases, residents complete a pen-and-paper questionnaire. However, in some instances, a census taker completes the questionnaire while talking with the individual. Although the distinction between a survey and a questionnaire is important for the sake of clear thinking, generally the word *survey* implies the use of a questionnaire.

Questionnaire Construction Considerations

Much has been written about ways to construct health questionnaires (e.g., Aday & Cornelius, 2006), ways to write individual questions on the questionnaire (Aday & Cornelius, 2006; Krosnick, 1999), and techniques to have sets of questions form a valid and reliable scale (DeVellis, 2003; Fowler, 1995; Pedhazur & Schmelkin, 1991). Several key points can be drawn from these resources that are paramount to developing a good health program evaluation.

To the extent possible, evaluators should use existing questionnaire items and valid and reliable scales so that they can avoid spending precious resources “reinventing the wheel.” An example of existing items is the U.S. Census Bureau race/ethnicity categories. These race/ethnicity items can be used rather than creating new race/ethnicity categories for a particular evaluation. An advantage of using existing items is that it provides some assurance that the items are understandable. Existing scales also can be used, which makes it possible to compare evaluation participants with those with whom the scale was previously used. However, if the intended audience has a unique characteristic—for example, a specific medical diagnosis—that is relevant to understanding the effect of the program, then comparison to existing scales may not be the optimal choice.

Instruments also need to be appropriate for diverse ethnicities and possibly multiple languages. In terms of instruments, cultural sensitivity has two dimensions: the surface structure, which consists of the superficial characteristics, and the deep structure, which consists of core values or meanings. This second dimension is sometimes called cultural relevance. Attention to both careful translation and cultural relevance is especially needed for questionnaires that are being used for the first time with a different cultural group, such as was done by Yu, Wu, and Mood (2005). Epstein, Osborne, Elsworth, Beaton, and Guillemin (2015), in studying translation for one health education questionnaire, identified five different translation errors: question style, frequency or time frame, breadth of what is meant, strength of emphasis, and actual meaning.

One type of scale that is likely to be discussed with regard to evaluating health programs is the use of client goals. MacKay, Somerville, and Lundie (1996) reported that this evaluation technique has been used since 1968. Program staff members may be inclined to count the number of client goals attained as an indicator of program success; indeed, the temptation is to consider this quantity as outcome data that are very readily available. Unfortunately, this crude measure of client outcome is highly problematic from an evaluation

TABLE 13-3 Example of Data Sources for Each Health and Well-Being Domain

Health Domain	Examples of Data Sources
Physical health	Survey data: self-report Secondary data: medical records for medical diagnoses Physical data: scale for weight, laboratory tests Observation: response to physical activity
Knowledge	Survey data: self-report, standardized tests Secondary data: school records Physical data: not applicable Observation: performance of task
Lifestyle behavior	Survey data: self-report Secondary data: police records Physical data: laboratory tests related to behaviors, such as nicotine or cocaine blood levels Observation: behaviors in natural settings
Cognitive processes	Survey data: self-report, standardized tests of cognitive development and problem solving Secondary data: school records Physical data: imaging of brain activity Observation: problem-solving tasks, narrative
Mental health	Survey data: self-reported motivation, values, attitudes Secondary data: medical records diagnostic category Physical data: self-inflicted wounds, lab values Observation: emotional bonding
Social health	Survey data: self-report, social network questionnaires, report of others Secondary data: attendance records of recreational activities Physical data: not applicable Observation: interpersonal interactions
Community-level well-being	Survey data: self-report of safety and community cohesion Secondary data: neighborhood crime rates, number of primary care clinics Physical data: number of vacant lots, number of grocery stores Observation: children and adults using parks, pedestrian traffic, amount of fresh produce in grocery stores
Resources	Survey data: self-report Secondary data: employer records, country marriage records, school records Physical data: address Observation: possessions

perspective. The main problem is that, unless the goals are highly standardized for specific health problems, there can be great variability in the goals set. Similarly, unless strict criteria have been established for determining whether a client goal was reached, biases among program staff members may influence client assessments. The use of goal attainment scaling, in which a Likert-type scale specific to each goal is used, still poses serious problems (MacKay et al., 1996) and therefore its use ought to be severely curtailed.

To assess the readability, ease of completion, and overall appeal of the questionnaire, a pretest or pilot test is advised. Involvement of stakeholders in this activity is encouraged for two reasons: It helps the evaluators have a better questionnaire given the intended audience, and it helps stakeholders anticipate what the evaluation data will include. Key considerations are to keep the language simple, use an easy-to-follow format and layout, and break down complex concepts into more easily understood ideas. Even if evaluation participants are expected to be well educated, people are more likely to complete questionnaires that are easy and quick to read.

Verify that what is in the questionnaire corresponds to the program outcome objectives. Evaluators are often tempted to add “just a few more questions” because the opportunity exists or because the information might be interesting to know. A shorter questionnaire is both better and more likely to be completed. A good rationale for going beyond the program objectives in what is collected is if those data will be used for subsequent program planning or to refine the current program.

Regardless of the care taken to construct a questionnaire, whatever can be misinterpreted or be done wrong will inevitably happen. For example, questionnaires that are copied double-sided and are stapled together are likely to have skipped pages. Unless the questionnaire is administered by an interviewer who is well trained, do not use skip patterns that direct respondents to skip questions based on a previous response. These complicated patterns quickly become confusing, and the well-intending respondent may answer

all questions, including those items that ought to have been skipped. Use of skip patterns is really appropriate only for questionnaires conducted in person or over the phone.

Survey Considerations

Any survey, whether done in person or via mail or e-mail, needs careful planning. The process by which the questionnaire is distributed and returned to the evaluator must be planned thoroughly to minimize nonresponse and non-participation rates. That process is as critical as the quality of the questionnaire to the success of the evaluation. One technique for developing a well-crafted survey plan is to imagine and role-play each step in the process, and to follow the paper from hand to hand.

Collection of questionnaire data is being done electronically more and more often, whether via handheld devices used in the field, Internet-based surveys, or client-accessed computers. The same advice about the need to have a carefully crafted data collection plan applies to the use of electronic data collection: Follow the answers from the asker to the responder through data entry to computer output. Each of these steps is needed to be able to collect the data accurately and feasibly, and constitutes the survey design.

Response Biases

A threat to the quality of questionnaire data, and especially to self-report data from individuals, comes from the various types of *response bias*, the intentional or unconscious systematic way in which individuals select responses. One of the most common types of response bias, known as *social desirability*, is answering questions in a manner intended to make a favorable impression (Krumpal, 2013). Social desirability is a powerful motivator and has been widely included in program evaluations in which there is the potential for the participant to want to please the evaluators or when the participants believe there is a socially correct answer they are supposed to give.

TABLE 13-4 Interaction of Response Bias and Variable Error

		Variable Error	
		Low	High
Bias	Low	Ideal: high range of honest responses on good measure	Questionable but acceptable data from high range of honest responses on poor measure
	High	Questionable but acceptable data from skewed responses (i.e., toward socially desirable responses) on good measure	Unusable data due to skewed responses on poor measure

Response bias can also occur as a result of the respondent falling into a pattern of just giving the same response, regardless of the question or his or her true opinion or feeling. Response bias can be difficult to anticipate. Nonetheless, evaluators would be wise to consider that both response bias and errors inherent in the way the variables are measured can interactively produce questionable or even totally undesirable data (TABLE 13-4).

Secondary Data

Secondary data are data that have already been collected and are now being used for a purpose that is secondary to their original purpose. Some sources of existing data are appropriately used to assess the effect of health programs; others are not. Each source of secondary data must be carefully considered with regard to its quality. Evaluators must decide whether the data are actually needed to answer the evaluation question.

Vital records—namely, birth certificates, death certificates, and disease registries—are a major source of secondary data for health program evaluators. Birth records contain a wealth of information on prenatal variables, delivery complications, and infant characteristics. These records are usually not available for up to a year past the date of the birth of the infant, so evaluations of prenatal programs that are designed to

affect birth outcomes will not be able to include data from birth records immediately following the program. If the evaluation is longitudinal and focuses on trends, then birth record data may be useful. However, pinpointing the time of the programmatic intervention may be challenging. In addition, for community-based interventions, sampling comparable communities for comparison of birth data will need to take into account how to select the two communities using the address information on the birth certificates. These same caveats to using birth data apply to data from death certificates or disease registries.

Medical records, case files, or insurance claims may also contain information desired for the evaluation. Several issues must be considered before embarking on data abstraction. First is the quality of the data as recorded and available for abstraction. Because the data in such records are collected for clinical purposes rather than evaluation purposes, the information can be inconsistent and may vary by the practitioner recording the data. If the evaluator has reason to believe that data in the record are recorded reliably, the evaluator must then devise a reliable way to abstract the data. This effort involves training individual data abstractors. If any interpretation of the record information is required, guidelines for what will be recorded and decision rules for interpretation must be understood and applied consistently by all of

the data abstractors. Typically, the goal is at least 80% agreement between any two abstractors on the coding of data from a single data source.

Another source of secondary data is national surveys, such as the National Health and Nutrition Examination Survey (NHANES) or the National Survey of Family Growth (NSFG). These and several other surveys are conducted periodically by various federal agencies with a health focus, including the Occupational Safety and Health Administration. These data sets have often been used for community assessment. Data from these surveys are accessible to the public through the Internet; they can be used for evaluation of population-level programs. Some data sets have restrictions or stipulations on their use that must be addressed before they can be used. A drawback to using these broad surveys is that the most recent data can be as much as two years old. As secondary data sets, they may be of limited value in determining the effect of small-scale programs. By contrast, they may be highly useful if the effect evaluation focuses on a population-level health program, such as a state program, and the timing is such that immediate information is not critical.

The use of large secondary data sets for the evaluation of programs faces the challenge of overcoming conceptual issues, such as associating the variables available in the data set to the program theory and determining the reliability and validity of the data. Other pragmatic considerations arise as well, such as selection of subsamples and the need to recode data. In addition, data from some national surveys may not generate results applicable to rural populations (Borders, Rohrer, & Vaughn, 2000). Overall, the evaluator needs to be cautious and have a specific rationale for using large secondary data sets for an effect evaluation.

Big Data

Large amounts of complex, diverse data from surveillance, program client records, electronic health records, and billings are increasingly available for public health evaluation. In addition, federal and public efforts to promote healthcare transparency have resulted in making decades

of data available for research (Groves, Kayyali, Knott, & Kuiken, 2013). These have commonly been referred to as big data and have been defined as “large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information” (Cottle et al., 2013).

Big data is characterized by the four V’s: volume, velocity, variety, and veracity (IBM, 2011). The continuous rapid generation and accumulation of health data from various sources creates a tremendous volume of data in structured, unstructured, and semistructured forms, all of which need credibility verification (Raghupathi & Raghupathi, 2014). Storage and management of digital, data such as those from billing and electronic health records, can overwhelm traditional data management methods (Raghupathi & Raghupathi, 2014), requiring customized applications and analytical tools (Groves et al., 2013). Examples of big data platforms and tools include Hadoop/MapReduce (open-source platforms on the cloud), Pig, Hive, Jaql, Zookeeper, HBase, Oozie, (Raghupathi & Raghupathi, 2014) and expandable infrastructures (i.e., data lakes, cloud data storage) (Roski, Bo-Linn, & Andrews, 2014).

Data analytics have numerous applications for health program monitoring. For instance, predictive analytics links numerous data sources (e.g., clinical, genetic, outcomes, claims, social data, and more) to predict outcomes within and/or outside established health systems (Groves et al., 2013; Schneeweiss, 2016). Monitoring can also occur through wearables such as the Fitbit and Apple Watch, as well as through smart clothing, that collect physiological data, integrated with technology from the Internet of Things (IoT) for daily care, rehabilitation, training, and chronic disease management (Chen, Ma, Song, Lai, & Hu, 2016). *Public health informatics* can be used to detect trends including risky sexual behavior, drug use, disease transmission and other public health threats, and events such as hospital readmissions suggesting some aspect of public health system failure (Lee et al., 2016; Luo, Wu, Gopukumar, & Zhao, 2016).

Issues that public health planners need to consider in relation to big data include availability, scalability, privacy, quality assurance, and ease of use (Raghupathi & Raghupathi, 2014). Big data are increasingly key components to informing public health as well as other sectors, and they are likely to affect the nature of evaluation in upcoming decades.

Physical Data

Biological samples, anthropometric measures, and environmental samples are examples of physical data that may be needed to evaluate a health program. Biological samples include things such as blood, urine, or hair; anthropometric measures are typically height, weight, and body mass index; and environmental samples range from ozone levels to bacteria counts in water supplies, to lead levels in fish. The decision regarding inclusion of physical data in the evaluation should be based on the health program goal and objectives, as well as the determination of whether the intervention and causal theories underlying the health program are substantiated sufficiently to justify the cost and effort needed to collect physical data, especially if laboratory tests are necessary.

As with the collection of other types of data, physical data need to be collected consistently. Evaluators may not have control over laboratory processing, so they need some assurance that any laboratory results are reliable. Evaluators need to be familiar with the laboratory standards for processing the samples and take steps to minimize factors that would lead to erroneous variation in results. Another consideration with regard to physical data, and specifically biological data, is the cost involved in collecting, storing, and processing the data. Generally, use of biological data in an evaluation can be quite an expensive proposition, and evaluators need to be proactive in budgeting for these expenses.

► Across the Pyramid

At the direct services level of the public health pyramid, health programs focus on individuals.

As a consequence, the sample frame is more likely to be accessible and knowable to the evaluators. If this is the case and the program is sufficiently large, a simple probability sample of the program participants is possible. However, getting a probability sample of nonparticipants may be more difficult.

At the enabling services level, health programs focus on groups of individuals and are provided in a wider range of contexts. Programs at this level of the pyramid will be most challenging to evaluate in terms of their effect, for several reasons. First, the sampling frame is less likely to be knowable and accessible to the evaluators, which necessitates that the evaluation sampling plan be creative and tailored carefully to the program realities. Second, some of the options may require the use of statistically constructed control groups. Third, response biases are more likely because of the challenge in accessing or collecting data from participants in enabling services programs.

At the population-based level of the public health pyramid, health programs are provided to entire populations. Although this does not preclude the use of probability sampling designs, it does limit the evaluation options to those sampling methods that can be implemented reasonably with populations.

At the infrastructure level, programs focus on changing the workings of the healthcare organization or the public health system. The effect evaluation question determines whether the evaluation concentrates on the changes in the infrastructure or the changes to the health status of clients. This distinction in turn influences the sample that is needed to answer the intervention effect question.

► Discussion Questions and Activities

1. Identify at least one sampling issue that would be particularly relevant at each level of the public health pyramid. Which strategy could be used to minimize these problems?

2. Conduct an Internet search for power analysis software. From among the programs you find, select two (that are freeware) and experiment with them. What information is required to conduct the analysis? In what ways does the need for that information affect how you would proceed with planning the outcome evaluation?
3. For a violence prevention program delivered in classrooms, two middle schools had all seventh- and eighth-grade students eligible to participate. In that district, there were five middle schools. Given that information, which type of effect evaluation samples would you suggest for this program, and why?
4. Issues related to obtaining good response rates were discussed in this chapter. List those issues. Which ones could be anticipated? Which are preventable? For the problems you listed, give at least one possible means of minimizing or preventing the response problem.
5. Suppose you plan on conducting a paired *t*-test with program participants, using their pretest and posttest data. Alpha significance is set at .05, and the standard deviation on the outcome measure was reported in the literature to be 3.5, with a mean of 12.0. Use this information to determine the sample size needed.

► Internet Resources

Sampling and Questionnaire Design

The Pew Research Center on U.S. Survey Research has a site for questionnaire design (<http://www.pewresearch.org/methodology/u-s-survey-research/questionnaire-design/>). Coursera offers a free online class called “Questionnaire Design for Social Surveys” offered by the University of Michigan, at <https://www.coursera.org/learn/questionnaire-design>

Harvard University’s Program on Survey Research offers a number of links to resources for survey research, available at <http://psr.iq.harvard.edu/book/psr-resources>

William Trochim’s website for research includes a discussion of sample types, with both graphics and formulas: www.socialresearchmethods.net/kb/sampling.php. Another website, www.surveysystem.com/sdesign.htm, gives tips and brief explanations for designing questionnaires.

American Association for Public Opinion Research

For some program evaluations, the response rate is very important. The website <http://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx> has a link to an Excel spreadsheet already set up to calculate response rates. However, it does require having counts for ineligible and refusals.

Power Analysis

Power analysis and determination of sample size can be intimidating. However, using programs can provide some reasonable estimates and allow you to understand the trade-offs between sample size and power. A Google search yields many power analysis resources. The two listed here are examples of what can be found. A free power analysis program, GPower (<http://www.gpower.hhu.de/>) is accessible over the Internet and is based on the work of Cohen (called the father of power analysis). A slightly more primitive but very functional calculator comes from a website from a Harvard course (http://hedwig.mgh.harvard.edu/sample_size/size.html); this calculator is intended more for clinical experimental designs that would be applicable for evaluation of small pilot programs.

References

- Aday, L. A., & Cornelius, L. J. (2006). *Designing and conducting health surveys: A comprehensive guide* (3rd ed.). San Francisco, CA: Jossey-Bass.

- Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement and statistics*. London, England: Sage.
- Blair, J., & Czaja, R. (1982). Locating a special population using random digit dialing. *Public Opinion Quarterly*, 46(4), 585–591. doi:10.1086/268755
- Borders, T. F., Rohrer, J. E., & Vaughn, T. E. (2000). Limitations of secondary data for strategic marketing in rural areas. *Health Services Management Research*, 13(4), 216–222. doi:10.1177/095148480001300402
- Chen, M., Ma, Y., Song, J., Lai, C., & Hu, B. (2016). Smart clothing: Connecting human with clouds and big data for sustainable health monitoring. *Mobile Networks and Applications*, 21(5), 825–845. doi:10.1007/s11036-016-0745-1
- Cottle, M., Hoover, W., Kanwal, S., Kohn, M., Strome, T., & Treister, N. (2013). Transforming health care through big data strategies for leveraging big data in the health care industry. Institute for Health Technology Transformation. Retrieved from <http://thehealthtran.Com/Big-Data-in-Healthcare>
- DeVellis, R. (2003). *Scale development: Theory and application* (2nd ed.). Thousand Oaks, CA: Sage.
- Epstein, J., Osborne, R. H., Elnsworth, G. R., Beaton, D. E., & Guillemain, F. (2015). Cross-cultural adaptation of the Health Education Impact Questionnaire: Experimental study showed expert committee, not back-translation, added value. *Journal of Clinical Epidemiology*, 68(4), 360–369.
- Faugier, J., & Sargeant, M. (1997). Sampling hard-to-reach populations. *Journal of Advanced Nursing*, 26(4), 790–797. doi:10.1046/j.1365-2648.1997.00371.x
- Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation*. Applied social research methods series: Volume 38. Thousand Oaks, CA: Sage.
- Groves, P., Kayyali, B., Knott, D., & Kuiken, S. V. (2013). The “big data” revolution in healthcare: Accelerating value and innovation. Summit, NJ: McKinsey & Company.
- IBM. (2011). *IBM big data and information management*. Retrieved from <http://www-01.ibm.com/software/data/bigdata/>
- Krosnick, J. A. (1999). Maximizing measurement quality: Principles of good questionnaire design. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of political attitudes* (pp. 37–57). New York, NY: Academic Press.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4), 2025–2047. doi:10.1007/s11135-011-9640-9
- Lee, E. C., Asher, J. M., Goldlust, S., Kraemer, J. D., Lawson, A. B., & Bansal, S. (2016). Mind the scales: Harnessing spatial big data for infectious disease surveillance and inference. *Journal of Infectious Diseases*, 214(suppl 4), S409–S413. doi:10.1093/infdis/jiw344
- Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big data application in biomedical research and health care: A literature review. *Biomedical Informatics Insights*, 8, 1. doi:10.4137/BII.S31559.
- MacKay, G., Somerville, W., & Lundie, J. (1996). Reflections on goal attainment scaling (GAS): Cautionary notes and proposals for development. *Educational Research*, 38(2), 161–172. <http://dx.doi.org/10.1080/0013188960380204>
- Manalo, V., & Meezan, W. (2000). Toward building a typology for the evaluation of services in family support programs. *Child Welfare*, 79(4), 405–429.
- Marpsat, M., & Razafindratsima, N. (2010). Survey methods for hard-to-reach populations: Introduction to the special issue. *Methodological Innovations Online*, 5(2), 3–16.
- McCreech, N., Frost, S. D. W., Seeley, J., Katongole, J., Tarsh, M. N., Ndunguse, R., et al. (2012). Evaluation of respondent-driven sampling. *Epidemiology*, 23, 138–147. doi:10.1097/EDE.0b013e31823ac17c
- Muhib, F. B., Lin, L. S., & Stueve, A. (2001). A venue-based method for sampling hard to reach populations. *Public Health Reports*, 116(suppl 1), 216–222.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum.
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, 2(1), 3. doi:10.1186/2047-2501-2-3
- Ramachandran, H. J., Wu, W. X., He, H., Jian, Y., & Wang, W. (2016). Awareness, knowledge, healthy lifestyle behaviors, and their correlates to coronary heart disease among working women in Singapore. *Heart and Lung*, 45, 341–349. <http://dx.doi.org/10.1016/j.hrtlng.2016.05.034>
- Rodgers, M., Miesel, Z., Wiebe, D., Crits-Christoph, P., & Rhodes, K. V. (2016). Wireless participant incentives using reloadable bank card to increase clinical trial retention with abused women drinkers: A natural experiment. *Journal of Interpersonal Violence*, epub ahead of print, August 7, 2016. doi:10.1177/0886260516662849
- Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: Opportunities and policy implications. *Health Affairs (Project Hope)*, 33(7), 1115.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Rothbart, G. S., Fine, M., & Sudman, S. (1982). On finding and interviewing the needles in the haystack: The use of multiplicity sampling. *Public Opinion Quarterly*, 46(3), 408–421. doi:10.1086/268737
- Schneeweiss, S. (2016). Improving therapeutic effectiveness and safety through big healthcare data. *Clinical Pharmacology & Therapeutics*, 99(3), 262–265. doi:10.1002/cpt.316.
- Slomka, J., McCurdy, S., Ratliff, E. A., Timpson, S., & Williams, M. L. (2007). Perceptions of financial payment for research participation among African-American drug users in HIV studies. *Journal of General Internal Medicine*, 22(10), 1403–1409. doi:10.1007/s11606-007-0319-9

- Sokol, R. & Fisher, E. (2016). Peer support for the hardy reached: A systematic review. *American Journal of Public Health, 106*, e1–e8. doi:10.2105/AJPH.2016.303180
- Soule, M. C., Beale, E. E., Suarez, L., Beach, S. R., Mastro-mauro, C. A., Celano, C. M., Moore, S., & Huffman, J. C. (2016). Understanding motivations to participate in an observational research study: Why do patients enroll? *Social Work in Health Care, 55*(3), 231–246. doi:10.1080/00981389.2015.1114064
- Thompson, S. K. (2012). *Sampling* (3rd ed.). Hoboken, NJ: John Wiley.
- Wilkerson, J. M., Iantaffi, A., Grey, J. A., Bockting, W. O., & Rosser, B. S. (2014). Recommendations for internet-based qualitative health research with hard-to-reach populations. *Qualitative health research, 24*(4), 561–574. doi:10.1177/1049732314524635
- Xu, Y., Fyfe, M., Walker, L., & Cowen, L. L. (2014). Estimating the number of injection drug users in greater Victoria, Canada using capture-recapture methods. *Harm Reduction Journal, 11*(1), 1. doi:10.1186/1477-7517-11-9
- Yin, R. (2013). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.
- Yu, M. Y., Wu, T. Y., & Mood, D. W. (2005). Cultural affiliation and mammography screening of Chinese women in an urban county of Michigan. *Journal of Transcultural Nursing, 16*(2), 107–116. doi:10.1177/1043659605274745

CHAPTER 14

Quantitative Data Analysis and Interpretation

Patton (2008) emphasized the need for evaluators to think carefully in advance about what stakeholders want to know from data and which actions or decisions they may make based on the findings. This process serves two functions: It allows the evaluator to educate the stakeholders with regard to the evaluation design, and it helps establish realistic expectations regarding the findings of the evaluation. Then, when data analysis is in progress and the stakeholders are involved in interpretation of the data, they will be better prepared to understand both the data and its limitations. Patton's advice extends the involvement of the stakeholders from the program planning into its evaluation.

Needless to say, both stakeholders and evaluators will be focused on whether the program makes a difference. Program effect evaluations are essentially efforts to identify whether a significant degree of change occurred among program participants, preferably compared to the rest of the intended audience. Attention to the design of the effect evaluation, sample selection, and methods used to collect data lays the foundation for the subsequent statistical

tests and affects the overall trustworthiness of the statistical findings. This chapter provides a rudimentary review of statistical tests, but it is not intended to duplicate material available in statistical textbooks. The focus here is on the relationship of the statistical test to the design and sample, with an emphasis on understanding the implications of the statistical results in terms of program effect.

► Data Entry and Management

Data collected for the effect evaluation need to be electronic if they were not collected in a computerized format. Data collected on paper, as is often the case in survey methods, need to be entered into a computer database or spreadsheet, so that evaluators can conduct statistical analyses. The software to be used for data analysis must be chosen before beginning data entry.

A key consideration in the choice of statistical software is the sustainability of the evaluation.

Put simply, if program staff members and stakeholders are expected to be involved in the data entry and analysis, then their computer skills and interests need to be considered in choosing software. For the majority of program evaluations done by agencies, today's widely used spreadsheet and database programs are both adequate and convenient. In fact, such software is typically included in software packages, such as Microsoft Office.

The convention for entering data into a spreadsheet is that each row represents a person (a participant in the evaluation) and each column represents one variable. A variable in this situation is one survey question—that is, one item of data. Each column must contain no more than one discrete, distinct question. If a question contains several items with yes or no responses, then each item is actually a question. In this way, the number of yes or no responses for each item can be counted. Setting up the spreadsheet is a crucial component of data management.

If the evaluation plan calls for outsourcing the data analysis, entering the data into a standard spreadsheet is recommended. The data files can then be read by more sophisticated statistical software. The commercially available statistical software programs, such as SPSS, Stata, and SAS, have become increasingly user friendly and are marketed widely to larger organizations interested in ongoing evaluations and data management. These statistical software packages include components that facilitate doing highly complex statistical tests used in research that may not be available in software intended for business.

Another choice is to use free statistical shareware, such as EpiInfo™. EpiInfo™ may be downloaded via the Internet from the Centers for Disease Control and Prevention (CDC). This software is particularly helpful for program evaluations in which relative risks and odds ratios need to be calculated. EpiInfo™ was designed to be used internationally and to be compatible with as many systems as possible. Its ease of use and inclusion of statistics needed for some health program evaluations can make it an appropriate statistical software choice.

Data management includes not only the choice of software but also the management of the flow of paper, data preparation, and the oversight of the electronic files. The paper flow issues include tracking data entry, storage of original paper questionnaires and consent forms, and destruction of paper records. A rule of thumb is that paper ought to be kept for as long as the evaluation is active and until the final report has been distributed. This rule also applies to the electronic files. Of course, creating backup files and having a standard procedure for naming the files are mandatory.

Options for online survey software include Red Hat, SurveyMonkey, and Qualtrics. These programs allow survey results to be downloaded in raw data formats, such as Comma Separated Values (CSV), eXcel Spreadsheet (XLS), and fixed field that are compatible with common data analysis software. Preparing these data for analysis can be time intensive and needs to be taken into account in planning the evaluation.

Data cleaning is a critical step that can take a noticeable amount of time and effort. This process involves checking the data for obvious data entry errors. Data cleaning is important because, without good data, the statistical results are meaningless. The need for data cleaning can be minimized by planning during the development of the instrument or questionnaire and by careful adherence to the data collection procedures. Data cleaning begins with reviewing the frequency distributions of all variables. First, look for values that do not seem reasonable or plausible, such as a participant's age being 45 when the program is for adolescents and is based in a regular high school, or a negative value for the number of days in a program that was calculated based on subtracting dates. If an unrealistic value is found, the next step is to review the data for the individual and determine whether (1) the data were incorrectly entered, (2) the value is plausible, or (3) the value is so unlikely that the data ought to be considered as missing. Keeping an incorrect value in the data can drastically alter the mean, standard deviation, and subsequent statistical tests.

The data also need to be reviewed for *skip patterns*—that is, systematic nonresponses to items on a questionnaire. If specific items have low response rates, then their use in subsequent analyses should be questioned. As with unreasonable values, including an item with a low response rate has implications for the subsequent statistical analysis. Including items that have missing data decreases the total number of respondents included in any analyses that incorporate that item, which can result in distorted and unstable statistics.

Outliers

Outliers are those variables with reasonable, plausible, yet extraordinary values; they lie outside the normal or at the extreme ends on the tails of a distribution curve. Outlying values can occur as a result of errors in the measures or instruments, data entry errors, or unusual but accurate data. Common examples of outliers in health care include the rare patients who incur extremely high hospitalization costs, as would be the case for a very low birthweight infant and for the long lengths of stay for individuals with a rare but serious illness or with complications from a procedure.

Authors of statistical textbooks warn of the effects on statistical results when outliers are included in the data analysis (e.g., Kleinbaum, Kupper, Muller, & Nizan, 2013). Outliers dramatically influence the results of statistical tests by shifting the mean and increasing the variance. Although complex statistical methods can correct for the effects of outliers, these methods are more complicated than necessary for most program evaluations. Evaluators need to decide whether to keep the outlier in the analysis or to exclude the outlier based on some defensible rationale. The rationale for exclusion is often based on determining some cutoff point for the values to be excluded.

The decision to include or exclude an outlier is made on a case-by-case basis. One factor that can influence the decision is the sample size in the evaluation study. Statistics based on

smaller samples are more dramatically affected by outliers. For example, an evaluation of change in HbA1c levels might be studied in 20 people. If one person had a decrease of more than 50% in this level while everyone else had a decrease of between 0% and 10%, the average percentage decrease will be larger than if that one individual were excluded from the analysis. In contrast, if data from 200 people are collected, with the same range in decrease, the one individual with a 50% decrease in HbA1c level will have less effect on the average decrease.

Linked Data

Linked data refer to a data set that results from merging data from more than one source, so that a more comprehensive set of variables becomes available for the subjects in the data set. Recall that, for health programs, individuals, neighborhoods, or states could be the subjects of an effect evaluation. The types of data that are linked for health program evaluation can therefore include survey data with survey data, vital records data with survey data, vital records data with population surveys, survey data with administrative data, or population survey data with population survey data. Use of linked data may be necessary if the evaluation question focuses on outcomes on which data exist from a variety of sources. Linked data can be helpful throughout the stages of program planning and evaluation, from community assessment to effect evaluation.

One reason to use linked data is to connect program participation to outcome data. The study by Reichman and Hade (2001) is an example of linking participation data with outcome data; these authors matched a list of participants in a prenatal program with birth data from vital records to evaluate the outcome of the program. Another reason to use linked data is to have program outcome data associated with services utilization (process) data. For example, Meuleners, Hendrie, and Lee (2008) linked mortality data, hospital data, and data from mental health information systems to study readmission rates

for persons who had experienced interpersonal violence. A third reason to use linked data is to validate self-report responses. For example, a systematic review found low to moderate correspondence between adults' self-reported physical activity and direct measures, such as accelerometry (Prince et al., 2008). This strategy often leaves the evaluator with the dilemma of which data to believe and use.

The basic steps involved in linking data are simple, although their implementation is often far from simple. First, the data sets to be merged need to be in compatible software files. As software has become more standardized, it has become easier to create data files that are compatible. Nonetheless, software compatibility must be checked before beginning the linking process. Second, at least some variables about the individuals must be the same in both data sets. In other words, there needs to be a set of variables—called matching variables—that are the same in both data sets, and those variables must relate to only one individual. Thus, each file needs to be checked for having the matching variables. Third, matching variables are used as the criteria for linking the data from each file and merging the data into one file.

Two major issues arise when using linked data: confidentiality and accuracy. For the majority of health program evaluations, data about individuals are linked. This effort requires that some unique identifiers of the individual exist in both data sets, such as date of birth, Social Security number, or medical record number. Having data that identify individuals with their data can raise ethical concerns, however. As a consequence, there must be strict, comprehensive, and careful procedures to remove the unique person identifiers after completing the merge of data files. Given legal efforts to protect confidentiality, breaching legal and ethical guidelines can have dire negative consequences. Accuracy can be an issue in terms of correctly linking the files, so that all data for a specific person are really about that person. Achieving accurately linked files requires using a complex

algorithm for matching variables. For example, if there are two “Mary Smith” entries in the files, birth date and marital status may be needed to distinguish between the two Mary Smiths and thus link the files correctly.

► Sample Description

Once the data are clean, the statistical analyses can begin. Always begin the analysis with a careful examination of the sample or samples by reviewing frequency statistics on each group (participants only, control group only) for any obvious unexpected differences. If the frequencies appear to be as expected, the evaluator can proceed to statistical comparisons. If any frequencies are not as expected, the data should be reviewed more carefully. For example, if the mean age for the participants looks considerably higher than for the control group, use a comparison test to assess whether there is a statistical difference for the two groups.

As a general rule, if an experimental/exposed group and a control/unexposed group were used in the design, the evaluator should begin with a statistical comparison of the participant and the control groups on basic demographic variables. This step can be important as a means to convince stakeholders and others that the subsequent differences were not related to demographic differences. In other words, if no statistically significant differences are found, then it is safe to say that the participants were similar to the population from which they were selected. This speaks to the generalizability of the results and the external validity of the evaluation. If statistically significant differences between the participant and control groups are present, acknowledge and discuss possible reasons for the differences. This action speaks to the trustworthiness of the evaluator. It may be important to consider using the statistically different demographic factors in subsequent analyses, as moderating factors, to diminish their influence on the statistical findings about the effect of the program.

► Thinking About Change

In terms of program effect evaluation, *change* is measured as a difference. In many ways, change is a relative term because it is detectable only by comparison. The amount of change detected is influenced by what is being compared—a fundamental characteristic of change that must be understood by evaluators, program staff members, and stakeholders alike. Also, connecting the measure of change back to the outcome objective target values is important.

Change as a Difference Score

Change generally is measured three ways. The first way to measure change is to subtract an initial baseline score before the intervention from a subsequent score after the program. The term *score* is used generically to refer to a measure of the health outcome, whether it is a lab value, a cognitive test score, or a health services utilization rate. The first difference score is calculated with the following basic formula:

$$\text{Amount of change} = \frac{\text{sum (each posttest score} - \text{each pretest score)}}{\text{number of paired scores}} \quad (1)$$

To calculate change using this formula requires having data from only one group, making it the only option for evaluations that used a one-sample pretest/posttest design. For each individual, a difference score is calculated and then averaged across all the participants.

The second way to measure change is to subtract the mean score of the group who received the program from the mean score of the group who did not receive the program:

$$\text{Amount of change} = (\text{mean participants' posttest score}) - (\text{mean nonparticipants' pretest score}) \quad (2)$$

To calculate change using this formula requires having data from both the experimental group

and the control group after the intervention has been received. Given that a nonequivalent-group, posttest-only design is weak, however, this formula is seldom used.

Designs that include two groups and both pretest and posttest data are stronger and provide additional data for estimating the program effect. If data have been collected from both the experimental group and the control group before and after the intervention, as would be the case in some quasi-experimental designs and true experimental design, formulas (1) and (2) can be combined:

$$\begin{aligned} \text{Amount of change} = & (\text{mean participants' posttest score} - \\ & \text{mean participants' pretest score}) - \\ & (\text{mean nonparticipants' posttest score} - \\ & \text{mean nonparticipants' pretest score}) \end{aligned} \quad (3)$$

This formula gives a more precise measure of change because it uses all the relevant data. First calculating the amount of change in participants and in nonparticipants is especially important for health programs that have a longer duration and for intended audiences that experience natural changes over that time period. The remaining difference is likely due to the program. A between-group mean change is essentially determined by this third formula and is often used to report the findings of clinical trials. For example, Eser, Yavuzer, Karakus, and Karaoglan (2008) used this method to report the effect of balance training in a randomized trial.

Issues with Quantifying Change from the Program

Before evaluators can determine the program effect, four conceptual challenges to understanding change must be addressed. In this section, each challenge is explained, and possible approaches to resolving the challenge are presented. These challenges may influence the interpretation of the statistical findings.

Direction of Directed Outcome: Increase or Decrease

The direction of the change becomes important for selecting appropriate calculation techniques, interpreting the statistical results, and presenting the findings. For example, in Layetteville, the adolescent pregnancy prevention program sought to decrease the birth rate for adolescents from a baseline of 37.6 births per 1,000 adolescent girls, whereas the congenital anomalies prevention program sought to increase the use of prenatal vitamins from a baseline of 80% to 98% use by pregnant women. Thus, when a change score is calculated for each of the programs, the result of subtracting the pre-intervention value from the post-intervention value will have different signs for the two programs. The decrease in adolescent pregnancy takes the form of a positive number, whereas the increase in prenatal vitamin use appears as a negative number. If a multicomponent program has objectives with targets that go in both directions, confusion is likely to ensue. There may be no way to avoid this problem, so taking care in the reporting and presentation may be the only path. One solution to this dilemma is to report the change as an absolute number. Alternatively, the findings may be characterized as phrases (i.e., “improvement”) in a way that gets around having a mix of positive and negative numbers, both of which reflect improvement.

High Level of Desired Outcome at Baseline or Pretest

One difficulty faced by population-focused programs is that behavioral change in the population follows the diffusion of innovation curve (Rogers, 1983). Innovations, as new and novel ideas or products, become adopted over time by a greater number of people. This process is referred to as *diffusion* of the innovation through the population. The difficulty created by diffusion is that, with a higher prevalence of the desired health behavior before the program, it becomes more challenging to increase the prevalence of the health behavior. This dilemma arises because

those who are the last to adopt a change are known to be the most resistant to change.

One approach for addressing this issue is to take into account the initial prevalence of the health or behavior when estimating the effectiveness of the intervention (Hovland, Lumsdaine, & Scheffield, 1949, as cited in Green & Lewis, 1986). To do so, the change in the percentage of a population with the behavior after the program is compared to the percentage of the population without the behavior before the program. Because it is more difficult to achieve an equal increase in a behavior when the baseline value is already high, this approach is weighted in favor of improvements in population with a high baseline. The result is an increasingly higher number as the pervasiveness of the desired health outcome increases at baseline. In general, a higher number is better, although it can be misleading. A major shortcoming of this approach is that it does not take into account the desired outcome level, as reflected in the outcome objectives.

Relationship of Change to an Outcome Objective Target Value

Mohr (1992) argued that the simple difference scores as given between the participants and nonparticipants or between pre-program and post-program scores or rates do not provide information on how effective the program was in terms of whether it was weak or strong. To overcome this shortcoming, Mohr proposed a ratio that captures the amount of change achieved in relationship to the amount of change planned as given in the objective target value:

$$\text{Effectiveness ratio} = \frac{\text{posttest score} - \text{pretest score}}{\text{target score} - \text{pretest score}}$$

The target score is the level that the program intended to achieve, as stated in the outcome objectives. Keep in mind that this formula is applied to one group at a time, so it does not compare experimental and control groups.

The formula is useful, however, in contrasting the program effect at two sites. The effectiveness ratio can also be useful in making revisions to the extent element of the TREW (Timeframe, what portion of Recipients experience what Extent of Which type of change) effect objectives.

Consider, for example, two clinics that are participating in Bowe County's program to increase folic acid intake among women of childbearing age, where the change in prenatal vitamin use is one outcome indicator. In Clinic A, after the counseling intervention, 70% of the women take the vitamins—an increase of 10%.

The effectiveness ratio for this clinic is .26 (TABLE 14-1). Clinic B has a baseline rate of 85% and also has an increase of 10% of women taking the vitamins, but its ratio is .77. Although both clinics had a 10% increase or improvement, Clinic B appears more successful because of the higher baseline rate. Although the two clinics had a 10% increase, Clinic A started much farther from the target value, so it has a much lower effectiveness ratio than Clinic B, whose baseline was closer to the target value. Of course, this ratio still does not help us understand the extent to which the program addressed the health problem.

TABLE 14-1 Calculation of Effectiveness and Adequacy Indices: An Example

	Bowe County		Interpretation
	Clinic A	Clinic B	
Outcome objective target value	98%	98%	Value established for the outcome objective
Percentage taking prenatal vitamins at baseline	60%	85%	Baseline or control values
Percentage taking prenatal vitamins post-program	70%	95%	Outcome values
Pre-program to post-program change	$70\% - 60\% = 10\%$	$95\% - 85\% = 10\%$	Simple amount of change in outcome variable
Effectiveness ratio	$\frac{70\% - 60\%}{98\% - 60\%} = .26$	$\frac{95\% - 85\%}{98\% - 85\%} = .77$	Ratio of actual to planned effect; reveals that the selection of the target rate influences interpretation of effectiveness
Intervention efficiency	$\frac{70\% - 60\%}{10 \text{ min} - 5 \text{ min}} = 2\% \text{ per minute}$	$\frac{95\% - 85\%}{15 \text{ min} - 5 \text{ min}} = 1\% \text{ per min}$	Gives a measure of improvement per unit of intervention effort (using minutes of counseling intervention)

Relationship of Change to Intervention Effort

The effectiveness of a program can also be thought of in terms of the ability of the intervention to generate change. Abelson (1995) proposed a formula that takes into account both the intervention effect size and the size of the factor that caused the effect—namely, the intervention effort or dosage. The effect size is the mean value of the outcome variable for the experimental group minus the mean value of the outcome variable for the control groups. The intervention causal size is the amount of the intervention each group received, generally thought of as dosage. The intervention efficiency score, or what Abelson calls the causal efficacy, uses both the effect size and the cause size:

$$\text{Intervention efficiency} = \frac{\left(\begin{matrix} \text{mean postscore} \\ \text{experimental group} \end{matrix} - \begin{matrix} \text{mean postscore} \\ \text{control group} \end{matrix} \right)}{\left(\begin{matrix} \text{amount of intervention} \\ \text{experimental group} \end{matrix} - \begin{matrix} \text{amount of intervention} \\ \text{control group} \end{matrix} \right)}$$

The intervention efficiency score provides insight into the amount of effort required to produce the amount of change observed. In Table 14-1, Clinic A has a higher intervention efficiency score (percentage increase per minute of counseling) compared to Clinic B because more intervention time was given in Clinic B for the same amount of increase in women taking prenatal vitamins. Ideally, an intervention theory that is evidence based and carefully aligns

interventions with desired outcomes should lead to a higher intervention efficiency index. But as the Clinic A versus Clinic B example shows, the intervention dosage (minutes)—and not the intervention type (counseling)—made Clinic B appear less efficacious. Only by examining the process evaluation data would the program evaluator or program manager be able to determine the source of the variation between the two clinics.

The intervention efficiency score is critical for linking the outcomes and the processes. **TABLE 14-2** shows a 2 × 2 matrix demonstrating how the intervention efficiency score gives useful planning information. As shown in Table 14-2, the ideal outcome would be to have large effects from small causes, and the least desirable outcome would be to have small effects for large causes. The intervention efficiency score is calculated using the effect evaluation data. However, during the planning stage, hypothetical data could be used, thereby providing some parameters for setting expectations. In any event, the intervention efficiency score would be one source of information for making decisions regarding continuing, modifying, or ceasing a program.

Unmeasured Factors Adding to the Difference

Calculating change as the difference between pre-program and post-program data gives a net program effect, which is the amount of intervention effect on participants compared to any effect on the comparison group, given the

TABLE 14-2 Intervention Efficiency as a Relation of Effect Size and Causal Size

		Effect Size	
		Small Effect	Large Effect
Causal Size	Small Cause	Neutral intervention efficiency	High intervention efficiency
	Large Cause	Low intervention efficiency	Neutral intervention efficiency

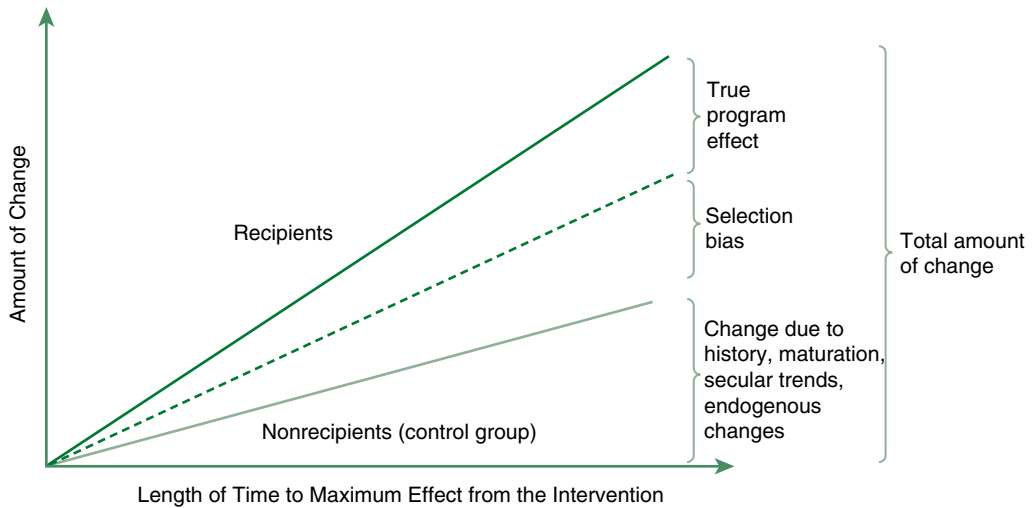


FIGURE 14-1 Contributing Factors to the Total Amount of Change

amount of error due to the design and measures used. Under ideal conditions, there would be no design or measurement error, so the amount of difference between participants and members of the comparison group would be the true amount of change from the program. Of course, real conditions make it more difficult to know the true and complete amount of change attributable to the program (**FIGURE 14-1**). Any of a variety of knowable and measurable but not controllable factors may influence the amount of difference or change that is statistically found, such as growth or aging and recent media reports. For example, media reports might influence study participants to change a behavior such that the change cannot be attributed solely to the program. In addition, unknowable or unanticipated, and hence unmeasured, factors can influence the amount of difference or change that is statistically found, such as sudden disasters, epidemics, and policy changes. Clearly, statistical findings are only as accurate and trustworthy as the design and measurement allows.

Clinical and Statistical Significance

The presence of statistical significance does not necessarily equate to practical or clinical

significance. An excellent example of this phenomenon is the statistically significant increase in birthweight frequently found in a variety of prenatal programs. The amount of additional birthweight is often in the range of 5 to 10 grams, which is a very minor—not clinically important—increase, except for very small newborns. *Statistical significance* indicates how likely one would be to get the result by pure chance, whereas *clinical significance* relates to how likely the intervention is to have a noticeable benefit to participants. Effect evaluations of health programs ideally seek to establish both the statistical and clinical significance of the program. Statistical significance may not translate directly into practical importance, as is the case in making programmatic or policy decisions. For these reasons, it is important to distinguish between significance tests and tests that indicate the degree of effect from the program, although both are often meaningful to program stakeholders.

► Across Levels of Analysis

Health programs can be designed to produce effects across the various levels of the public health pyramid. Effect evaluations of programs

designed for and delivered to individuals yield the most straightforward data for analysis. When programs are designed and delivered to a unit of individuals, such as a family, a school within a district, or a work unit within an agency, the effect evaluation will have data collected about that unit; items on a questionnaire will ask about things that happen “in our family,” “in my school,” or “in my department.” Analysis of data collected from individuals about a unit must take the unit into account, which is accomplished by aggregating the data. Aggregation means summarizing data from across participants within one nested unit to create a variable at the unit level of analysis. The advantage of using aggregated data is that a different pattern exists for aggregates. Thus, the evaluator can describe the characteristics of the aggregates, compare results across aggregates rather than across individuals, and identify associations between characteristics of aggregate and program or other variables. Epidemiological patterns can be more noticeable when comparisons are made across aggregates. For example, comparing data about different clinics, schools, or work sites may provide more useful information than focusing on individuals within those units.

Although aggregation of data makes sense for public and community health program evaluation, the notion of not maintaining an individual level of analysis may be uncomfortable or unfamiliar to clinicians who are trained to rely on data about each individual. Neither aggregated nor individual-level analysis is right or wrong. Instead, each results in different information about program effects.

If the evaluation is focused on an aggregate, such as a family, community, population, or work unit, the first step in aggregation is to analyze the data from the perspective of the within-unit variable. If those within the unit are more like one another than they are like the rest of the sample—in other words, if there is a low amount of variability among those within the unit—then it is acceptable to create a score per unit by aggregating the unit members’ data to form a unit-level variable. The unit is then statistically considered as though it were a single participant.

Various statistical tests, such as the intraclass correlation (ICC) or the eta-squared value, are used to determine the validity of aggregating the data (Burke et al., 2005). Alternatively, some statistical software may use data from individuals for analysis across units and analyses of units within units, such as students in classrooms in schools in districts. Use and interpretation of these complex statistical tests require statistical consultation and guidance. The important point is that statistical procedures for dealing with and analyzing aggregated data exist and may be appropriate for effect evaluations that are conducted for an evaluation research purpose.

If the analysis focuses on a health outcome and a characteristic of a unit (i.e., a school, department, or community), then analysis might potentially find a higher correlation between the units than among the individuals. This outcome is known as an *ecological correlation*. Black (1999) stressed that these ecological correlations are important to consider in terms of the intent of the research. The correlation between units is acceptable as a finding if the intervention was aimed at the unit and the outcome being assessed occurred at the unit level. Conversely, if the evaluation was intended to identify changes in participants across units, then the analysis should remain at the individual level.

In summary, statistical analysis of effect evaluation data follows all the conventions of statistical analysis of research data. Evaluators need to be cognizant of the interplay between how the health program is delivered, how the evaluation data are collected, and which statistical tests are carried out. A word of caution is in order, however. If large amounts of data have been collected and the statistical tests first used to assess for a change revealed no program effect, there is an inclination to begin data dredging. Data dredging is the process of continuing to analyze the data in ways not originally planned—in other words, looking and looking and looking until some significant finding is found. Data dredging eventually yields statistically significant findings, if only by chance. Therefore, unnecessary additional statistical tests ought to be avoided.

► Statistical Answers to the Questions

Before beginning statistical analysis, program evaluators should review the program outcome objectives and the questions for the effect evaluation. The questions help keep the analysis focused, thereby minimizing evaluators' inclination to search for significant findings (i.e., data dredging). Staying close to the evaluation questions also ensures that the key concerns of stakeholders are addressed first. Unfortunately, the process of choosing the best statistical procedure and tests can quickly devolve into a guessing game. This review is not intended as a comprehensive review of basic statistical principles; rather, it emphasizes the practical relationship between the evaluation questions and the statistical tests performed. The purpose of this section is to provide guidelines for choosing statistical tests and decreasing the amount of guesswork involved. Given that spreadsheet and database software greatly facilitates doing the mathematical calculations, it is imperative to have a framework for choosing the best and most appropriate statistical test.

TABLE 14-3 lists a set of questions that need to be answered to arrive at an appropriate statistical analysis plan. **FIGURE 14-2** shows the factors that

determine a statistical approach. For the purposes of this discussion, the focus of analysis refers to whether the effect evaluation is seeking to answer questions about comparison of groups, associations between two or more variables, or prediction of outcomes. This review of statistical data analysis is organized by those three dominant analytic approaches used in effect evaluations.

Using the level of intervention and level of analysis as two dimensions, the possible foci of analysis can be assigned. **TABLE 14-4** is based on the assumption that the evaluation data were collected at the same level as that of the program intervention. A careful look at this table reveals that statistical tests of comparison and association can be used widely and that prediction tests are recommended only when the level of intervention and the level of analysis are the same, unless the help of a statistician is available. In addition, analyses cannot be done at levels where the data are more discrete (lower) than the level at which the programmatic intervention was delivered—hence the cells labeled “Not appropriate” in Table 14-4.

When discussing the types of statistical analyses that are appropriate for comparison, association, and prediction, a useful distinction is the difference among analysis procedure, measure of magnitude, and test for significance. This distinction is rarely emphasized in statistical

TABLE 14-3 Factors That Affect the Choice of a Statistical Test: Questions to Be Answered

1. How many groups did the design result in: one, two, or more? Were they paired/matched, independent, or a population?
2. What is the focus of the evaluation question: comparison, association among variables, or prediction of outcomes?
3. Which level of measurement was used for the dependent variable and for the independent variables: nominal, ordinal, interval/ratio? If interval/ratio measures were used, do the data have a parametric or nonparametric distribution?
4. What is the interest and capacity of the stakeholders with regard to understanding statistical analyses?

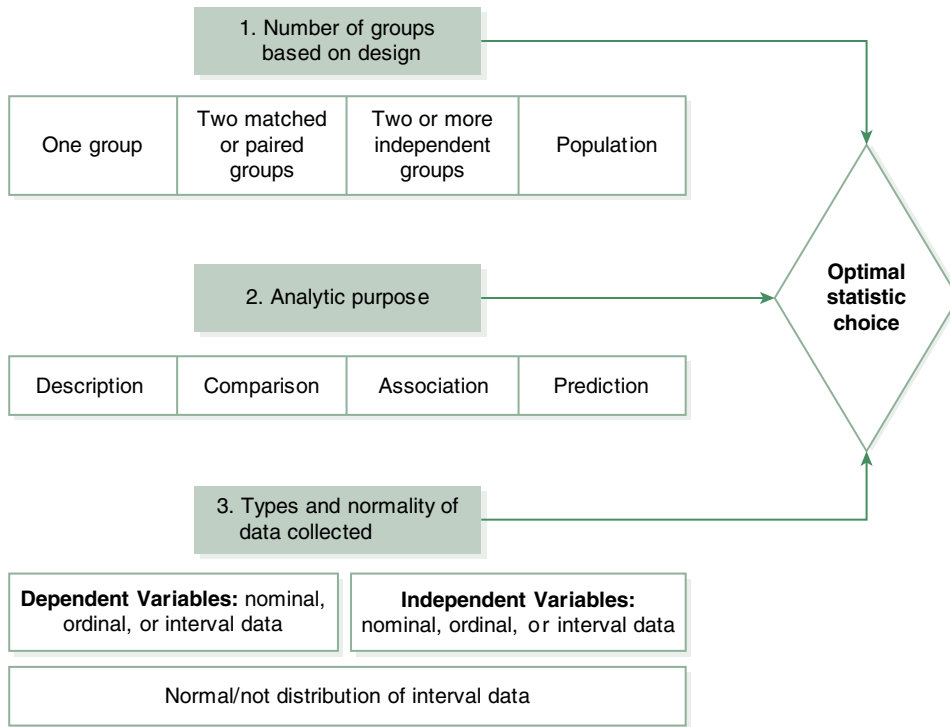


FIGURE 14-2 Summary of the Three Decisions for Choosing an Analytic Approach

textbooks. The essential question for effect evaluations is either How much difference did the program make? or How strongly was the program related to the effects? Measures of magnitude answer this question. In contrast, the question, Was the difference more than would happen by chance? is answered with tests of statistical significance. For each measure of magnitude, there is a corresponding test of significance. Both the magnitude and the statistical significance of results are relevant when evaluating programs (Rossi, Lipsey, & Howard, 2004).

Description

The first step in any statistical analysis focuses on description. *Descriptive statistics*, also called exploratory data analysis or univariate statistics, yields information of the most basic nature, such

as frequency counts, mean, mode, percentages, and dispersion of values. Do not underestimate the power and information contained in descriptive statistics (Tukey, 1977) or the ease with which most stakeholders can understand them. Descriptive statistics can be applied to all types of data, regardless of the design. These statistics are used to answer evaluation questions, such as, What were the demographic characteristics of program participants and comparison groups? How were the scores on the pretest and the posttest distributed? and What percentage of the program participants reached the target set in the objectives? If the level of measure involves ordinal and interval variables, descriptive statistics can include measures of central tendency (mean, mode, and median) as well as measures of variation (range, variance, and standard deviation). These

TABLE 14-4 Analysis Procedures by Level of Intervention and Level of Analysis (Assuming Data Collected at Same Level as Analysis)

Level of Analysis	Level of Program Intervention			
	Individual	Aggregate	Population	Infrastructure
Individual	Comparison tests, association tests, prediction tests	If individuals can be identified, then: comparison tests, association tests	If individuals can be identified, then: comparison tests, association tests	If individuals can be identified, then: comparison tests, association tests
Aggregate	Not appropriate	Comparison tests, association tests, prediction tests	If subgroups can be identified, then: comparison tests, association tests	If subgroups can be identified, then: comparison tests, association tests
Population	Not appropriate	Not appropriate	Comparison tests, association tests, prediction tests	Comparison tests, association tests, prediction tests
Infrastructure	Not appropriate	Not appropriate	Comparison tests, association tests, prediction tests	Comparison tests, association tests, prediction tests

statistics were reviewed elsewhere in the text, in the discussion of defining the health problem. Spreadsheet and database software programs can readily compute descriptive statistics, and the results can be displayed easily with the associated graphics package.

The central tendency, frequency distributions, and variation of variables form the foundation for further statistical analyses. The review of the variance needs to determine whether the amount of variance is similar in both the experimental and control groups. This information helps the evaluator select the appropriate statistical tests

for comparisons and correlations. Evaluators should spend time reviewing these descriptive statistics because the distribution of values often gives the first understanding about a program's participants and effects. Also, when the statistics for each variable are reviewed, evaluators may be able to obtain insights into unique characteristics of participants or unexpected distributions. Such observations may lead to additional and sometimes unplanned analyses.

Another reason to do descriptive analyses is to assess whether the data are normally distributed. Nonnormally distributed data,

TABLE 14-5 Commonly Used Parametric and Nonparametric Statistical Tests for Comparison, Association, and Prediction

Type of Data	Complexity of Question About Effect		
	Comparison	Association	Prediction
Parametric	Difference scores, <i>t</i> -tests of difference of means, variance analyses (Analysis of variance [ANOVA], Analysis of covariance [ANCOVA])	Correlations (Pearson's), hierarchical analyses	Time series, regression analyses, logistic regression analyses
Nonparametric	Chi-square tests based on contingency tables	Chi-square tests based on contingency tables, odds ratio, relative risk, other (i.e., sign test, Wilcoxon, Kruskal–Wallis)	Log-linear and probit regression analyses

including nominal and ordinal measures, can be used only in *nonparametric* statistical tests. When interval measures follow the normal distribution curve, *parametric* tests can be used. However, if the interval or ratio-level data are not normally distributed, nonparametric statistical tests must be used. **TABLE 14-5** summarizes the major nonparametric and parametric statistical tests that are used for comparison, association, and prediction.

Comparison

Comparison questions can ask for a within-group comparison—that is, determinations of whether baseline (pretest) scores are different from the follow-up (posttest) scores. Comparison questions are also appropriate for between-group comparisons—that is, determination of whether program participants are different from nonparticipants or from members of the control group. These comparison questions can be posed at the individual level of analysis as well as at the aggregate and population levels. **TABLE 14-6** summarizes the major types of comparison

analysis procedures that are appropriate for each level of analysis. Of course, which level of measurement is employed and whether the variables are parametric influence the final choice of statistical procedure.

Comparisons between groups using only nominal data (for both the outcome variable and the independent variables) are performed by using one of the various versions of the chi-square test. Whether the groups are related/matched and how many groups are compared (Black, 1999, p. 436) can make a difference in which statistic is optimal. Between-groups questions addressed with such analyses focus on some version of this question: Are participants statistically more likely to have characteristic *y* compared to nonparticipants? Some health program evaluation questions focus on comparing populations, such as those that seek to determine rates in different states of health behaviors derived from the Behavioral Risk Factor Surveillance System (BRFSS). These simple comparisons are made by using the same statistical tests used with individual-level data. However, because the population, rather than a sample, is used, slightly different equations are used.

TABLE 14-6 Main Types of Comparison Analyses Used by Level of Analysis and Assuming That the Variables Are at the Same Level of Measurement

Level of Measurement	Comparison-Focused Analyses		
	Analysis Procedures	Measures of Magnitude	Tests of Significance
Nominal by nominal data	Difference scores, chi-square tests based on contingency tables (i.e., McNemar; Fisher's exact)	Percent or mean difference, phi coefficient, Cramer's V , lambda	p -value
Ordinal by ordinal data	Median test, Mann–Whitney U test, Kruskal–Wallis test for three or more groups, sign test, Wilcoxon matched pairs signed rank test, Friedman two-way analysis for three or more groups	Lambda, uncertainty coefficient, Goodman and Kruskal's gamma, Somers' d , eta coefficient	p -value
Interval by interval data	t -test (independent samples), paired t -test (related samples)	Difference between means	p -value, confidence interval

Association

Most evaluations aspire to answer more than comparative questions. Questions about the relationship or association among variables are asked, such as whether receiving more interventions is related to a greater amount of change or whether the amount of change is associated with a specific characteristic of program participants. Correlational statistics or other statistics of association do not provide information on the temporal sequence of variables and therefore do not provide information about causation. Instead, *correlational* analyses indicate the strength of the relationship and whether the relationship is such that the variables vary directly or inversely. An inverse relationship between the variables, such as increasing age and decreasing tissue resilience is indicated by a negative correlation.

TABLE 14-7 provides a summary of the main tests of association. To choose the appropriate

statistical test, evaluators must consider the level of measurement needed, the parametric character, and the number of groups used in the analysis. **TABLE 14-8** provides an example of the tests of the strength of association that could be used with data at different levels of measurement. Using the antiviolence program to prevent adolescent deaths due to gunshot wounds as an example, the dependent variables are the ICD-10 code for the injury, the rank of the school in terms of the number of injuries, and the number of school days missed due to gunshot wounds; the independent variables are the race of students, the rank of the school in terms of standardized test scores, and the number of hours of anger management education. **TABLE 14-9** continues with the antiviolence example by showing how the design of the evaluation further affects the statistical analysis plan.

At the individual level, in addition to contingency table analysis procedures for nonparametric data, correlation analyses are possible

TABLE 14-7 Main Types of Association Analyses Used by Level of Analysis, Assuming That Variables Are the Same Level of Measurement

Level of Measurement	Association-Focused Analyses		
	Analysis Procedures	Measures of Magnitude	Tests of Significance
Nominal by nominal data	Fisher's exact for 2×2 table, chi-square or independent samples, McNemar or Cochran Q for related samples	Relative risk, coefficient of contingency, phi coefficient, Cramer's V , lambda	p -value, confidence intervals
Ordinal by ordinal data	Chi-square, Spearman rank order	Kendall's coefficient of concordance, Kendall's tau a, tau b, tau c, Somers' d , Spearman rank order coefficient	p -value
Interval by interval data	Multiple regression analyses	Pearson correlation coefficient, intraclass coefficient	p -value
Mixed	One-way analysis of variance (ANOVA) for nominal by interval data	Eta coefficient	p -value for F statistic

with parametric data. When the correlational analysis includes a copious number of variables, the likelihood increases that some pairs will be significantly related. Therefore, it is wise to lower the alpha value from .05 to either .01 or .001 to provide a more conservative statement about what was statistically significant. One approach to reducing the number of variables in the analysis is to exclude variables on a logical basis, such as not possibly being related (i.e., hair color and height).

Moderating and Mediating Variables

The causal theory for the program may include either moderating or mediating variables. If data on these variables were collected as part of the

effect evaluation, then those data can be used to assess whether their presence changes the strength of correlation between the intervention and the health outcome. Although inclusion of moderating and mediating variables can quickly complicate the statistical analysis, there is at least one simple way to use those data: A moderating or mediating variable can be used as a control variable in the correlation between the health outcome variable and the intervention variable. This approach yields a partial correlation that has been adjusted for the effects of the control variable.

Prediction

Questions about how much of an effect a programmatic intervention might have on

TABLE 14-8 Example of Statistical Tests for Strength of Association by Level of Measurement, Using Layetteville Adolescent Antiviolence Program

Independent Predictor Variable	Dependent Outcome Variables		
	Nominal	Ordinal	Interval
	(e.g., International Classification of Diseases, 10th revision [ICD-10] for injury)	(e.g., rank among schools in terms of number of injuries)	(e.g., number of students with emergency room admissions)
Nominal, categorical, discrete (e.g., race of student, group assignment)	Cramer's <i>C</i> (two or more groups), chi-square	Mann-Whitney, Kruskal-Wallis	<i>t</i> -test, ANOVA
Ordinal (e.g., rank among schools in terms of test scores)	Probit regression	Spearman's rho, Kendall's tau, Spearman rank order, eta coefficient	Linear regression
Interval (e.g., number of hours of anger management education)	Logistic regression	Pearson product <i>r</i>	Pearson's <i>r</i> ; multiple correlation coefficient, linear regression

individuals, aggregates, or populations are basically questions of causation. Causal questions are the most difficult to answer, despite the fact that most stakeholders want an answer to the most fundamental causal question: Did our program cause the health improvement? **TABLE 14-10** summarizes key statistics that can help predict future outcomes.

To answer causal questions, it is absolutely necessary to have data from a rigorous nonexperimental or true experimental design. In other words, answering causal questions is not only a matter of statistical analysis but one of design. Unless a design has been used that enables the evaluator to eliminate alternative causes and that provides a time line for the effects, the cause-and-effect relationship between the program

and the health outcome cannot be substantiated. The statistics used for causal evaluation questions are essentially the same as those used for assessing relationships and associations, except that the findings can be interpreted as causal rather than only as indicating relationship. If there was no randomization, then the statistical approaches yield association information only. However, because the statistical procedures for prediction forecast a trend, that trend is often interpreted as prediction.

One major variation on the basic correlational analysis is regression analysis, sometimes called trend analysis (Montgomery, Peck, & Vining, 2015). Regression analysis allows for prediction, in terms of extrapolation and interpolation, based on a best-fit line (Montgomery et al., 2015). As

TABLE 14-9 Examples of Statistical Tests by Evaluation Design and Level of Measurement, with Examples of Variables

Design	Nominal by Nominal	Interval by Interval	Interval by Nominal
	(e.g., ICD-10 for injury by race)	(e.g., number of students with emergency room admissions by hours of work anger management education)	(e.g., number of students with emergency room admissions by student race)
One group pretest/posttest only	Chi-square	Pearson correlation, logistic regression	Point biserial coefficient
Two groups independent	Chi-square, $k = 2$ tables	One-way analysis of variance (ANOVA)	Chi-square
Two groups related or matched (i.e., case control, cohort)	McNemar change test	t -test for related or matched samples	McNemar

TABLE 14-10 Main Types of Prediction Analyses Used by Level of Analysis, Assuming That Variables Are at the Same Level of Measurement

Level of Measurement	Prediction-Focused Analysis		
	Analysis Procedures	Measures of Magnitude	Tests of Significance
Nominal/categorical/discrete by nominal categorical/discrete data	Probit regression analyses	Correlation coefficient (r^2)	F -test
Ordinal by ordinal data	Trend analyses	Correlation coefficient (r^2)	F -test
Interval by interval data	Time series, regression discontinuity	Beta coefficient	F -test, confidence intervals
Nominal by interval data	Logistic regression	Entropy concentration	F -test

a tool, regression analysis answers evaluation questions, such as How much more improvement might occur with more intervention? and As participant characteristics 1 and 2 increase, how much change occurs in the health outcome? Regression analyses are based on the correlation of independent variables with the dependent variable and on the strength of the association among independent variables.

Time series analysis requires collecting data on the same sample, using the same measure at multiple time points. Such analyses, based on the multiple regression model, can be used with data, such as number of motor vehicle accidents per year, number of Medicare enrollees per month, number of infants diagnosed with ear infections per week, or, prototypically, closing value of the New York Stock Exchange on each day. If a health program was expected to have a long-term or delayed effect, then the time series analysis would provide insights into whether a change was observed in the pattern across time.

Many health program evaluation questions focus on changes across time rather than on change at a single point in time among participants. For example, for evaluations of full-coverage programs, for which no control group may exist, a change in the recipient population may be observable only across a longer time period. Typical across-time evaluation questions are, To what extent is there a decline in the health problem from year to year? and To what extent was there a change in the health outcome from before the program or policy was implemented to after it was established? To answer such questions, a longitudinal design for data collection is commonly used such that data are collected from the same individuals, on the same variables, at multiple points in time. Analysis of the repeated measures collected from the same sample becomes complex because an underlying association inevitably exists between the individual and the data; each data set is not independent of the subsequent data set because the same people provided the data. If the level of measurement is at the interval level, then ANOVA tests are appropriate.

► Interpretation

Depending on the sample size, evaluator's analytic resources, and stakeholder preferences, comparisons can be done with or without statistical test findings and any of the following: the program goals, the impact objectives, benchmarks from other programs, results of similar programs, professional standards regarding the health problem, and intuition or educated guesses. These less scientific but practical comparisons do provide insights into the relative success of the health program, and they may highlight general areas of success and inadequacy.

In any study, spurious findings and surprises can occur. A *spurious finding* is one that is incidental to the evaluation question or that is an artifact of poorly understood factors. Spurious findings are generally curiosities that can be discounted as the result of measurement error and random chance. In contrast, *surprise findings* are not related to the evaluation questions but cause one to pause. Surprise findings can either support the health program interventions or not; in some cases, they may lead to new descriptive insights. For example, if the evaluation measured x , but the finding is that the average value of x is considerably higher or lower than in the literature or than common sense would dictate, then the evaluator may be surprised. Surprises can be used as a basis for further exploration or for making revisions to the program or the evaluation plan. Surprises, while not common, are important and need to be valued and acknowledged because they can inspire additional questions and lead to new knowledge.

Four Fallacies of Interpretation

Green and Lewis (1986) suggest that four types of errors can occur when interpreting evaluation data. One fallacy is equating effectiveness with efficiency. *Effectiveness* is the extent to which results are achieved, whereas *efficiency* is a ratio of amount of effort or resources to the amount of effect achieved. Another way to think of this fallacy is in terms of the cost per unit of health

improvement. There may be a point of diminishing returns at which application of additional resources and efforts may not result in meaningful gains in effect. Here, the use of the intervention efficiency ratio becomes important: The issue is how much more is required in the organizational or service utilization plan to obtain each additional unit of effect from the program.

Another fallacy is assuming that a constant rate of progress or health improvement will or has occurred. The rate of change may be variable or sporadic for many reasons. For example, for population-focused services, such as mass-media awareness campaigns, the diffusion of innovation curve (Rogers, 1983) will be evident if population adoption of a new behavior is part of the effect theory. The rate of change may also vary depending on the characteristics of the recipient audience.

The assumption that ongoing improvement—that is, the inexorable forward movement of change—can be achieved is another fallacy. Program effects can be affected by time, such as relapse; lack of reinforcement or follow-up; memory failure; and program dropouts. Any one of these or similar factors can cause the health problem or condition to return. The extent to which it is possible to identify ongoing improvement can be influenced by the timing of data collection relative to the expected beginning of the effect. A lack of forward movement in the degree of change can occur in at least five ways; each can be visualized graphically (FIGURE 14-3). In Figure 14-3a, the change over time reflects a delayed change called a *sleeper effect*. In Figure 14-3b, the program effect is seen following the intervention, but then a gradual return to the baseline rate, known as *backsliding*, occurs. As shown in Figure 14-3c, anticipation of the intervention sometimes leads to adopting the behavior before the program has actually started (*trigger effect*). If factors influence the rate regardless of the program, this is reflected in *historical effects* (Figure 14-3d). A *backlash* can occur if the program is discontinued such that the long-term rates are worse than the rates before the program began (Figure 14-3e).

The fourth fallacy is underestimating the complexity of the change process. Behavioral change within individuals involves multiple stages (Prochaska, DiClemente, & Norcross, 1992) and is in turn influenced by multiple factors. Similarly, achieving physical changes through a health intervention can be complex, involving medications, procedures, and behaviors. Underestimating the complexity inherent in achieving change in the intended audience—whether that audience consists of individuals, families, communities, or organizations—can lead to oversimplification of the interpretation of the findings. In addition, what is noticed in the findings is affected by what was articulated in the program effect theory, especially the theory of intervention, partly because the effect theory guides the decisions regarding what is measured and when. Therefore, better causal and intervention theories that explain or predict the change and possible contributing factors to either maintaining the status quo or to changing are more likely to lead to an evaluation that can minimize the risk of falling prey to this fallacy.

Ecological Fallacy

The ecological correlation mentioned previously should not be confused with the ecological fallacy, which has been an issue in program planning and evaluation for more than 30 years (Milcarek & Link, 1981). The *ecological fallacy* is the assumption that a group characteristic applies to all individuals within that group. For example, if a group of participants had an average of a high school education, the ecological fallacy would be to assume that any member of the group had a high school education. This fallacy involves interpretation of the findings in a way that is not being completely truthful to the data. Perhaps the group of participants actually includes many persons with college degrees and many persons with only an eighth-grade education; the presence of both subgroups could yield an average of 12 years of school.

Johnson and Caron (2006) found that the ecological fallacy was used by members of the

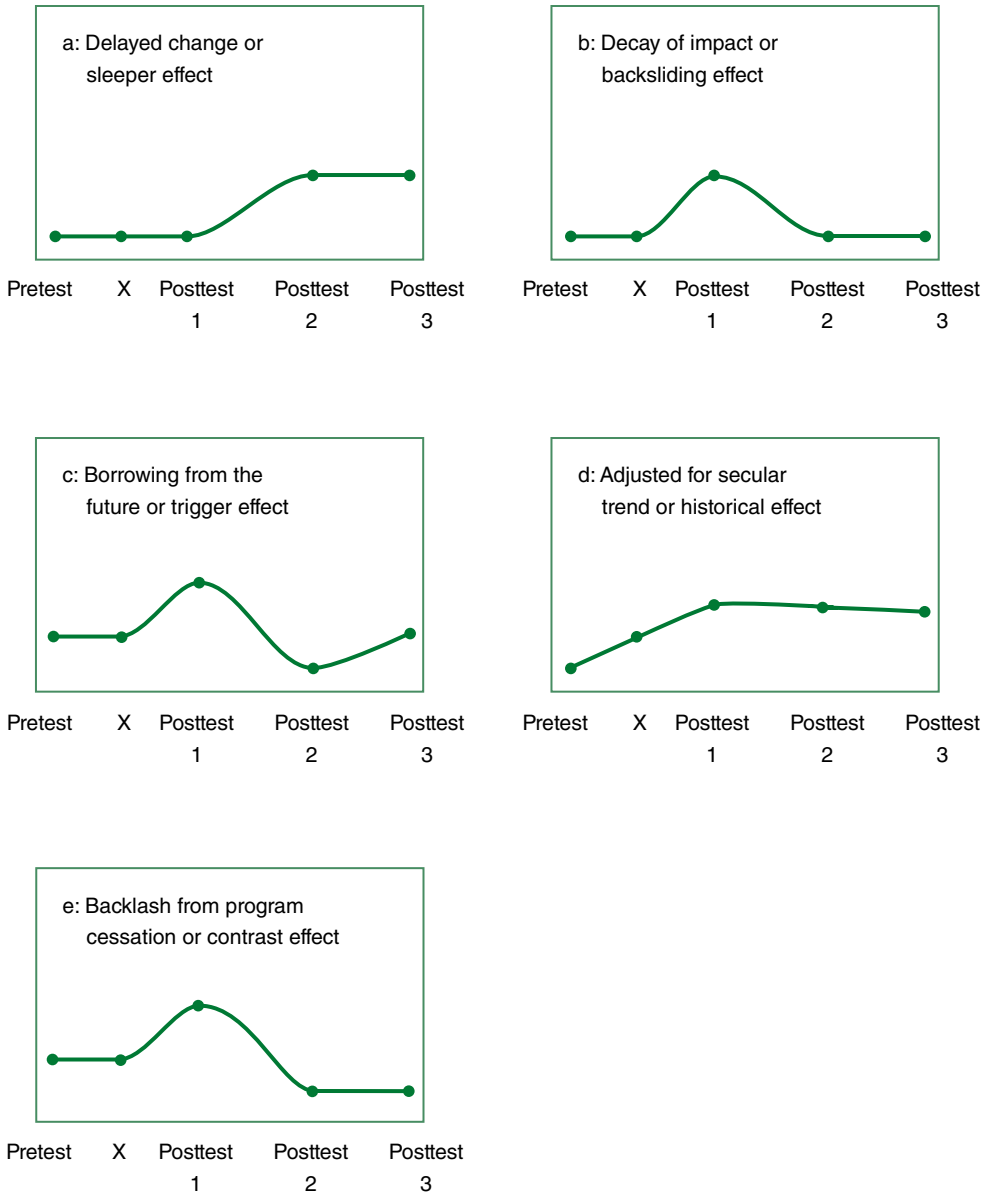


FIGURE 14-3 Five Ways That the Rate of Change Can Be Altered

Note: X indicates time of the intervention.

public in estimating the risks for their community based on state data. Their findings hint at a natural human tendency to draw erroneous conclusions based on the flawed logic of an ecological inference. Various statistical approaches, such as multilevel modeling, can be helpful in

understanding the extent to which a group characteristic truly applies to individuals within a group. Also, a move away from ecological or observational designs produces data that do not require making assumptions about individuals based on group characteristics.

► Across the Pyramid

At the direct services level of the public health pyramid, health programs focus on individual change. Therefore, most of the analysis revolves around identifying changes within subjects, such as pretest/posttest changes in program participants. Health programs at the direct services level are likely to be highly tailored and relatively small. The small number of participants should prompt concerns about having a sufficient sample size for statistical analysis. If the sample size is less than 30, the choice of statistical tests is affected. Because it may be easier to identify nonparticipants in the program (controls), however, statistical tests may be available that can compare those groups.

At the enabling services level, health programs focus on groups and changes evident in those groups. Issues related to levels of analysis are of major importance at this level of the public health pyramid. If data from individuals within units (whether families, schools, social networks, or agencies) are to be aggregated, then statistical tests for within- and between-group variance need to be performed. Most of the data analysis needs to focus on identifying differences between groups in the amount of change for each group.

At the population services level of the public health pyramid, it becomes more difficult to have nonparticipants. Therefore, it is more likely that statistical tests will involve trend analysis of population data or time series analysis. In some cases, it may be possible to compare populations, such as counties or states. Because the population mean is known, as is the mean of the sample, a different set of statistical tests becomes possible.

At the infrastructure level, the outcome evaluation is likely to have two major foci: the change among participants who are part of the infrastructure, and the change in the population before and after the infrastructure change. In the first case, the statistical tests chosen depend on the design of the program. For example, public health workers in several different counties may

receive the same training program. The analysis could occur at the individual level of analysis and compare program participants with nonparticipants. Alternatively, data from the workers might be aggregated to have a higher level of analysis (i.e., county), and participating counties would then be compared with nonparticipating counties. If the evaluation was concerned with the health status of those served by the infrastructure, a trends analysis could be carried out to identify whether a change occurred in the health status of county residents after the training program was finished. These are just a few examples of the interaction between the impact evaluation question, the design, and the statistical analysis.

► Discussion Questions and Activities

1. What are the pros and cons of aggregating data to have a higher level of analysis? Under which conditions would aggregation be the preferred approach to handling the impact evaluation data?
2. The adult immunization program of Bowe County has reached the effect evaluation stage. Data have been collected from individuals on their motivation to get immunized, use of the enabling service of transportation to the clinics has been tracked, and surveyors have approached people outside grocery stores about recollection of the mass-media campaigns. The evaluation team is now developing a plan for the statistical analysis. What would be your questions to the team about the data? Depending on the answers, which recommendations would you make?
3. It is common for stakeholders to ask, How can you prove that our program had an effect? What would be your standard, simple, sound-bite response to this question?

4. What is the importance of having effect size and significance, respectively? Which would be more important to subsequent decision making about continuing a program?
5. Evaluations of the programs in Layetteville and Bowe County could potentially benefit from having linked data. For at least two health programs, identify at least two databases for linking the data and give a brief outline of how you would link the data.

► Internet Resources

Have fun with statistics—really! The websites identified here can help make statistics real, and statistics can sometimes be easier to understand when you see the numbers change.

Seeing Statistics

The Java applets at the Seeing Statistics webbook (<http://www.seeingstatistics.com/>) are, as the developers put it, both “easy” and “fun.”

Statistical Education Through Problem Solving (STEPS)

The STEPS webpage (www.stats.gla.ac.uk/steps/glossary/) gives brief definitions of statistical tests.

Steve’s Attempt to Teach Statistics (StATS)

If you prefer a practical approach to statistics, try Steven Simon’s Statistical Koans (www.pmean.com/06/StatisticalKoan1.html). His newsletter makes statistics fascinating (www.pmean.com/archive12.html).

Rice Virtual Lab in Statistics

From this homepage (http://onlinestatbook.com/stat_sim/index.html), you can access links to applets that allow visualizing what happens when you take different samples from your data. You will need to have Java enabled.

Decision Tree for Statistics

The online decision tree for making statistical choices found at www.microsirris.com/Statistical%20Decision%20Tree/ takes you through a series of decision questions. The following websites provide alternative visual flowcharts for choosing a statistic: http://www.muhlenberg.edu/media/contentassets/pdf/academics/psychology/stats_decision.pdf, or go through a series of questions at this page, <http://www.insilicase.com/Web/StatPicker.aspx>, and access a downloadable app of the same series of questions.

GraphPad.com: Intuitive Biostatistics

This website clearly links the number of groups to some of the possible statistical tests: www.graphpad.com/www/book/Choose.htm.

References

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum.
- Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics*. London, England: Sage.
- Burke, S. M., Carron, A. V., Patterson, M. M., Estabrooks, P. A., Hill, J. L., Loughhead, T. M., Rosenkranz, S. R., & Spink, K. S. (2005). Cohesion as shared beliefs in exercise classes. *Small Group Research*, 36(3), 267–288. doi:10.1177/1046496404267942
- Eser, F., Yavuzer, G., Karakus, D., & Karaoglan, B. (2008). The effect of balance training on motor recovery and ambulation after stroke: A randomized controlled trial. *European Journal of Physical and Rehabilitation Medicine*, 44, 19–25.
- Green, L. W., & Lewis, F. M. (1986). *Measurement and evaluation in health education and health promotion*. Palo Alto, CA: Mayfield.
- Hovland, C. F., Lumsdaine, A., & Scheffeld, F. D. (1949). Experiments on mass communication. New York, NY: John Wiley. In L. W. Green & F. M. Lewis (Eds.) (1986). *Measurement and evaluation in health education and health promotion*. Palo Alto, CA: Mayfield.
- Johnson, B. B., & Caron, C. (2006). Evaluating public responses to environmental trend indicators. *Science Communication*, 28, 64–92. doi:10.1177/1075547006291346

- Kleinbaum, D., Kupper, L., Muller, K. E., & Nizan, A. (2013). *Applied regression analysis and other multivariable methods* (5th ed.). Boston, MA: Duxbury Press.
- Meuleners, L. B., Hendrie, D., & Lee, A. H. (2008). Hospitalisations due to interpersonal violence: A population-based study in Western Australia. *Medical Journal of Australia*, *188*, 572–575.
- Milcarek, B. I., & Link, B. G. (1981). Handling problems of ecological fallacy in program planning and evaluation. *Evaluation & Program Planning*, *4*(1), 23–28. doi:10.1016/0149-7189(81)90050-1
- Mohr, L. B. (1992). *Impact analysis for program evaluation*. Newbury Park, CA: Sage.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). *Introduction to linear regression analysis*. Hoboken, NJ: John Wiley & Sons.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Gorber, S. C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *5*(1), 56. doi:10.1186/1479-5868-5-56
- Prochaska, J., DiClemente, C., & Norcross, J. (1992). In search of how people change: Applications of addictive behaviors. *American Psychologist*, *47*, 1102–1114. <http://dx.doi.org/10.1037/0003-066X.47.9.1102>
- Reichman, N. E., & Hade, E. M. (2001). Validation of birth certificate data: A study of women in New Jersey's Healthy Start program. *Annals of Epidemiology*, *11*, 186–193. [http://dx.doi.org/10.1016/S1047-2797\(00\)00209-X](http://dx.doi.org/10.1016/S1047-2797(00)00209-X)
- Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). New York, NY: Free Press.
- Rossi, P. H., Lipsey, M. W., & Howard, E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

CHAPTER 15

Qualitative Methods for Planning and Evaluation

Elsewhere in this text, the approach to health program planning and evaluation has been essentially quantitative in approach. Numbers are used in the needs assessment, budgeting and documenting the implementation of the services utilization plan, and quantifying the amount of effect on individuals receiving the program. However, stories are equally valid and important in health program planning and evaluation. This chapter focuses on qualitative designs and methods. Much has been written on ways to conduct qualitative research and on strategies to ensure rigor in such research (Lincoln & Guba, 1985; Miles, Huberman, & Saldaña, 2014; Strauss & Corbin, 2014), and there is no pretense of covering all that is currently known about qualitative research. As with the chapters reviewing quantitative methods, this chapter reviews qualitative methods, with particular emphasis on the use of these methods for health program planning and evaluation.

► Qualitative Methods Throughout the Planning and Evaluation Cycle

Data are collected throughout the planning and evaluation cycle. The use of qualitative methods is an option for that data collection. During the assessment stage of the planning process, qualitative methods are suitable for gaining insights into the needs, problems, barriers, and issues faced by the target population. Also, during the assessment and planning stages, qualitative methods yield data essential to develop and delineate program elements before initiating a program. Similarly, qualitative methods can help generate the program theory, especially the effect theory, so that it incorporates the explanatory models of the intended audience and

the theory-in-use of program staff members if a program is being revised. During the process monitoring phase, qualitative methods can take the place of quantitative methods, particularly for sensitive programs that function at the direct services level of the public health pyramid, and can broaden the observational field in terms of adding new types of information. During the effect evaluation phase, qualitative methods can enhance the explanatory power of the evaluation design and data. Also, by providing analyses of processes and individual cases, qualitative data can help explain why or how outcomes occurred. In other words, information from the qualitative data can be used to improve the effect theory on which the program was based.

Qualitative approaches give a voice, both literally and metaphorically, to stakeholders. This effect is important because it makes stakeholders feel valued. The act of speaking to someone and telling a story generates feelings that are substantively different from the feeling generated by responding to a survey questionnaire.

► Qualitative Methods

The first challenge faced by evaluators and program planners is to select the best qualitative method to answer the questions at hand. Qualitative methods have emerged from a variety of philosophical perspectives, each of which has influenced the development of the associated methodology. These underlying perspectives influence which evaluation questions might be answered from that perspective (**TABLE 15-1**). Thus, the choice of a qualitative method is based on whether it is the best method given the question being asked and with the understanding that several methods might be applicable to more than one perspective (**TABLE 15-2**). The following sections review the major and most widely used qualitative methods, with particular attention being paid to how each method is useful in health program planning and evaluation.

Across the qualitative methods, a major concern is the amount of time it takes first to gather

TABLE 15-1 Comparison of Qualitative Perspectives with Regard to the Basic Question Addressed and the Relevance to Health Program Planning and Evaluation

Perspective	Basic Question Addressed	Planning and Evaluation
Content analysis	Which themes are in the text?	Thoughts and perspectives revealed in the text and dialogues
Critical analysis	How has power shaped it?	Participants' views of their ability to be in control of the health problem and the solutions; staff members' view of their autonomy in improving the program
Ethnography	What are the norms and values (culture)?	Participants' cultural forces that contribute to the problem and acceptance of the program
Grounded theory	What are the relationships (theory)?	Explanations that participants and staff members have for the health problem and possible solutions
Phenomenology	What does it mean to the person?	Participants' meaning of content and the problem being addressed

TABLE 15-2 Comparison of Major Qualitative Perspectives with Regard to the Method Used

Perspective	Typical Methods
Content analysis	Focus groups, surveys with open-ended questions, narrative designs
Critical analysis	Individual in-depth interviews
Ethnography	Case study, participant observation, observations
Grounded theory	Individual in-depth interviews
Phenomenology	Individual in-depth interviews

and then to analyze the data. The nature of the qualitative methods is such that the amount of time necessary to interview, observe, or interact to gather the data tends to be longer than the amount of time required to gather data via a survey or other quantitative methods typically used in program planning or evaluation. In addition, the iterative and intensive data analysis process is longer than quantitative data analysis. These factors make qualitative methods more costly to use than a survey method. **TABLE 15-3** summarizes the benefits and challenges for each of the qualitative methods discussed here.

Individual In-Depth Interview

The most widely known and used of the qualitative methods is the in-depth, open-ended questioning of an individual. This involves using questions that require more than a yes or no response; encouraging explanation; and

allowing for storytelling that leads to information that is detailed, personal, and often reflective. This type of data is preferred when meanings and attributions are the focus and when new or fresh insights into a poorly understood phenomenon are sought. Typically, individuals who are interviewed are purposively selected based on a narrow set of criteria. Interviewers may be program recipients, stakeholders in the program, or members of the intended audience.

In-depth interviews can be used throughout the planning and evaluation cycle. In-depth interviews can also be used to understand the effects of the program. For example, Strachan et al. (2016) used this approach to conclude that positive prior experiences, awareness of malaria prevention benefits, and community social support encouraged use of long-lasting insecticidal nets in Uganda.

Data Collection

Conducting in-depth interviews requires skill on the part of the interviewer. The interviewer needs to be sensitive to cues of the interviewee about when to stop the interview as well as when and how to ask for clarification when the interviewee is either using terminology the interviewer does not understand or otherwise not providing an easily interpreted story. The ability and skill of the interviewer are critical to conducting a “good” interview that gathers sufficient details to make analysis less complicated yet has not disrupted the flow of ideas from the interviewee.

In-depth interviews can be done almost anywhere that both the interviewer and the interviewee feel comfortable and secure in terms of maintaining confidentiality. Whenever possible, people should not be interviewed in the presence of anyone who has power over them. This includes both frontline staff relative to managers and patients relative to providers. They may also need to be interviewed in a place that feels safe and private.

Sometimes the interviewee is referred to as a *key informant*, denoting that the individual is unique in possessing information of a specific

TABLE 15-3 Summary of Key Benefits and Challenges to Using Qualitative Methods in Planning and Evaluation

Method	Key Benefits	Key Challenges
Case study	Allows for understanding of context as an influence on the program or participant	Complex, overwhelming amount of data: definition of case
Observations	Can identify sequence of causes and effects; may identify new behaviors or events	Difficult to obtain reliable data unless recording devices are used; sampling frame difficult to establish
Individual in-depth interviews	Provides rich insights into personal thoughts, values, meanings, and attributions	Identifying individuals who are willing to be open
Focus groups	Inexpensive given the amount and type of data collected; get collective views rather than individual views	Need training in managing the group process; need good data-recording method
Survey with open-ended questions	Very inexpensive method of data collection	Poor handwriting and unclear statements can make data useless; often underutilized
Narrative designs	Very inexpensive methods of data collection; provide insights into social and cultural influences on thoughts and actions	Require special training in data analysis; may not have credibility with stakeholders; difficult to select the text most relevant to the health problem or program

nature. Key informants are more likely to be useful during the community assessment and planning stages because they would have knowledge of system processes and barriers to services and can draw on a range of experiences to comment on any generic issues. Data collected from in-depth interviews with key informants are analyzed in the same manner as data collected from any other interviewee.

Written Open-Ended Questions

Sometimes surveys include an open-ended question or two to which respondents can provide qualitative information. Typically, questions,

such as *Anything else? What was the best/worst? and What suggestions do you have?* are used. Such questions can be added to community assessment, process monitoring, or impact survey questionnaires. Because responses to such questions are generated by the individual rather than by the researcher, the responses are of a qualitative nature.

The major limitations associated with collecting qualitative data with written open-ended questions are that people often do not write legibly, are not articulate, or provide extremely brief responses. These factors can make it impossible to interpret their responses, rendering their data useless. Giving more explicit

instructions and providing sufficient space to write detailed responses can help minimize these problems. However, if the information from the questions is considered crucial, then other methods for collecting the data should be considered. Nonetheless, open-ended questions on a survey can be beneficial. Such questions, compared to a survey of close-ended questions, provide respondents with an opportunity to express themselves; as a consequence, they are likely to feel “heard.” Also, the response may provide sufficient information to generate an initial set of categories for a close-ended question. Ultimately, it is the discretion of the evaluator to decide whether the data are analyzable and useful. The analysis steps described later in this chapter can be applied to this rather limited but potentially fruitful set of qualitative data.

Focus Group

The *focus group* method of collecting qualitative data involves conducting an interview with a group of individuals, thereby taking advantage of the group dynamic, which itself can lead to discussions and revelations of new information. For some individuals, a group setting for interviews is less intimidating. By participating in focus groups, individuals within the group can confirm or disconfirm the experiences and views of others. This “instant feedback” aspect contributes to the focus group method being somewhat more efficient than in-depth individual interviews, which require a separate interview for validation of interpretations.

Data analysis follows the general steps used for analyzing in-depth interviews. Thus, those familiar with that data analysis procedure will be capable of analyzing focus group data. The difference is that the findings are not person-specific but are considered already validated during the group process.

The benefits of the focus group method have led to its use in health program planning and evaluation. For example, Esangbedo et al. (2016) concluded from a combination of interviews and focus groups with healthcare workers

in Nigeria that many had factually unfounded fears of the Ebola virus vaccine. Such information would help the program planners in developing an effective intervention. Cabassa, Contreras, Aragón, Molina, and Baron (2011) used focus groups to evaluate outcomes from an educational program regarding depression that was delivered as a telenovela. They found that the mode of delivering the education was effective.

Because of their exploratory nature, focus groups are better suited to use during the community assessment and program planning phases. This method of qualitative data collection may not be as useful for outcome assessments because it may be inappropriate to ask individuals to reveal in a group setting what difference the program made to them. In some situations, focus groups can be appropriate as one method to collect process monitoring from staff members.

Data Collection

The steps involved in selecting focus group members, conducting the focus groups, managing the group process, and analyzing the data have been explicated by various authors (Krueger & Casey, 2014; Stewart & Shamdasani, 2015). The types of questions posed to a focus group are constructed as open-ended questions. To maximize the data provided by the group members, the focus group leader actively directs the discussion by seeking clarification and by eliciting views and opinions from members more inclined to be quiet. Data are collected as either audio recordings or extensive written notes, and sometimes both; audiovisual recordings are rarely used. The process of recording can be somewhat intrusive, especially if the group members are not comfortable with a recorder or are distracted by the note taker. Thus, the physical location of the recorder and the note taker with regard to the focus group members becomes a conscious choice of how not to influence the interactions of the group members.

The most immediately noticeable challenge is the logistics of scheduling a meeting with the 8 to 12 members of a focus group. Along these lines,

it can be difficult to predict which combination of individuals will be optimal for producing insightful data. A balance between homogeneity and heterogeneity of group members is needed but can be difficult to anticipate. The selection of focus group members, otherwise called the sampling strategy, should be guided by the notion that people who share commonalities are more likely to communicate with one another. At the same time, some degree of differences among the members stimulates discussion and promotes clarification.

Another challenge relates to capturing the data produced by a focus group. People frequently talk when others talk, and they often have voices that are difficult to identify on audio recordings. Similarly, when taking handwritten notes, the note taker inevitably misses some comments. For this reason, having more than one person conduct a focus group and having more than one data-recording technique enhance the accuracy of the data captured.

Observation

Using one's own eyes to collect data is another qualitative method. When a person is looking at the behavior of others and making interpretations about that behavior, the nature of the data becomes more subjective—that is, more qualitative. Observational methods vary widely and range from nonparticipatory techniques, such as using rooms with one-way mirrors to observe parent–child interactions, to participatory techniques, such as assuming the identity of a program participant to experience what participants experience. Observations also occur across a wide variety of settings—from natural, such as a home or street corner in the neighborhood of the intended audience, to less naturalistic, such as a clinic or lab.

Data Collection

Depending on the purpose for using an observational method, the specific procedures used to conduct the observation and record

the data will vary. Nonetheless, there are some commonalities. First, training is needed with regard to what will be observed. For example, if the purpose is to assess the quality of interaction between program staff members and participants, then dyadic or triadic actions and reactions will be the focus. If the purpose is to assess whether program participants have acquired a skill taught in the program, then the sequence of actions taken by the participant when in the situation requiring the skill will be the focus. In both process monitoring and the effect evaluation examples, many other events, interactions, and factors that are not immediately relevant to the purpose occur simultaneously and therefore are not recorded or noted during the observation. To help data collectors distinguish between relevant and irrelevant data, and collect data consistently, it is helpful to pretest data collection processes and keep notes from the lessons thereby learned as guides for subsequent data collection.

Data collection can be accomplished in several different ways. One technique relies on audiovisual recordings. Although this data collection method is likely to capture all events or interactions of interest, it is also the most intrusive from the point of view of the person being observed. The advantage of using recording is that data can be analyzed by several different analysts. Repeated viewings of the recordings can also help ensure that an important event or interaction of interest is not missed.

Another approach to collecting data involves coding events or interactions as they occur, preferably with a standardized data collection tool. For example, personal wearable devices have become popular ways of tracking physical activity as an alternate to the more traditional and more limited pedometer. Given their increasing use and with the most common wearable devices simply worn on the wrist, a well-trained data collector can often use these devices discreetly. However, the devices do not always yield complete data or measure accurately (Takacs et al., 2014; Wallen, Gomersall, Keating, Wisløff, & Coombes, 2016). Verifying the accuracy of recordings can also be

complicated and expensive. Evaluators can use the evidence from such studies to inform them of the extent and nature of likely data biases, and interpret the results accordingly.

Another option for collecting observational data entails making detailed notes and keeping a log of occurrences and observations after one is no longer doing the observation. This method relies heavily on the recall accuracy and memory of the person who made the observations as well as his or her ability to record what was observed with minimal interpretations in the description. While this approach to data collection is most susceptible to biases and inaccuracies, it is also the least intrusive.

Behaviors or events can also be measured sequentially, which can yield unique insights into program implementation processes and participant behaviors. At the interpersonal level, suppose a program staff member initiates a question (call it Q1) and the participant looks away (call it LA1), followed by the question being repeated (Q1). The Q1' LA1' Q1 sequence can then be analyzed (Bakeman & Quera, 2011). At the project level, it may be informative to track how projects unfold across months or years, from initial planning to maturity (Miles et al., 2014). A sequential analysis of very specific program processes could reveal patterns of hindering or supporting the successful program implementation.

Depending on the method used to record the observations, observational data may be susceptible to recall or observational bias. This is particularly the case if the observer is a participant-observer who is not at liberty to record the observations as they actually occur. Observation has other potential disadvantages. For example, deciding on a sampling frame can be difficult. In other words, the number of observations needed to be confident that the behavior could have occurred at a usual rate and in a usual manner may not be readily knowable. As with most qualitative methods, another disadvantage of the observation method is that the amount of time and resources required for data collection can be problematic. These

disadvantages are offset by the advantage of potentially identifying behaviors and events that might become known only through observation. The flexibility and variety of techniques that can be used to collect the data can be viewed as another advantage.

Case Study

A *case study* is an empirical inquiry into existing phenomena in their real-life contexts when the boundary between what is being studied and its context is not clearly evident. Typically, case studies use multiple sources of data (Yin, 2013). Hence, this is not necessarily a purely qualitative technique; however, the data and analyses tend to be more qualitative than quantitative. The case is determined based on the question being asked; thus, it could consist of an individual, a classroom, an organization, a program, or an event. The case study methodology is particularly useful for gaining insights into an entire program (Veney & Kaluzny, 1998) because the program is the unit of analysis (i.e., case). Typically, case studies address questions of how or why something occurred, and there is no attempt to control what happens experimentally. If the case study method is used during the assessment phase, then there is not yet a program and thus what constitutes the case is unclear. If the case study method is used for the effect evaluation, then either the program implementation is an event that is the focus of the case study or the program itself is the case.

The methodology for conducting a case study is given by Yin (2013) and Stake (2005). Both authors argue that, as with any design, attention needs to be paid to the generation of the research question and rival hypotheses, and to the use of methods and techniques that enhance scientific rigor of the study. Stake (2005) notes that one distinctive challenge is clarifying what constitutes a case. For health evaluators, the case is likely to be a program but not the reasons for the program or relevant policies, which are viewed as being too general to be examined as a case.

Data Collection

The choice of the case is critical and must be based on carefully considered criteria for what makes the case either uniquely typical or extraordinarily different. Generally, when the case study method is used for program process monitoring, the health program is the case. Sometimes an embedded design best fits the data and evaluation questions. For example, for some effect evaluations, evaluators may select a few individual program participants to be the cases in order to refine the effect theory based on their experiences and processes of change. Lochman, Boxmeyer, Powell, Wojnaroski, and Yaros (2007) chose one participant in a program for children with disruptive behavior. With the child as the case, they interviewed the parents, teachers, classmates, and the child over a period of several months. In this way, they were able to assess both the program implementation and the program effect on the child. Where multiple institutions or organizations are providing the intervention, the case study may examine the effects of the different contexts on the intervention. A more costly design that can yield additional insights is the multiple case study, in which two or more, generally very similar programs in different settings are compared to discern factors associated with desired outcomes.

A key feature of the case study methodology is the use of multiple sources of data. Typically, data collected as part of a case study include both primary data—generally collected via interviews, observations of behavior, and surveys—and secondary data. Secondary data collection might include reviewing agency or program documents; existing data collected for other evaluation or program monitoring purposes; and program-related materials, such as promotional materials, policies, and procedures. A case study then involves considering all these data to arrive at some answers to the evaluation question. The amount of data collected may prove overwhelming to evaluators, resulting in “analysis paralysis” or delays in arriving at sound conclusions.

Despite their challenges, case studies are common in program planning and evaluation because of their utility in explaining how programs achieve intended results. From published reports, the value of such studies becomes evident. For example, Glik et al. (2016) describe how three communities across the United States developed differing approaches to decrease cardiovascular disease based on their distinct priorities and resources. Laws et al. (2016) used a comparative case study to trace how five areas in Victoria, Australia, implemented an obesity prevention program for infants and toddlers, and found facilitative factors to include prepackaged programs relevant to local needs, along with training by external evaluators and local coordinators who helped align the new program with existing services. A case study in New Mexico found that a coalition successfully advocated for comprehensive sex education in schools, despite losing funding due to the federal administration’s abstinence-only stipulation (Mendes, Plaza, & Wallerstein, 2014). This study reinforces the idea that program planning and evaluation occurs at the infrastructure level of the public health pyramid.

Innovative Methods

New qualitative methods and approaches continue to be developed. The qualitative methods reviewed in this chapter have included the major types, which form a broad base for development of other methods. The three additional qualitative methods introduced in this section are examples of the evolution of qualitative methods and their applicability to program planning and evaluation.

Photovoice

Photovoice involves the use of photography by the participants and analysis of their photographs to understand the phenomena and lives of the participants who took the photographs. The photovoice methodology involves not only taking the photographs as the data collection

phase but also interpreting those photographs through dialogue or discussion. The ease in using disposable and smartphone cameras make photovoice feasible to use with diverse study participants. For instance, photovoice is an attractive alternative to qualitative methods that rely heavily on language. For this reason, it has been used with individuals with intellectual disabilities (Jurkowski, Rivera, & Hammel, 2009). The interactive, interpretive aspect of photovoice also makes it appealing for use in community-based participatory research. In one study, parents explained pictures they had taken to identify factors facilitating and impeding healthy eating and physical activity for their families (Mareno, 2015). Photos and videos enabled middle school students to describe factors affecting their asthma, after which participants had higher asthma-related self-confidence, and caregivers scored higher in measures of asthma control (Blumenstock, Gupta, & Warren, 2015).

Narrative Designs

A qualitative approach that is used rarely in health program planning and evaluation entails the use of narrative methods. Narrative methods use text as the data. The text can come from personal sources, such as diaries, and from public or agency sources, such as existing agency records, memos, reports, videos, newspapers, and other print media. Narrative designs focus on the linguistics of the texts and often follow a more straightforward content analysis approach (Krippendorff, 2012).

Narrative designs can be helpful during the community assessment phase for understanding how a community presents itself through the local media and other writings. For process monitoring, such designs can be very helpful in tracking program drift as reflected in writings about the program or program procedures. Narrative analyses have also been used to document how norms and power affect health-related interactions, such as how addiction treatment requires participants to develop “stories of change” to be considered

adequately engaged (Andersen, 2015, p. 678). Narrative analyses have also been used recently to understand public skepticism about the safety and efficacy of vaccinations used to respond to the H1N1 epidemic (Abeysinghe, 2015).

The obvious limitation of a narrative design is the inherently limited nature of text sources and the lack of an opportunity to clarify meanings or intent. On the positive side, narrative designs can be less expensive, particularly if the texts are publicly and readily available. In addition, stories, as narrative texts, are easy to obtain during interviews.

Mixed Methods and Triangulation

Each data collection technique has limitations. To compensate for some limitations, it may be helpful to use more than one data collection method. The notion of mixed methods encompasses using a combination of qualitative and quantitative methods. For example, Hacking et al. (2016) used both quantitative and qualitative data to evaluate text messaging related to health behaviors. Quantitative pre–post data did not show improvements in health behavioral knowledge. Participants explained in focus groups that these text-message prompts provided valuable behavioral reinforcement but not new health information (Hacking et al., 2016). In a study of factors affecting community-based research, Native American focus group participants’ qualitative explanations of varying trust toward academic partners informed development of a nationwide quantitative survey (Lucero et al., 2016). Such inclusion of different types of data helps evaluators draw more realistic conclusions about factors affecting program implementation and results.

Use of multiple methods can sometimes yield quite different results, as happened in the study carried out by Quinn, Detman, and Bell-Ellison (2008). They used quantitative survey and qualitative interviews to understand barriers to use of prenatal care. The findings from the

two methods were quite different. According to the survey, transportation and child care were the major barriers, whereas embarrassment and other emotional reasons were the major barriers identified in the interviews. Differences in findings from qualitative and quantitative data collection have been found even in the reporting of deaths (Huy, Johansson, & Long, 2007).

In some instances, multiple methods are used for triangulation. The term *triangulation* refers generally to the use of multiple methods for the purpose of confirming, disconfirming, or modifying information gained through one of the methods. In some way, perspectives can be reconciled via this approach (Thurmond, 2001), but debates continue on how best to achieve rigorous triangulation. Triangulation has been used to understand program implementation better. One example of triangulation used for program planning was a needs assessment conducted for a program in Chicago (Levy et al., 2004). These researchers used focus groups, key informant interviews, observational data, and quantitative data from the Behavioral Risk Factor Surveillance System (BRFSS) to arrive at a set of concerns regarding diabetes and cardiovascular disease in a Latino and African American community. The authors demonstrated areas of overlap and disagreement in the data, which contributed to better understanding of both the communities and the optimal programmatic interventions.

The major challenge to using multiple methods is the analysis and synthesis of the data. Although it may seem more straightforward to analyze the data from each source independently, ultimately those findings need to be synthesized and reconciled with the findings based on the data collected with the other methods. In addition, it is important that each data collection method be done as rigorously as possible. Cost and time factors are additional considerations in choosing to use multiple methods. Given these substantial challenges, the multiple methods approach is more likely to be used in evaluation research or to be used on a very small scale.

► Scientific Rigor

Just as with quantitative methods, scientific rigor is important with qualitative methods. The terminology used to describe aspects of rigor is different; nonetheless, the underlying concepts are comparable. The following four elements of scientific rigor in qualitative methods (Lincoln & Guba, 1985) need to be present for the data to be considered of high quality.

Credibility is roughly equivalent to internal validity; it refers to whether one can have confidence in the truth of the findings. Credibility is established through several activities. First, evaluators must invest sufficient time in the process so that the findings can be triangulated. Second, evaluators may use outsiders—that is, individuals other than those who provided or analyzed the data—to gain insights into the meaning and interpretation of the data. This practice is sometimes referred to as peer debriefing. Third, evaluators should refine working hypotheses that are generated during data analysis with negative cases. This activity requires seeking out data that would be evidence that the working hypothesis is not always accurate. With qualitative data, working hypotheses are more like hunches that may be influenced by the researcher's bias. Therefore, in qualitative analysis, efforts are made to both confirm and disconfirm the working hypothesis. The act of seeking disconfirming data ensures that the researcher's biases are minimized in the results. A fourth activity that enhances credibility is checking findings against raw data, by reviewing again the raw data in light of the findings. Finally, credibility is increased when those who provided the data are asked to review the findings and provide feedback about the accuracy of the interpretations.

Transferability (also known as applicability) is similar to external validity in that the findings should have applicability to other contexts and respondents. The main technique for increasing transferability is to provide thick (detailed, comprehensive) descriptions in reports so that others can assess independently the possibility

of transferability of the findings to other groups. This stands in contrast to the quantitative technique, which focuses on sample selection as a means of increasing generalizability of findings.

Dependability is roughly equivalent to reliability. The notion is that other researchers and evaluators should be able to arrive at the same results if they repeat the study or analyze the same set of data. Ensuring dependability is done primarily by leaving a paper trail of steps in the analysis so that others can see that the findings are supported by the data. Another major aspect of dependability is the use of reliability statistics to demonstrate that, given the same set of data, two researchers would arrive at the same coding of the data. The two most widely used reliability statistics in qualitative research are the percentage agreement and the kappa statistic; both are determined after the codes for the data have been finalized, including having definitions. Percentage agreement is the simple percentage of a given data set that two researchers independently code into the same preestablished categories. The numerator is the number of data units that are similarly coded, and the denominator is the total number of data units to be coded. While this percentage is simple to calculate, it does not take into account the possibility that both researchers might have assigned a “wrong” code to the same data. To compensate for this shortcoming, the kappa statistic can be used (Cohen, 1960). The kappa statistic is based on the ratio of observed (actual) codes used to the expected codes to be used for a given set of data, for any two independently coding researchers.

Confirmability, sometimes referred to as objectivity, connotes that the findings are truly from the respondents and do not reflect the researcher’s perceptions or biases. As with dependability, the main technique for ensuring confirmability is to leave an audit trail that others can follow to arrive at the same findings. A variety of techniques to document the researcher’s impressions, biases, and interpretations are also available, such as making theoretical memos or taking field notes. Use of such techniques adds to

the confirmability of the findings by providing greater assurance that the researcher has taken steps to be self-aware of factors that might have influenced the interpretation of the data, and subsequently the results.

► Sampling for Qualitative Methods

Qualitative approaches to sampling are guided by a different set of driving forces. The first consideration is the design. **TABLE 15-4** summarizes the general considerations for selecting a sample for each of the qualitative methods. The second consideration is the specific sampling strategy to be used. The purpose of any qualitative design is to understand a phenomenon, so the sample is purposefully selected in reference to that phenomenon. In program planning and evaluation, the phenomenon in the community assessment or planning phases would be a characteristic of the intended population that is the focus of the qualitative assessment, whereas in the evaluation phase, the phenomenon would be the health outcome of participants or the implementation of the intervention by the program staff members.

The need to have a purposive sample means that there is no attempt to make the sample statistically representative of the intended population. Instead, the goal is to represent conditions of interest (Yin, 2013). Thus, random selection of a sample is not relevant. Power analyses are also not appropriate because there is no attempt to quantify an effect size or to find statistical significance. However, Morse (1994) provides some guidelines in terms of numbers. She suggests at least six participants are needed when trying to understand the experience of individuals, and 30 to 50 interviews are necessary for qualitative studies that focus on culture (ethnographies) and generation of a theory about a health problem (grounded theory studies).

Several different types of purposeful samples are used in qualitative research (**TABLE 15-5**). In

TABLE 15-4 Sampling Considerations for Each of the Qualitative Methods Discussed

Method	Sampling Considerations
Case study	Choice of case based on being either “usual” or “unusual”; maximum number of cases feasible to conduct
Observations	Ability to sample behaviors or events without altering their quality; need to obtain saturation of categories of findings
Individual in-depth interviews	Need to obtain category saturation; choice of individuals based on theoretical sampling
Focus groups	Representativeness of participants within and across the groups; appropriate size of each focus group for within-group discussion; minimum number of focus groups needed to capture diversity of views
Survey with open-ended questions	Linked to sampling strategy for the survey; likelihood of write-in responses
Narrative designs	Quantity and quality of existing documents available for review; access to existing documents

addition to those that are specific to qualitative research, the sampling strategies for accessing hard-to-reach populations would be appropriate. The qualitative sampling strategies are built around having either homogeneity or heterogeneity of participants. Homogeneity of participants is preferred if the phenomenon is not well understood and is being described for the first time or in-depth for a specific aggregate. To have homogeneity of the sample, either individuals are selected based on a set of inclusion/exclusion criteria or they are considered to be typical. Heterogeneity of a qualitative sample provides a range and points of possible contrast. The latter approach is helpful if the intent of the study is to expand the applicability of the findings.

Heterogeneity of a qualitative sample is achieved through selecting individuals purposefully based on their deviation from the norm or purposefully to maximize variation. If it is

unclear what might be “normal” or “deviant,” then randomly selecting individuals is an acceptable sampling strategy. The purpose of selecting a sample to have heterogeneity is to uncover as many as possible different responses and perspectives, which then become categories of information found in the data. Additional individuals are added until no new information is being gained from their participation, a strategy called sampling for category saturation. At some point, adding more data from additional individuals will not provide more or new information. This point, referred to as saturation, can be desirable as evidence of having sufficient data to minimize the possibility of missing any key concepts. Sampling for category saturation requires some degree of flexibility in the number of participants, whether cases or individuals, in the evaluation study. Category saturation can be achieved sooner or later than expected,

TABLE 15-5 Summary of Types of Sampling Strategies Used with Qualitative Designs

Sampling Strategy	Types of Cases Used	Use
Convenience	Participants or cases that are accessible and willing	Saves time and recruitment money
Critical cases	Exemplar participants or cases; cases important in some unique way	Permits generalization to similar individuals or cases
Deviant cases	Highly unusual participants or cases	Reveals the factors associated with unique and extreme conditions and may lead to a new theory or parameters
Maximum variation	Participants or cases with differing experiences	Fosters category saturation with most possible categories
Random purposeful	Participants or cases randomly selected from a large sampling pool	Adds credibility to the sample and thus some indication of generalizability
Typical cases	Usual or normal participants or cases	Leads to a broadly applicable theory or categories but does not address the full breadth of the program effects
Theory based	Participants or cases with the theoretical construct	Elaborates or refines the theory

allowing a sooner-than-expected conclusion to enrollment or necessitating the recruitment of additional participants, respectively.

Recall that the effect theory is a theory about how the intervention changes the causal theory. If the qualitative design is being used either to refine the causal theory of the factors that lead to the health problem or to assess the intervention theory, then a theoretical sample might be chosen. A theoretical sample is one that is purposefully chosen to complete or refine a theory. Theoretical sampling, which comes from the grounded theory tradition (Charmaz, 2014), entails sampling for ideas and constructs so that theories can be further

developed and refined. To accomplish this, individuals are selected based on whether they are anticipated to be reflective of or divergent from the theory. A theoretical sample would be useful, for example, with participants in a pilot program because those participants should have experienced the same theoretical effect from the intervention. Like sampling for category saturation, theoretical sampling requires flexibility in determining when enough data have been collected. Key difficulties in implementing a theoretical sample include knowing who might fit the theory and determining when the theory is no longer likely to change with the inclusion of additional participants in the qualitative study.

► Analysis of Qualitative Data

Numerous texts provide detailed instructions on analyzing various types of qualitative data, such as the classics by Miles and Huberman (1994); Miles et al., (2014); Krippendorf (2012); Weber (1990); and Lincoln and Guba (1985). These and other texts should be consulted for detailed instructions on the process of analyzing qualitative data, so that scientific rigor is maintained. Remember that program planners and evaluators, whether engaged in qualitative or quantitative data collection, need to make every reasonable effort to achieve scientific rigor as a basis from which claims can be made more solidly and confidently. This section offers a brief summary of the steps involved in the immersion into the data and surfacing from the depths of analysis that occur when working with qualitative data.

A note of caution is warranted. While the steps in this procedure are presented in a sequential order here, most of the steps occur iteratively in real-world qualitative data analysis. It's a bit like taking two giant steps forward and then needing to take three baby steps backward before stepping forward again. This is a normal and expected part of the process of qualitative data analysis.

Overview of Analytic Process

Before analysis can begin, the data need to be transformed into a format amenable to manipulation and analysis. This effort may involve transcribing audio-recorded interviews or entering text, audio, or visual data into qualitative analysis software. Once the data are transformed reliably and accurately, analysis can begin.

The first step is to decide what factors are codable units of data and then identify those units within the data. Codable units from in-depth interviews can consist of words, phrases, or paragraphs, whereas codable units from observation may be facial movements or interactions. In contrast to quantitative data, in which the numerical

response to a single question or a lab value is an obvious unit of data, what constitutes a unit of data is not always so obvious in qualitative data. There is a constant struggle to identify the units of data so that they can be categorized based on their properties. The unit of analysis varies by method, question being asked, and the underlying perspective from which the question is being asked. Also, the codable unit of analysis can evolve as the analysis proceeds, becoming either larger or smaller in terms of amount or complexity.

The next step is to understand the meaning of what was said, observed, or read. Certain terminology is unique to qualitative methods and its data analysis. Manifest versus implied meanings are critical in qualitative data. *Manifest meanings* are the obvious, unambiguous meanings, whereas *implied meanings* are the unspoken innuendos, metaphors, and references that color the meaning and interpretation. Data analysis relies on making this distinction and being faithful to the decision to code the data based on manifest or implied meanings.

Based on the meanings, a discovery process begins in which groups of data with similar meanings begin to form categories. This process is rather idiosyncratic, with each researcher having a preferred style. Some begin with broad, overarching categories and gradually generate more specific subcategories. Others prefer to begin with many discrete categories; then, based on grouping similar categories, they evolve toward broader, overarching categories. Either approach is fine because both result in a nested set of categories, not unlike an outline of nested ideas that contains broad topics followed by corresponding, more specific subheadings. The process of generating categories is one of constantly comparing the data to be coded with the data that have already been categorized. Constant comparison is a trademark of qualitative data analysis. The process results in the feeling of being immersed in data, lost among the trees in the forest. As the categories evolve, memos are generated that are necessary to have dependability.

At the point at which sufficient data have been analyzed for categories to develop, the categories

are named. This step is likely to occur along with the sorting of data into groups or categories. A category is a classification of concepts in the data. As more data are reviewed for their meanings and grouped accordingly, the properties and dimensions of the categories begin to surface. A dimension implies that the quality of the data exists along a continuum, whereas a property is a discrete attribute or characteristic. As more data are added to the categories, new data are constantly compared with data already in the category, and adjustments are made. **TABLE 15-6**

shows an example of interview text, the coded units, and the categories to which those data units have been assigned.

When the categories are reasonably well established, definitions of the categories are developed. Based on the properties and dimensions, definitions help establish criteria for whether data belong to that category and are used as the guide for whether new data will be added to that category or a different category. Definitions are considered good if they enable the categories to meet two standard criteria for

TABLE 15-6 Example of Interview Text with Final Coding

Interview Text	Final Coding
<p>Interviewer: Okay. Thinking specifically about how case management may have affected you or made a difference to you, how do you think case management has affected your health or your pregnancy or the baby's health? How has it made a difference?</p> <p>Respondent: It really didn't make me no difference.</p>	<p>Outcome code: Case management made no difference to me</p>
<p>Interviewer: You couldn't identify any ways that it made a difference?</p> <p>Respondent: No.</p>	
<p>Interviewer: How has case management affected your thoughts or feelings about yourself?</p> <p>Respondent: I have had <i>when I was pregnant. I was depressed and everything</i>, and <u>talking to my case manager</u> made me feel better about myself.</p>	<p>Risk code: <i>expressed psychosocial state, depression</i></p> <p>Intervention code: <u>talking and listening</u></p> <p>Outcome code: increased self-care attributed to case manager involvement</p>
<p>Interviewer: How did that happen? How did she make you feel better about yourself?</p> <p>Respondent: She just told me that don't think about nothing else and let it upset you and just think about what you got inside of you because <i>I was under a lot of stress</i> and <i>she said that if I be stressed out too much I could harm the baby.</i> So, I thought about that, and that was serious. If my pregnancy hadn't been stressed out and depressed like that.</p>	<p>Risk code: <i>expressed psychosocial state, stress</i></p> <p>Intervention code: <u>gave information</u></p> <p>Outcome code: patient demonstrates a change in knowledge</p>

category development: mutually exclusive and exhaustive. *Mutually exclusive* category definitions allow for data to be in one and only one category; there is no other choice for which category best reflects the data. *Exhaustive category* definitions prohibit the use of “other” as a category; all data have a place, a category in which they belong. Using the concepts of mutually exclusive and exhaustive categories is very helpful as a gauge of definition clarity and for deciding whether more data need to be collected. If considerable amounts of data cannot be categorized, then either the categories need to be changed (broadened or narrowed) or more data need to be collected so that the category properties, characteristics, and dimensions can be understood more fully.

Depending on the qualitative method and perspective taken, a step sometimes taken at this point is to present the findings to the participants in the study. The study participants are offered the opportunity to confirm the results and discuss alternative interpretations of the data. This step adds to the confirmability of the final results.

In the final step in qualitative analysis, researchers typically generate explanations or working theories based on the data. In terms of program planning and evaluation, this may involve revisiting the process theory or the effect theory or any element of those theories. For example, findings from a qualitative evaluation may cause researchers to revise their view of the key contributing or determinant factors of the health problem or to rethink their perception of which intervention is more effective. In addition, a separate model of the health problem might be developed based on the findings. Ultimately, the results need to be considered in reference to the purpose of the study and the questions that were asked at the outset.

Software

Several commercial software programs are available for use in managing qualitative data and facilitating analysis of such data. Many evaluators use software such as NVivo, although manual methods are also still common (Malone &

Coyne, 2016). Software can facilitate analyses when there are a large number of documents by making it easy, with a given code, to pull into one place all the text segments across all documents. Qualitative software also facilitates coding by multiple team members, albeit sometimes with technical problems that can be time consuming to address. One major consideration may be the price, which tends to start at several hundred dollars per user for the leading software programs, although there are some free software programs, such as Qualitative Data Analysis Program (QDAP). The “Internet Resources” section at the end of this chapter gives links to the websites for several leading qualitative software programs. These programs have features that facilitate diagramming relationships among categories, counting units of analysis per category, and coding within coded text. The functions of software for qualitative data analysis include searching for text, text-based management in terms of linking text to information about that text, coding and retrieving, support theory building by allowing for higher-order coding, and building networks of concepts in graphic representations.

Software programs can greatly facilitate managing the potentially overwhelming amounts of data. Nevertheless, the conceptual work remains for the evaluator to do. Interpretation of meanings so that codes are applied to appropriate codable units, and ongoing development and refinement of categories, domains, and properties are some of the tasks that cannot be done by the computer software. Also, these programs, like any software program, require training prior to their use.

Issues to Consider

Given the traditional and normal reliance on numbers, one issue when reporting qualitative findings is whether to report numbers—in other words, “to count or not to count.” When themes and codes are developed, there is an inclination to count the number of occurrences or generate percentages related to the codes. If data are presented as counts and percentages,

however, these statistics must be interpreted very cautiously; only one or two vocal individuals with many comments could easily change the count or skew the percentages.

The challenge in reporting numbers from qualitative data is determining what is an appropriate denominator. The denominator may consist of either the number of all coded comments or the number of individuals from whom data were collected. Equally important is what is an appropriate numerator: It could be the number of participants who mentioned a category or the number of times a category is mentioned throughout the study. These choices greatly influence how the findings are portrayed and interpreted.

Qualitative methods are notorious for being messy, confusing, and repetitive. While many of these qualities can be minimized using available software for qualitative data management, the nature of the data requires iterative category development. This can prove challenging because the quantities of data are often overwhelming and because the various stakeholders and evaluation participants might potentially provide conflicting interpretations of the data. The evaluator, program planner, and relevant stakeholders need to be committed to obtaining and using the qualitative data. Otherwise, the delays and frustrations will overcome the ultimate value of the data, and the data will be set aside to collect dust.

Then there is the cost. When budgeting for collection and analysis of qualitative data, numerous elements must be included as possible expenses. There is travel time and mileage to and from the sites where the data are collected. Interviews that are recorded need to be transcribed and typed verbatim. This activity is roughly equivalent to data entry from a close-ended survey. However, every hour of interview generates approximately three hours of transcription. If independent coders are used to establish the reliability of coding data, they may need to be paid for their time and efforts.

A major issue for conducting any qualitative research is the need to have highly

trained data collectors and analyzers. Training encompasses not only how to achieve consistent, reliable, and unbiased collection of data but also how to acknowledge personal preferences and interpretations that may influence the analysis of the data. Training for data collection is important for in-depth interviewing to minimize the interviewer's use of nonverbal or verbal cues that might lead the interviewee to respond in the way he or she thinks the interviewer wants. Humans are very good at reading other humans, and subtle facial movements and body language can alter the response of an interviewee. Training for data analysis is more complicated and is best accomplished by using a team approach. Team members can challenge one another, seek clarification on interpretations of data, and provide checks and balances during the data analysis process.

► Presentation of Findings

As with findings based on quantitative data, the findings based on qualitative data need to be communicated to the relevant stakeholders and presented in a manner that both conveys the scientific soundness of the findings and is understandable. One aspect of conveying the scientific soundness of the findings is to include descriptions of the context in which the data were collected to show transferability of the results. Tables can be constructed that show the evolution of category development and can help demonstrate dependability and confirmation. Most important is including the words of the study participants alongside the category; this helps show confirmation. Stories, descriptions, explanations, and statements provided in the words of the participants are more powerful than numbers, and they make the numbers "more human." These statements can be quite powerful as tools for marketing the program and for fundraising. The final step of generating

an explanation may result in diagrams of relationships among categories that provide a visual means of understanding the overall findings. To the extent that the data can be associated with the logic model of the program or the program theory, the findings will be viewed as immediately relevant.

► Across the Pyramid

At the direct services level of the public health pyramid, qualitative methods—and particularly interviews—are used to answer questions about individual perspectives, interpretations, perceptions, and meanings. Observational methods can also be readily used for process monitoring evaluations (TABLE 15-7). Photovoice is well suited for use by individuals for community assessment data collection.

At the enabling services level of the public health pyramid, qualitative methods are likely to focus on questions about individual perceptions and interpretations as well as on more aggregate-wide perceptions. To obtain individual perceptions for either planning or impact evaluations, in-depth interviews would be appropriate. To ascertain the more common

perceptions of members of an aggregate, focus groups would be quite useful.

At the population services level, qualitative methods are more likely to be used during the assessment phase of the planning and evaluation cycle. Qualitative methods can provide detailed and specific information on cultural understandings related to health, illness, and prevention. These population-level, or aggregate-wide, cultural findings can then be used to design better population-level services.

At the infrastructure level of the public health pyramid, qualitative methods have been used for a variety of purposes. Ryan et al. (2016) identified public health infrastructure priorities related to flood-related disasters from qualitative analyses of prior articles. van den Driessen Mareeuw, Vaandrager, Klerkx, Naaldenberg, and Koelen (2015) interviewed a range of stakeholders to identify dynamics undermining the translation of knowledge into action in the Dutch public health system, including a tendency for powerful entities such as the Ministry of Health to set the research agenda rather than the intended users of new knowledge. As these examples illustrate, qualitative methods can be used to address the role of public health infrastructure at different stages of the planning and evaluation cycle.

TABLE 15-7 Suggested Qualitative Methods by Pyramid Level and Planning Cycle

Services Level	Assessment	Planning	Monitoring	Outcome Evaluation
Direct	Interview; photovoice	Interview	Observation	Interview; observation
Enabling	Focus group; observation	Focus group	Observation	Interview
Population	Focus group	Focus group	Focus group	Focus group
Infrastructure	Case study; narrative	Case study	Case study	Case study; interview; focus group

► Discussion Questions and Activities

1. Which qualitative method(s) do you believe would be ideally suited to use at each stage of the planning and evaluation cycle? Justify your answer.
2. If you had a very limited budget but were committed to collecting and using some type of qualitative data to get at the perceptions of program participants, what might be your best option? Explain.
3. Apply the four criteria for assessing the rigor of qualitative studies to an evaluation of a program to increase adult immunizations in Bowe County that includes financing for individuals who are uninsured or underinsured.
4. Discuss under what circumstances you would and would not use numbers in the presentation of results from each of the types of qualitative methods reviewed.

► Internet Resources

Qualitative Evaluation Checklist

Michael Quinn Patton's qualitative evaluation checklist posted at the Western Michigan University Evaluation Center's website summarizes when and how to conduct a qualitative evaluation; see <https://www.wmich.edu/sites/default/files/attachments/u350/2014/qualitativeevalchecklist.pdf>.

The Pell Institute

This site provides a six-step guide to qualitative analyses. The guide draws heavily on the second edition of Miles and Huberman's guide to qualitative analyses (1994), which is less expensive than the most recent edition and still relevant: <http://toolkit.pellinstitute.org/evaluation-guide/analyze/analyze-qualitative-data/>.

National Science Foundation

This guide also draws on the 1994 edition of Miles and Huberman's influential text for suggestions on qualitative data analysis: https://www.nsf.gov/pubs/1997/nsf97153/chap_4.htm.

Qualitative Data Analysis Software

Three commercial software packages available to assist in data analysis are NVivo (www.qsrinternational.com/), ATLAS.ti (www.atlasti.com/), and Ethnograph (www.qualisresearch.com/). At each website, the overview of product features gives insights into the possibilities of qualitative analyses. If expense is an issue, however, you can try QDAP (<http://www.umass.edu/qdap/>). A comparison of qualitative data analysis software packages is available at <http://dlab.berkeley.edu/blog/choosing-qda-software>.

References

- Abeyasinghe, S. (2015). Vaccine narratives and public health: Investigating criticisms of H1N1 pandemic vaccination. *PLoS Currents*, 7, ecurrents.outbreaks. doi:10.1371/currents.outbreaks.17b6007099e92486483872ff39ede178
- Andersen, D. (2015). Stories of change in drug treatment: A narrative analysis of "whats" and "hows" in institutional storytelling. *Sociology of Health & Illness*, 37(5), 668–682. doi:10.1111/1467-9566.12228
- Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. New York, NY: Cambridge University Press.
- Blumenstock, J., Gupta, R., & Warren, C. (2015). Empowering students with asthma in Chicago schools through photovoice and videovoice. *Journal of Allergy and Clinical Immunology*, 135(2), AB157. <http://dx.doi.org/utsph.idm.oclc.org/10.1016/j.amepre.2015.03.011>
- Cabassa, L. J., Contreras, S., Aragón, R., Molina, G. B., & Baron, M. (2011). Focus group evaluation of "secret feelings": A depression fotonovela for Latinos with limited English proficiency. *Health Promotion and Practice*, 12, 840–847. doi:10.1177/1524839911399430
- Charmaz, K. (2014). *Constructing grounded theory*. Thousand Oaks, CA: Sage.
- Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational Psychology and Measurement*, 20(1), 37–46.

- Esangbedo, D. O., Ughasoro, M. D., Tagbo, B. N., Olowu, A., Anikene, C., & Iwegbulam, C. C. (2016). Health-care workers' perspectives on Ebola virus vaccine: A focus group and in-depth interview interventional study. *The American Journal of Tropical Medicine and Hygiene*, 95(3), 654–662. doi:10.4269/ajtmh.16-0206
- Glik, D. C., Sharif, M. Z., Tucker, K. L., Tejada, S. A., Prellip, M. L., Ammerman, A. S., Keyserling, T. C., Torres, S. E., & Jilcott Pitts, S. (2016). Community-engagement to support cardiovascular disease prevention in disparities populations: Three case studies. *Journal of Health Disparities Research and Practice*, 9(1), 6.
- Hacking, D., Haricharan, H. J., Brittain, K., Lau, Y. K., Cassidy, T., & Heap, M. (2016). Hypertension health promotion via text messaging at a community health center in South Africa: A mixed methods study. *JMIR mHealth and uHealth*, 4(1). doi:10.2196/mhealth.4569
- Huy, T. Q., Johansson, A., & Long, N. H. (2007). Reasons for not reporting deaths: A qualitative study in rural Vietnam. *World Health & Population*, 9(1), 14–23.
- Jurkowski, J. M., Rivera, Y., & Hammel, J. (2009). Health perceptions of Latinos with intellectual disabilities: The results of a qualitative pilot study. *Health Promotion Practice*, 10(1), 144–155. doi:10.1177/1524839907309045
- Krippendorff, K. (2012). *Content analysis: An introduction to its methodology* (3rd ed.). Beverly Hills, CA: Sage.
- Krueger, R. A., & Casey, M. A. (2014). *Focus groups: A practical guide to applied research* (5th ed.). Thousand Oaks, CA: Sage.
- Laws, R., Heskeith, K., Ball, K., Cooper, C., Vrljic, K., & Campbell, K. (2016). Translating an early childhood obesity prevention program for local community implementation: A case study of the Melbourne InFANT Program. *BMC Public Health*, 16(1), 748. doi:10.1186/s12889-016-3361-x
- Levy, S. R., Anderson, E., Issel, L. M., Willis, M. A., Dancy, B. L., Jacobson, K. M. . . . Hebert-Beirne J. (2004). Using multilevel, multisource needs assessment data for planning community interventions. *Health Promotion Practice*, 5, 59–68. doi: 10.1177/1524839903257688
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Lochman, J. E., Boxmeyer, C., Powell, N., Wojnarowski, M., & Yaros, A. (2007). The use of the coping power program to treat a 10-year-old girl with disruptive behaviors. *Journal of Clinical Child and Adolescent Psychology*, 36(4), 677–687. doi:10.1080/15374410701662881
- Lucero, J., Wallerstein, N., Duran, B., Alegria, M., Greene-Moton, E., Israel, B., . . . White Hat, E. R. (2016). Development of a mixed methods investigation of process and outcomes of community-based participatory research. *Journal of Mixed Methods Research*. doi:10.1177/1558689816633309
- Malone, H. E., & Coyne, I. (2016). A review of software applications and non-software approaches in JAN. *Journal of Advanced Nursing*. doi:10.1111/jan.13150
- Mareno, N. (2015). Parental perception of healthy eating and physical activity: Results from a preliminary photovoice study. *Journal of Clinical Nursing*, 24(9–10), 1440–1443. doi:10.1111/jocn.12726
- Mendes, R., Plaza, V., & Wallerstein, N. (2014). Sustainability and power in health promotion: Community-based participatory research in a reproductive health policy case study in New Mexico. *Global Health Promotion*. doi:10.1177/1757975914550255
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.
- Morse, J. M. (1994). Designing funded qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 220–235). Thousand Oaks, CA: Sage.
- Quinn, G. P., Detman, L. A., & Bell-Ellison, B. A. (2008). Missed appointments in perinatal care: Response variations in quantitative versus qualitative instruments. *Journal of Medical Practice Management*, 23(5), 307–313.
- Ryan, B. J., Franklin, R. C., Burkle, F. M., Watt, K., Aitken, P., Smith, E. C., & Leggat, P. (2016). Defining, describing, and categorizing public health infrastructure priorities for tropical cyclone, flood, storm, tornado, and tsunami-related disasters. *Disaster Medicine and Public Health Preparedness*, 10(4), 598–610. doi:10.1017/dmp.2016.3
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443–466). Thousand Oaks, CA: Sage.
- Stewart, D. W., & Shamdasani, P. N. (2015). *Focus groups: Theory and practice* (3rd ed.). Thousand Oaks, CA: Sage.
- Strachan, C. E., Nuwa, A., Muhangi, D., Okui, A. P., Helinski, M. E., & Tibenderana, J. K. (2016). What drives the consistent use of long-lasting insecticidal nets over time? A multi-method qualitative study in mid-western Uganda. *Malaria Journal*, 15(1), 1. doi:10.1186/s12936-016-1101-4
- Strauss, A., & Corbin, J. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Thousand Oaks, CA: Sage.
- Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C., & Hunt, M. A. (2014). Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, 17(5), 496–500. http://dx.doi.org/10.1016/j.jsams.2013.10.241
- Thurmond, V. A. (2001). The point of triangulation. *Journal of Nursing Scholarship*, 33, 253–258. doi:10.1111/j.1547-5069.2001.00253.x

- van den Driessen Mareeuw, F., Vaandrager, L., Klerkx, L., Naaldenberg, J., & Koelen, M. (2015). Beyond bridging the know-do gap: A qualitative study of systemic interaction to foster knowledge exchange in the public health sector in The Netherlands. *BMC Public Health*, 15(1), 922. doi:10.1186/s12889-015-2271-7
- Veney, J., & Kaluzny, A. (1998). *Evaluation and decision making for health services*. Chicago, IL: Health Administration Press.
- Wallen, M. P., Gomersall, S. R., Keating, S. E., Wisløff, U., & Coombes, J. S. (2016). Accuracy of Heart Rate Watches: Implications for Weight Management. *PloS one*, 11(5), e0154420. doi:10.1371/journal.pone.0154420
- Weber, R. (1990). *Basic content analysis* (2nd ed.). Newbury Park, CA: Sage.
- Yin, R. K. (2013). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.



SECTION VI

Additional Considerations for Evaluators

CHAPTER 16

Program Evaluators' Responsibilities

A premise throughout this text is that the program planning and evaluation process is a cycle with feedback loops throughout, from assessment and planning through outcome evaluation. Data and information generated by an evaluation do not automatically become a feedback loop; rather, intention and effort are needed to create and maintain the connection between evaluation findings and subsequent planning. Influencing the strength of the feedback loops is a set of ethical evaluator responsibilities, including responsibly presenting the data, reporting, communicating with people who may be affected by the evaluation outcomes, ensuring the evaluation quality, and staying current in evaluation practice. These responsibilities of the evaluator, if fulfilled, help ensure that the planning and evaluation cycle continues in a productive manner.

► Ethical Responsibilities

Ethics remains a foremost concern in health care. *Ethics* is the discipline or study of rights, morals, and principles that guide human behavior. Issues become ethical when basic human rights are

involved or when dilemmas arise about what might be the moral and principled course of action. In terms of health program development and program evaluation, the potential for ethical concerns is omnipresent.

Choices regarding who will or will not receive the health program can raise ethical questions. For example, if random assignment is planned, then the ethical question is what to do for those who will not receive the program. Also, the choice of the intended population can be subject to ethical questioning. What makes one high-risk or vulnerable group of individuals more or less worthy of receiving a health program can be an ethical issue. It is possible that conflicts regarding the development of a health program are implicitly conflicts of ethical perspectives. The following discussion focuses on the ethics and corresponding legal considerations that are directly related to program evaluation and delivery.

Ethics and Planning

Bioethics, or the application of ethical and philosophical principles to medical decision making, tends to focus on decision making

within the context of acute care or end-of-life decisions. However, global concerns over the possibility of *pandemics*, the spread of diseases over a large portion or an entire population, and the need for greater disaster preparedness have led to the application of ethical principles to population-focused health planning. Ethical principles, as derived from the philosophy of moral behavior and used as guides to health planning, determine the priority given to health problems and the approaches to deal with those problems. The choice of which ethical principles and frameworks to apply from among the variety that exists has consequences throughout the program planning and evaluation cycle.

For this reason, a brief overview of some of the frequently applied ethical frameworks and principles (TABLE 16-1) is in order. This overview

is not intended to be comprehensive or thorough. Rather, the intent here is to introduce the major ethical principles that play crucial roles in health program planning.

Autonomy, as an ethical principle, is the right to make a decision on one's own behalf. Applied to health programs, autonomy is the decision that each potential participant or participant group makes about whether or not to join the health program. It is such a basic ethical principle that it is often taken for granted as being an option. However, for some groups, such as those with intellectual disabilities, language deficits, or persons who believe they may lose services if they do not participate in an evaluation, autonomy is compromised.

One ethical approach, *criticality*, is to focus on providing for those who have the greatest

TABLE 16-1 Ethical Frameworks and Principles for Planning Health Programs

Approach	Principle	Health Application	Examples in Health
Autonomy	Personal right to self-determination and choice	Individual choice takes priority; avoidance of coercion	Pro-life and pro-choice; living will
Criticality (contractarian)	The worst off benefit the most	Greatest problem; severest health risk	Neighborhood prioritization for enhancements of built environment, such as parks and walkways; eligibility for dual Medicare–Medicaid
Egalitarian	All persons of equal value; minimize disparities	Treat everyone equally with regard to respect, rights, and opportunity	Medicare; universal screening; Patient's Bill of Rights
Resource sensitive	Resources are scarce	Cost-effectiveness as the standard	Oregon Health Plan; cost-effectiveness as part of health program evaluations
Utilitarian	The greatest good for the greatest number; the ends justify the means	Collective benefits outweigh individual choices	Required immunization; <i>Healthy People 2020</i> goals; prohibitions of smoking in public places

or severest health problem or risk. Under this ethical philosophy, persons with life-threatening illnesses or individuals at highest risk for exposure to air particulate pollution would receive high priority. This approach would also lead to an emphasis on addressing the health problems of groups, such as substance-abusing sex workers, isolated women suffering physical abuse, and undocumented farm workers. Taking this approach is likely to be costly for the healthcare organization and for society.

In contrast, the *egalitarian* ethical approach asserts that all individuals have the same fundamental rights and deserve the same (equal) access to opportunities and resources. In theory, the same treatment and opportunities should result in equity across groups. The egalitarian ethic underlies the Patient's Bill of Rights and the drive to have universal access to health care. The *resource-sensitive* ethical approach emphasizes being cost-effective, so that the least resources are expended for the greatest amount of health gain. A cost-effectiveness ethic might underlie targeting common health problems with inexpensive solutions, such as primary prevention; focusing on preventing high-cost outcomes, such as premature births; or seeking to reduce duplication of programs.

Prioritization could be approached from a *utilitarian* philosophy in which achieving the greatest good or benefit for the greatest number of persons is the guiding principle. The utilitarian ethic gives a high priority to the reduction in air pollutants and quarantine of individuals with highly contagious infectious disease. The vast majority of public health programs are predicated on the utilitarian principle, making it an integral element of public policy making. This also makes it appear to be a core value rather than an ethical principle. The *Healthy People 2020* objectives operationalize both needs-based and utilitarian ethics by directing attention, and thus efforts, toward those groups that are farthest from national target objective levels and recognizing the benefit to society as a whole.

There is growing global, national, and local awareness of the need for preparedness

for disasters with both natural and human causes. Along with planning efforts geared toward being prepared to respond to disasters has come a parallel awareness that planning for disasters has profound ethical implications. Measures focusing on isolation via quarantine, degree of public health authority, and extent of response force all have the potential to conflict with individual autonomy and freedom. Thus, public health disaster preparedness planning is increasingly including explicit statements of ethical principles to be followed during a response and to guide the planning.

Institutional Review Board Approval and Informed Consent

Ethical issues are most likely to surface with regard to the need to have participants in the evaluation provide informed consent. Informed consent is the agreement to participate voluntarily and willingly in a study based on a full disclosure of what constitutes participation in the study and what are the risks and benefits involved in participating. Whether informed consent is used in the evaluation is based on a set of factors: the requirements of the funding agency, the requirements of the agency providing the program, and the intent of the evaluation. If the evaluation will be used only for internal managerial purposes and not to generate knowledge, then the evaluation is not research and informed consent is not required.

In contrast, if the evaluation is done with the intent of generating generalizable knowledge, then the evaluation is considered research. For example, if the findings of the evaluation are to be shared beyond the program agency, as in an academic journal, the evaluation qualifies as research. In that case, evaluators, as researchers, are obligated to comply with federal regulations regarding obtaining informed consent. Universities and healthcare organizations that receive federal research grants have *institutional review boards* (IRBs)—that is, groups that review the proposed research for compliance with the federal

regulations governing research involving human subjects. IRBs are composed of researchers, nonresearchers, and representatives from the community at large.

The process of gaining IRB approval for conducting an evaluation is not a trivial issue but rather warrants serious attention. Even students who are conducting evaluation research need to obtain IRB approval of their research. Three levels of IRB review are distinguished (TABLE 16-2), with full review being the most comprehensive and involving all members of the IRB. Research may either qualify for expedited review by two members of the IRB or be exempt from IRB review but still be required to register as human subjects research. Although the exact procedures and forms vary slightly among IRBs, the responsibilities and reviews are essentially the same because all IRBs follow the same federal regulations—specifically, 45 Code of Federal

Regulations (CFR) 46. According to these regulations, informed consent has eight elements (TABLE 16-3), each of which must be addressed in the consent form. Ideally, the consent form will be written at an eighth-grade reading level. The eight elements can be addressed in brief letters for anonymous surveys or may entail extensive, multiple-page details for studies with risky procedures.

The highly formalized and bureaucratic procedures involved in obtaining IRB approval often cloud the underlying need to ensure that researchers and evaluators act ethically and that all persons involved in evaluation research be protected from needless harm. Nonetheless, if the evaluation is expected to generate knowledge, the evaluator is obligated to gain IRB approval. IRB approval must be obtained before the evaluation begins, which can be a major issue in the time line of conducting a program evaluation.

TABLE 16-2 Comparison of Types of IRB Reviews

Type of Review	Definition Criteria	Process
Full (review by all IRB members)	Involves more than minimal risk; involves knowing the identity of the participants and whether data are sensitive or may put the participant at risk	Requires completing a full IRB application; researchers must provide copies of all materials (i.e., surveys, consent forms, recruitment flyers, data abstraction forms, interview questions) that will be used in the research
Expedited (review by two IRB members)	Involves no more than minimal risk and may involve knowing the identity of the participants	Requires completing a full IRB application; researcher must provide copies of all materials (i.e., surveys, consent forms, recruitment flyers, data abstraction forms, interview questions) that will be used in the research
Exempt (from IRB review)	Involves no more than minimal risk; the identity of participants is not known or knowable; routine educational research; food-tasting research	A brief form that describes the research is completed and submitted; usually reviewed by at least one member of the IRB

TABLE 16-3 Eight Elements of Informed Consent, as Required in 45 CFR 46

1. Statement that makes it clear that the participant is being asked to volunteer to be in a research study (and can withdraw if he or she chooses)
2. Explanation of the purpose of the research
3. Description of the research procedures that details what is expected of the participant (e.g., tasks, length of participation in the intervention and the research, type of data to be collected)
4. Specification of the risks or discomforts that are possible or likely from being in the study
5. Explanation of the direct benefits to the participant, if any are possible or likely, from being in the study
6. Statement of confidentiality and information on how the study data will be maintained
7. Description of any compensation or payment for being in the study
8. Phone numbers where the participant can get more information about the research, information about his or her right as a research participant, and assistance if there is an injury or other problem resulting from being in the research

Ethics and Evaluation

Marvin (1985) stressed that evaluators have a responsibility to program participants and to evaluation participants to explain how the evaluation has the potential to harm them and future program participants. One point he made is particularly relevant to IRBs and consent forms: If the evaluation has potential implications for future programs and their availability, participants have the right to be aware of this fact.

Many health programs are designed for and delivered to children. In such circumstances, the evaluation may need to include collecting data from children. This practice can pose special ethical and IRB problems. Children who are old enough to understand that they are being asked to participate in a study must provide assent to be in the study; this would include their participation in evaluation research. The refusal of either the child or the parent to be part of the evaluation study must be honored. Matters become more complicated when the child becomes older and is increasingly capable of making decisions independently of his or her parents. For example, adolescents in safe-sex education programs may be willing to participate in an evaluation study but may not want their parents to know they are in the program. In this

situation, the evaluator must carefully consider the consequences and creatively develop means of including children without denying parental or child rights.

One major consequence of including children in evaluation research is that participation rates may decline when parental consent is required for participation either in the program or in the evaluation. Hussemann, Mortimer, and Zhang (2016) examined factors influencing parental consent. Contrary to expectation, socioeconomic variables were not associated with consent, instead, prepaid monetary incentives, being a biological parent, and having a parental history of study participation were associated with parental consent for children to be in a study. Depending on the research, parents give active permission to be in the study (they opt in) or passive permission not to be in the study (they opt out). Shaw, Cross, Thomas, and Zubrick (2015) found that using active consent led to a biased sample of children, meaning that children with behavior problems, from single-parent households, and in older age groups were underrepresented. Dynarski (1997) identifies additional complications by reminding researchers that consent is needed in experimental evaluations, as occurs when schools are randomized to experimental and control conditions.

Just as children are considered a special vulnerable population, so are many other vulnerable groups for whom health programs are designed and offered. The special circumstances of these groups need to be considered with regard to obtaining consent. For example, Gondolf (2000) found that, in evaluating domestic violence programs focused on the batterers, several issues needed to be addressed: maintaining the safety of the victim, tracking in longitudinal studies, and obtaining consent from both the present and past batterers and their corresponding victims. Taking precautions to address these issues resulted in a low refusal rate and few safety problems for participants. Evaluators can use the results of this and similar studies to estimate more closely the highest possible participation rates and to identify ethical techniques for increasing the participation of children and other highly vulnerable groups in health program evaluations.

Another vulnerable population comprises those in need of emergency care. In anticipation of needing consent for the evaluation of an emergency treatment called PolyHeme, Longfield et al. (2008) took the unusual step of assessing the preferences of community members. They found that, despite knowing about the potential benefits of the treatment, 30% would choose not to participate in its evaluation. In population-based programs, the issue of obtaining consent will present challenges that requires creativity and sensitivity to overcome.

Evaluators and program managers may also need to deal with ethical issues related to program staff members. Such ethical issues emerge when the health program involves community outreach, and thus, the safety of program staff members is a major concern. Also, if program staff members are participants in the program evaluation, as is likely to occur in process monitoring evaluations, their rights as study participants need to be taken into account. These rights include confidentiality with regard to the information they provide. Program personnel may themselves face ethical issues with regard to legally mandated reporting of occurrences, such as child abuse and specific infectious diseases (e.g., syphilis, tuberculosis)

that must be balanced against their safety or the safety of others involved.

Other factors contributing to ethical issues are financial arrangements, conflicts of interest, level of competence, and deadlines. Each of these factors has the potential to create an ethical dilemma. Another consideration is highlighted by the findings of one study in which evaluators who were employed in private or consulting businesses had different views of what constituted unethical behavior compared to evaluators who were employed in academic settings (Morris & Jacobs, 2000). These results serve as a stark reminder that ethics is both contextual and highly individualized. While principles may be agreed upon, their application may not. Evaluators working closely with stakeholders are likely to encounter a wide range of opinions about ethical behavior. As always, open dialogue and discussion of the ethical issues inherent in the health program or evaluation are the optimal approaches to reaching either consensus on actions or comfortable disagreements.

HIPAA and Evaluations

In 2003, the Health Insurance Portability and Accountability Act (HIPAA) went into effect (Centers for Medicaid and Medicare Services, 2005; Department of Health and Human Services, 2008). The purpose of HIPAA is to protect personal information related to having received health care. Personal information includes the individual's birth date, any element of an address, dates on which services were received, and diagnoses. Health providers must take specified steps to protect personal identifier information, including having secure fax lines and e-mail transmission mechanisms, and getting written permission to share personal information with others, whether insurance companies or consulting providers.

The effects of HIPAA on the evaluation of health programs depend on whether the evaluator is an employee of the organization and the evaluation is part of routine care, or whether the evaluator is an outsider and the

findings of the evaluation will be made public in any way. In the first situation, the evaluation is likely to be small scale and involve an existing program being conducted by program staff members. Because personal information about clients or patients is not being shared beyond the organization, evaluators need not take any additional steps with regard to the HIPAA regulations. This situation is similar to the one in which an informed consent would not be needed because the evaluation is not research. However, if client data will be provided to an external evaluator or if the evaluation is federally funded (National Institutes of Health, 2003), then the HIPAA regulations require that each client be informed that his or her information is being provided to others and give a signed authorization for that information release, unless the IRB allows an exemption on the basis of topic importance and infeasibility of gaining individual consent.

HIPAA regulations apply only to a specific group of healthcare organizations, typically hospitals, clinics, and physician offices. They do not apply to organizations that do not provide medical care. Thus, health programs provided by the latter organizations do not fall under the scope of the regulations, and evaluations of programs in exempt organizations do not require steps beyond the basic ethical considerations discussed earlier.

► Responsible Spin of Data and Information

It is incumbent upon the evaluator to disseminate information about the evaluation to complete the feedback cycle. Several considerations influence the dissemination of program evaluation findings. The first concern focuses on making a persuasive argument based on the data. Once a set of persuasive statements about the program is developed, the next consideration is the logistics of the report format. Finally, evaluators must consider the possibility that the evaluation results

will be misused. Each of these considerations is discussed in the following subsections.

Persuasion and Information

A major factor determining the effectiveness of feedback loops is the persuasiveness of the data and the resulting information. Although the word *persuasion* has negative connotations, the reality is that, for change to occur and for decisions to be based on evaluation findings, individuals with decision-making authority need to be persuaded by the “facts.” Unless the statements made based on the process evaluation and outcome evaluation data are persuasive, no feedback loops will be possible; instead, the program will produce only discrete, disconnected, and unused information. As the availability of so-called fake news has increased, so has the ethical obligation of program evaluators to convey only true and substantively meaningful findings.

When using statistical results, five properties of statements affect the persuasiveness of statements (Abelson, 1995): effect magnitude, articulateness, generalizability, interest level, and credibility. Magnitude of the intervention effect (*effect magnitude*) is the amount of change attributable to the program. The amount of change that can be attributed to the program influences the extent to which the statistical data are perceived as persuasive. One way to think about this issue is to note that people are more inclined to believe good news; therefore, larger intervention effects increase the persuasiveness of evaluation findings. A corollary of this idea is that people are less likely to be persuaded by small program effects, regardless of the potential clinical significance of a small effect. Although it would be unethical to conduct an outcome evaluation that is highly biased in favor of the program, the design and methods need to be such that the largest effect theorized can be identified.

Another property of a statement is the degree to which it is *articulate* in explaining the findings (its *articulateness*). Articulate statements have both clarity and detail. This

speaks directly to the quality of the written or verbal report, sometimes referred to as style. Being articulate may require not using statistical terminology in reports intended for stakeholders and policy makers. Recipients of the findings must perceive that the report is clear, regardless of whether the producer of the report feels it is clear. Clarity can be achieved through style, language, use of graphs, and careful delineation of ideas and issues. Similarly, consumers of the evaluation report may have a different need for details compared to the producers of the evaluation report. One strategy for arriving at the optimal level of detail needed to be persuasive is to have someone not involved in the evaluation review the report and make suggestions about what needs to be included. In other words, evaluators should involve a few stakeholders in the development of any written reports that will be disseminated to their peers.

The third property of a statement that contributes to its persuasiveness is the *generalizability* of the conclusions. The wider the applicability of the findings, the more likely those findings are actually persuasive. The notion that having highly generalizable findings makes the findings more persuasive runs counter to a more popular line of reasoning in the health field—namely, that results must be tailored to the unique characteristics of the program participants, especially race, ethnicity, sexual orientation, and such. It is imperative to consider the effect of the unique needs and perspectives of the program recipients to have an effective program; however, when it comes to influencing decisions, policy makers want information that will make the decision widely applicable. As a consequence, they often seek results that are generic rather than specific. In addition, there is an intuitive sense that if something applies to many, it must reflect some underlying truth. Whether this line of reasoning is accurate is less important than acknowledging that, as human beings, we may instinctively think this way.

The fourth characteristic of persuasive statements is the degree to which the findings are *interesting* (their *interest level*) with regard

to challenging current beliefs. When the findings are a surprise, more attention is paid to them. Similarly, if the findings make important claims or deal with important behaviors, the statements will receive more attention and be more persuasive. Another way of thinking about this characteristic is that if the evaluation does not answer the question, So what? or merely confirms the obvious, the findings will be less persuasive. Findings that are surprising, unexpected, or not immediately explainable and those that are important become sound bites that are used to get attention and influence perceptions.

If the findings have *credibility* in terms of being based on rigorous and sound methods and a coherent theory, the findings will be more persuasive. The bottom line: If the evaluation methods, design, data collection, or analyses are flawed, the findings lose credibility and persuasiveness. Of course, a coherent theory is also needed for the findings to have credibility and persuasiveness. A coherent theory, such as a causal theory, explains the findings in terms of relationships among the program intervention and outcome and impacts. Such a theory decreases the possibility of claims that findings happen by chance.

Patton (2008) suggested that both the strength of the evaluation claim, like the first of Abelson's statement properties, and its importance to the decision makers are crucial to the evaluation being of value. If we extend Patton's two dimensions to encompass intervention effect and interest values, the result is a matrix demonstrating that the influence of these characteristics on decision making varies (TABLE 16-4). Unimportant claims that are not interesting and that reveal a low level of intervention effect based on weak rigor are quickly forgotten. Claims that are important and interesting, are based on good science, and show large program effects are readily used as the basis for decision making. The matrix hints at the challenges that must be overcome to generate claims that will be used in decision making, including subsequent program or evaluation decisions.

TABLE 16-4 Effect of Rigor and Importance of Claims on Decision Making

Strength of Claim	Quality of Claim	
	Major Importance, High Interest	Minor Importance, Low Interest
High Rigor	Ideal for making decisions	Becomes factoid
Low Rigor	Cause for concern, need to study further, tentativeness to decisions	Ignored, forgotten, unspoken

Modified from Patton, M. Q. (2008). *Utilization focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Information and Sensemaking

Steps taken to turn data into persuasive information are important, but so is the process by which sense is made of information. *Sensemaking* (Weick, 1979, 1995) involves attributing meaning to information. Understanding the psychological processes used by individuals to interpret information in order to make decisions is the other side of the persuasion coin. Data become information only when they are given meaning. The process of giving meaning involves various human perceptual and judgment-making processes; these processes in turn ultimately influence the decision-making process and, therefore, are important for health program planners and evaluators.

Several well-recognized phenomena are now widely accepted as affecting decision making, including the unconscious use of a set of heuristics in making judgments (Tversky & Kahneman, 1974) and the use of hindsight bias and consequences to make judgments (Chapman & Elstein, 2000; Fischhoff, 1975), which is also called retrospective sensemaking (Weick, 1979, 1995) or the one-of-us effect (Hastorf & Cantril, 1954). Each of these cognitive-processing phenomena has potential consequences for program planners and evaluators.

For example, Lipshitz, Gilad, and Suleiman (2000) studied the effect of having information about the outcomes on the perception of events.

They found that the perception that the subject was “one of us” was dramatically altered depending on whether the subject’s success or failure was met with reward or punishment. The findings from their study underscore the importance of the sociocultural context of decision making, especially with regard to situations in which there are clear failures and successes and when both rewards and sanctions are possible. These findings have relevance to the presentation of evaluation information. They suggest that if the evaluator is not viewed as “one of us,” then failures of either the program or the evaluation are more likely to be met with sanctions. In other words, evaluators who become aligned with the stakeholders are less likely to experience repercussions. This same psychological phenomenon applies to program staff members: If the policy and organizational decision makers view program staff members as “one of us,” then the program staff members are less likely to experience sanctions if the program is not successful.

Culture is another factor that can affect interpretation of data. Culturally based values, ways of thinking, and experiences influence the meaning attributed to results. One key area in which this factor may become evident is in discussions about which program changes need to occur. Just as there is a need for cultural competence in the design of the health program, so there is a need for cultural competence in

terms of organizational competence to make or guide the implementation of program changes based on data about the program theory or the effect theory.

A recognition of the cognitive elements in interpreting data, whether related to decision making or culture, highlights the intricate nature of facts and interpretations. It is easy to believe that having factual data will always result in a detached, logically derived decision. In reality, evidence from decision science and psychology (not to mention one's own experience) does not fully support this proposition. Evaluators and program planners need not only to attend to the science of their work but also to apply the science of decision making and attribution.

► Reporting Responsibly

Report Writing

Evaluators are required to generate written reports and perhaps to make oral presentations of the findings. The notions of persuasiveness and the malleable nature of interpretations are part of the context for these reports. Typically, the format of a report is tailored to the evaluation and overall situation. In some instances, however, the format is very specific and the evaluator has minimal flexibility in the report format, as is the case with state or federal grants. Often, reports related to federal programs are submitted over the Internet, using predetermined forms that place strict limits on the number of words or characters. If the evaluator is not the one submitting the forms, it is critical to work with program personnel in advance, so that they have the information required for reporting to the funding agency. In addition to the reporting required by the funding agency, the evaluator may be asked to prepare a longer narrative report.

To the extent possible, a report should relate findings to the program process and outcome objectives. This linkage might utilize a table format, in which each objective is listed and the

corresponding evaluation results are presented. Such a table provides a visual context in which the audience can clearly see in which areas the program was more and less successful in meeting its objectives.

Evaluation reports generally contain an executive summary, background on the program and evaluation, a description of the program, a description of the evaluation, and a summary and recommendations. The content and style used for writing the report should be tailored to its audience. Gabriel (2000) has argued that, because of the ever-evolving nature of the evaluations of health programs, evaluation reports and reporting strategies should forgo the usual research format that emphasizes data analysis and instead focus on providing evaluation information to the decision makers and program stakeholders as expediently and accurately as possible.

An *executive summary* is a one- or two-page synopsis of the program, its evaluation, and the recommendations. This portion of the report is most likely to be copied and distributed and, therefore, is likely to be the most widely read part. It should contain enough information to convince any reader that the program and evaluation designs were appropriate. The length of the executive summary depends on its anticipated use and reporting requirements.

As mentioned earlier, evaluators are often asked or required to give an oral presentation of the evaluation plan or results. The first consideration when preparing such a presentation is the audience: Will it be staff members, program participants, policy makers, or other researchers? The answer to this question influences the content to be covered as well as the choice of which details will be of greatest interest to the audience. Use of software, such as Microsoft PowerPoint, can greatly facilitate generating an outline of the presentation and can simplify making appropriately designed slides. The material on each slide should be displayed in large print, with no more than one idea per slide and no more than seven or eight lines of text. Audiences have also become accustomed

to receiving handouts of the presentation, but carefully consider what is on the handout and whether it is wise and timely to make that information widely available. If the evaluation has been done in a manner that has involved the stakeholders throughout the process, there should not be any surprises, and release of the information will not be an issue. In less ideal situations, caution may be in order.

The use of charts and graphics is important when writing reports. The old adage that a picture is worth a thousand words is true: Exhibits are more likely to be remembered than text. Although not every number needs to be graphed, having no graphic representations makes reports dense, potentially boring, and less persuasive. All exhibits—whether graphs, charts, figures, or tables—should contribute to the ease of understanding and the remembering critical points rather than being only fancy additions that demonstrate the author's skill with a software program. The rationale for including the exhibit needs to be made explicit. The choice of what is presented graphically is in part a way to influence perceptions, attention, and hence decision making.

A list of suggestions for creating good graphics, drawing on the recommendations given by Good and Hardin (2012), appears in **TABLE 16-5**. Findings that are surprising, confirming, alarming, or interesting are appropriate choices for being illustrated with exhibits. Readily available software makes generating exhibits and integrating them into written reports quite easy. Exhibits should follow convention regarding which variable is on which axis. For example, time is best plotted on the horizontal axis (*x*-axis). It is essential that some explanatory text accompany all exhibits because not all readers will be able to interpret graphic presentations. For some reports, it may be necessary to be creative in the presentation of findings. For example, Dyer, Goodwin, Pickens-Pace, Burnet, and Kelly (2007) used a Venn diagram to convey the proportion of individuals with abnormal assessment scale scores and to illustrate how proportions of individuals with overlapping (concurrent) abnormal scores were distributed. Their use of multiple overlapping circles was needed to show that, of the 22% of individuals with depression, 8% also had inadequate capacity for activities of daily living and impaired cognitive functions.

TABLE 16-5 List of Ways to Make Graphs More Interpretable

- Only display graphs for ideas that are easily explained with visual aids. Limit graphs to those that show key variables and crucial ideas.
- Use only as many dimensions in the graph as there are dimensions in the data. For this reason, do not use three-dimensional bar graphs.
- Place the labels where they do not interfere with seeing the data. For example, keep all axis labels and tick marks on the outermost sides.
- Select carefully the range of the axis values to be realistic and reasonable. Choose a range of values closest to those in the data.
- Use numerical labels only if the data are numerical. For categorical or nominal data, give the category title.
- Create line graphs only if the data points reflect a trend.
- Give your graphic a full title that explains the contents of the graphic. Titles need not always be short.
- Choose variables and scale carefully because these items are much more important than choice of color scheme, except to the extent to which it may be useful to choose colors that will remain distinct when printed in black and white.
- Spend most of your time choosing and adjusting the numeric information.

Making Recommendations

Making recommendations, whether in a report or verbally, can tax the interpersonal and political skills of an evaluator. Recommendations can focus on the positive aspects that were identified through the evaluation as well as on areas needing change or improvement. They must be based on, and be a direct and clear outgrowth of, the evaluation data. Recommendations should be

linked to the program theory and should address the organizational and service utilization plan (FIGURE 16-1), identifying specific elements in need of improvement. The recommendations can also address elements of the program theory (Figure 16-1) by specifying which of the hypotheses seem not to be supported by the evaluation data and indicating alternative hypotheses that could be used to explain the findings. Framing

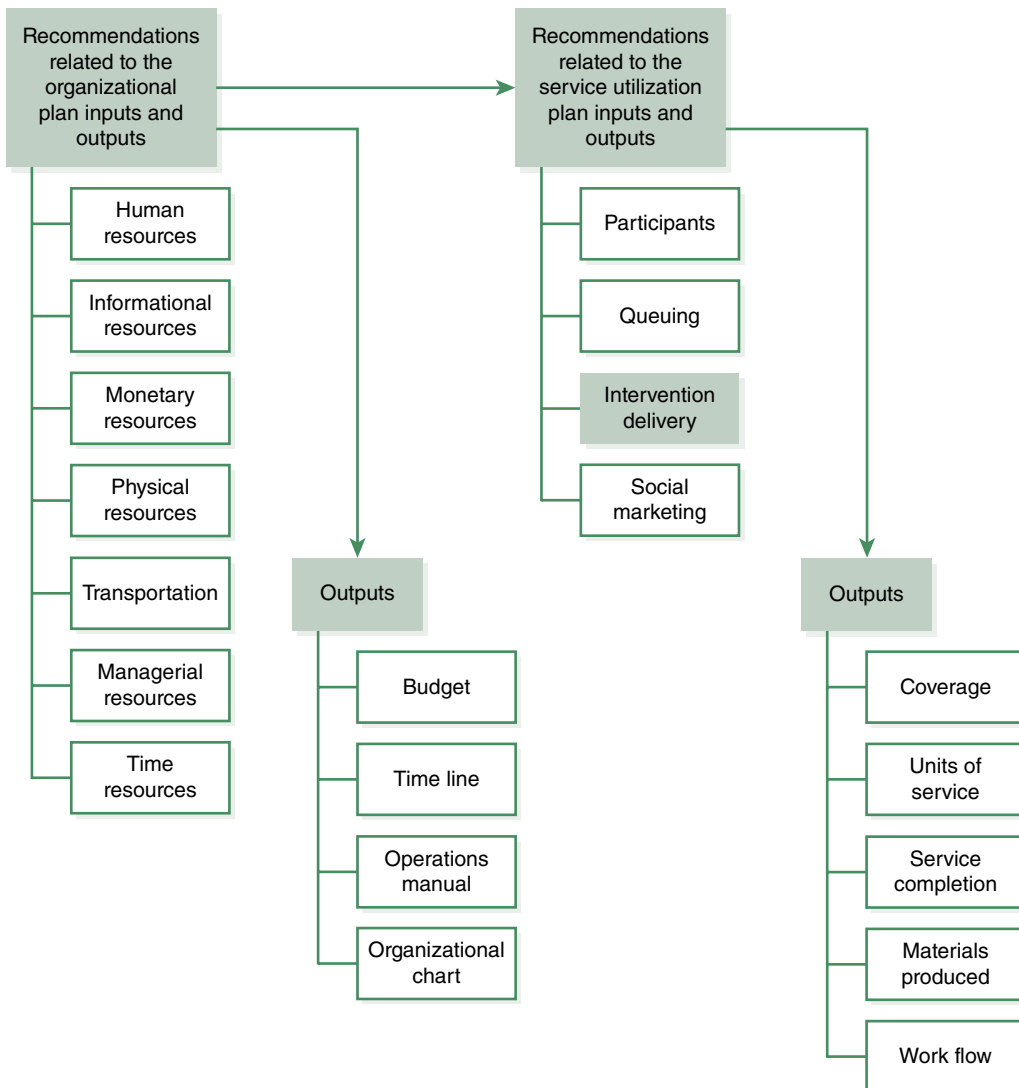


FIGURE 16-1 Making Recommendations Related to the Organizational and Services Utilization Plans

the recommendations according to the way the program was conceptualized helps program managers and other stakeholders make decisions regarding what needs to be or can be done and in what order of priority (FIGURE 16-2). Also, recommendations that are linked to elements of the program theory are more likely to be readily understood and hold higher credibility than those not linked to program theory.

The information in the following paragraphs regarding recommendations frames the dos and don'ts suggested by Hendricks and Papagiannis (1990) within the context of program theory.

Throughout the development and evaluation of the health program, recommendations can be collected, so that good ideas and insights are not lost or forgotten. Also, a wide variety of resources can be the basis of making recommendations,

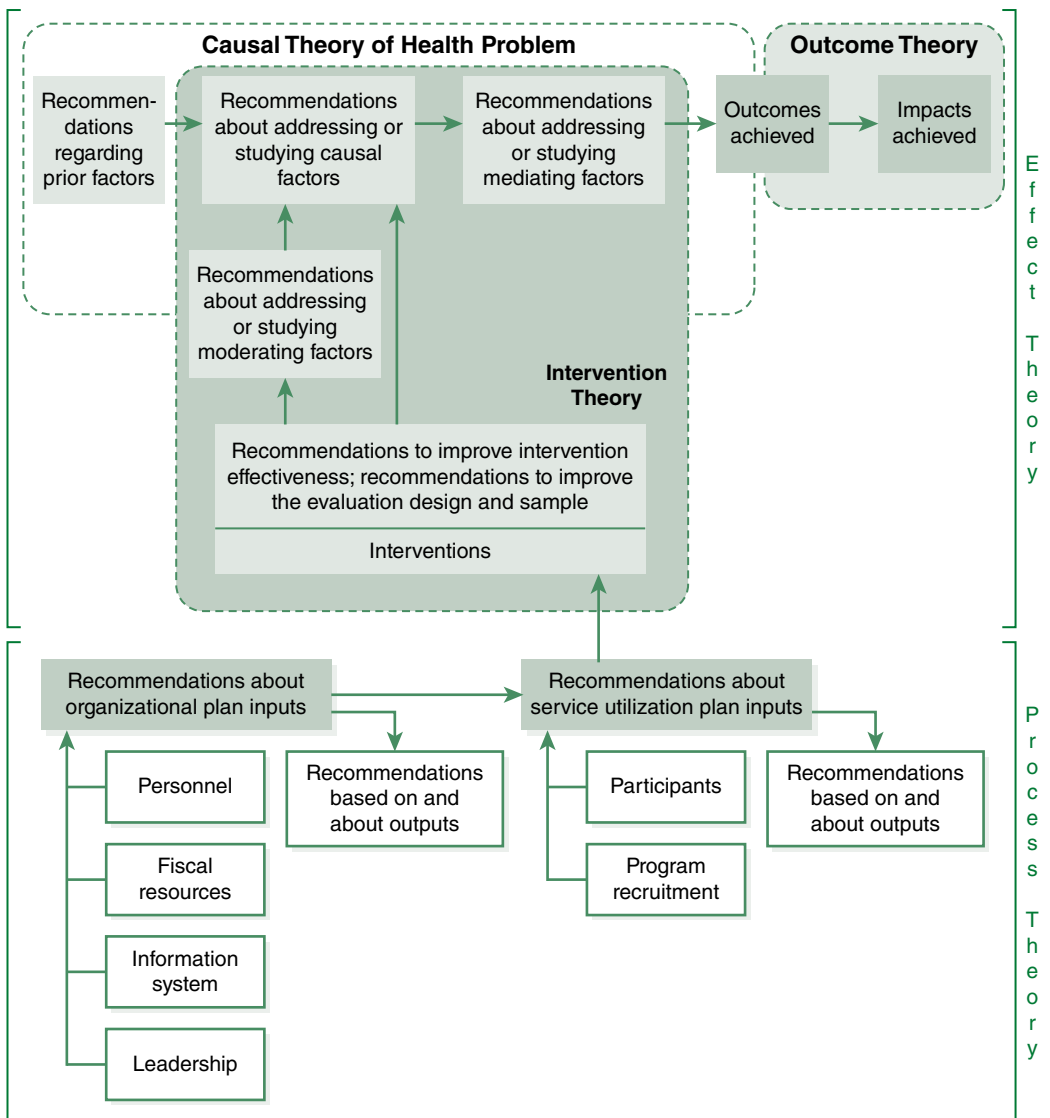


FIGURE 16-2 Making Recommendations Related to the Program Theory

including existing scientific literature, program staff members, and program stakeholders. Drawing on such sources for explanation or justification of the recommendations adds to their credibility. Working with the program decision makers, including during the process of developing the recommendations, minimizes the possibility of surprises being included in the report, thereby decreasing the risk of being perceived as undermining the program. Furthermore, working with the decision makers also begins to build support for the recommendations.

Recommendations should make sense in the larger context in which the program and the evaluation occurred, including in the social, political, and organizational contexts; in the context of the technical knowledge available when the program was conceived; and throughout the planning and evaluation cycle (FIGURE 16-3). Recommendations that show knowledge of the contextual influences can easily be accepted as

applicable and may lead to changes in various elements of the program theory. Recommendations that are realistic in the view of the stakeholders are more likely to be accepted and implemented than those not viewed as realistic. Of course, this does not negate the need to make appropriate recommendations that may take time to implement or that could be implemented under different circumstances. Iriti, Bickel, and Nelson (2005) proposed a continuum of recommendation from providing a set of options to supporting one set of recommendations over others, to recommending one option and how to implement that option. The levels of recommendations reflects that not all recommendations need to be equal, including in terms of the effort required to implement them, or in terms of specificity and generality. Decision makers in general prefer choosing from a set of recommendations so that they have options and can exert influence over the program.

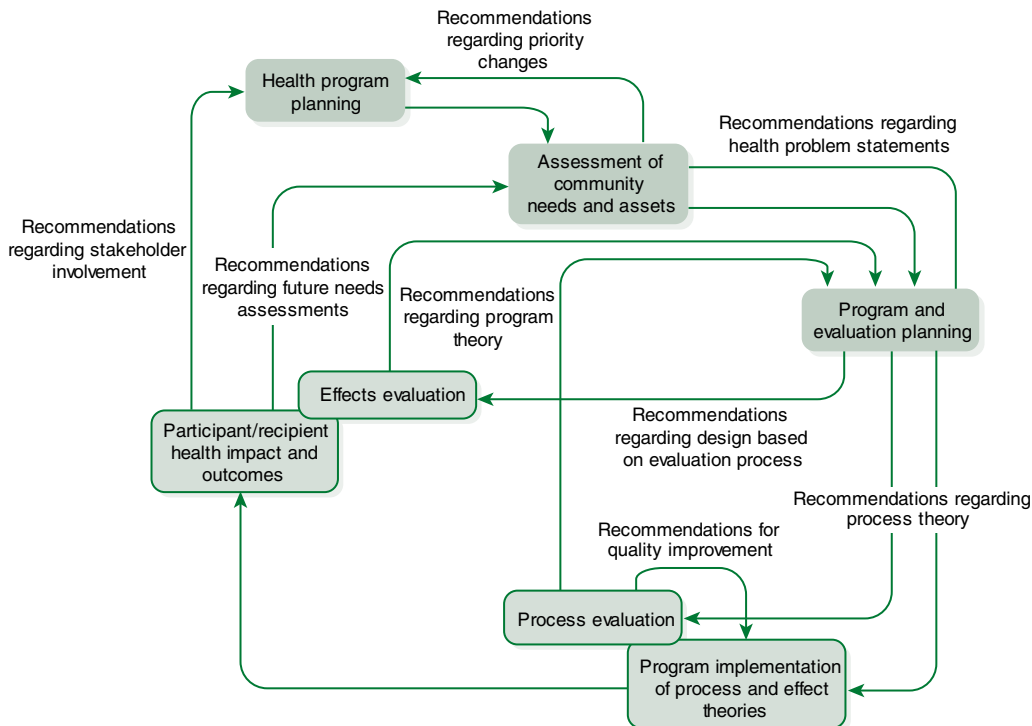


FIGURE 16-3 The Planning and Evaluation Cycle with Potential Points for Recommendations

Recommendations that require a fundamental change, such as to the program theory, are less likely to be accepted or implemented than recommendations that call for less extensive change. The former recommendations are best made in an incremental format so that they are simplified and hence are more acceptable and useful to the decision makers. It also helps to provide some accompanying statements or insights into future implications of the recommendations, such as potential benefits or implementation strategies. Again, this type of information facilitates the decision making regarding which recommendations to act upon. Of course, recommendations that are easy to understand will receive more attention and acceptance than those that are not. Techniques for improving the ease of comprehension include categorizing the recommendations in meaningful ways, drawing a boundary between the findings and the recommendations, placing the recommendations in a meaningful and appropriate position in the report, and presenting the recommendations in ways that the decision makers are accustomed to receiving information.

Misuse of Evaluations

Given the pressures to sustain programs and to meet funding mandates, evaluation results may sometimes be misused. Misuse can occur in many ways but is generally understood to consist of the manipulation of the evaluation in ways that distort the findings or that compromise the integrity of the evaluation. Misuse can occur at any point during the evaluation process, beginning with an inappropriately chosen design that biases findings and extending to interference with data collection and the outright alteration of the evaluation findings (Stevens & Dial, 1994). Evaluators, program managers, program staff members, stakeholders, and program participants involved in an evaluation can all be susceptible to a variety of pressures that may result in a less than unbiased evaluation. These pressures include biases based on resource constraints and self-interest, as well as external political and funding agency pressures. Pressures can also

come from within the organization providing the health program. Such pressures affect not only the evaluation design and execution but also the manner in which the results are interpreted, used, and disseminated or suppressed.

Patton (1997) has argued that increases in the use of evaluation information are accompanied by increases in the misuse of that information. This relationship is a function of the increased opportunity for the pressures leading to misuse to be present. Program planners and evaluators must recognize that misuse can be either intentional, such as falsifying client satisfaction responses, or unintentional, such as not having enough time to distribute client satisfaction surveys on busy days. Sometimes the pressures for misuse are subtle and the responses are almost subconscious; in fact, the intention might even be to help the program. Such complexities add to the difficulty of detecting and avoiding misuse.

One way to minimize the potential for misuse is to involve program stakeholders and educate them about the evaluation and its appropriate use. For example, it may be important not only to publicize favorable findings but also to acknowledge unfavorable findings and the plans to address those findings. In addition, sharing with the stakeholders and program staff members the standards to which evaluators are held by their professional organization may help them realize that the use of evaluation is not just a funding or sustainability issue but also an ethical issue.

Of course, it is equally possible that stakeholders may intentionally suppress, ignore, or discredit evaluations that are not favorable. This reality gains the most visibility in the health policy arena. For example, Mathematica, a private research firm, was hired by the federal Maternal and Child Health Bureau (MCHB) of the Health Resources Services Administration to evaluate the effect of the Healthy Start Initiative programs funded by the MCHB (Howell et al., 1997). The Healthy Start Initiative funded local programs designed to reduce infant mortality and the rate of low birthweight births; each local Healthy Start program had a local evaluation.

Mathematica evaluated a range of programmatic interventions in more than 20 locations, using much of the data from the local evaluations in addition to other data sources. The Mathematica meta-evaluation revealed a lack of evidence that the Healthy Start programs had an effect on the rates of infant mortality or low birthweight; however, these findings were not used by the MCHB in subsequent requests to Congress for funds for the Healthy Start Initiative. This story illustrates the tension that exists between health policies, which may be driven by contradictions between beliefs about what will work and the “cold, hard facts” of both poorly and well-done evaluations. Political considerations involved in situations like these can be problematic. Regardless of the source of the political issues, planners and evaluators will encounter the occasional unexpected ethical land mine.

Involving and Negotiating with Program Partners

Ensuring dissemination of both positive and null findings from evaluations may first require some negotiation with program partners. Specifically, in addition to addressing the ethics of obtaining informed consent for participation in the evaluation, evaluators need to negotiate authorship, publication outlet, and manuscript content before submitting a manuscript for publication. The negotiation on these and similar points proceed more smoothly if program partners and stakeholders have been involved throughout the evaluation and report preparation phases. Even so, egos and reputations may be at stake, so additional attention and negotiations may be necessary.

► Responsible Contracts

Organization–Evaluator Relationship

The relationship between the evaluator and the organization is one that requires attention, whether the evaluator is an employee or an external

consultant. Several factors can influence or shape this relationship. For example, if the evaluator is a third party between the program or organization and the funding agency, third-party dynamics come into play. That is, two parties may form a coalition against the third. The most likely scenario would be for the program manager to align with the funding agency in any dispute with the evaluator. Another factor influencing the organization–evaluator relationship is the different levels of priority given by the evaluator and the organization to the health program and the possible changes in these priorities. Such differences in priorities can lead to different time lines, different expectations, and different levels of commitment in terms of attention and resources. Yet another factor affecting the relationship between the evaluator and the organization is differences in professional and organizational cultures and in the language used to describe the health program and the evaluation. Divergences in any of these areas can become a source of tension and conflict, especially if the divergence is not explicitly acknowledged. This list of possible factors affecting the relationship of the evaluator and the organization of the health program, although certainly not exhaustive, reveals the complexities of the relationship, which deserve attention and merit the development of strategies and dialogues to optimize the relationship.

Another area of possible divergence is the broader purpose of the evaluation. Owen and Lambert (1998) have suggested that evaluation for management differs from evaluation for leadership. They claim that evaluation for management focuses on individual units or programs for the purpose of making ongoing adjustments with an emphasis on performance and accountability; this is essentially a process-monitoring evaluation. In contrast, evaluation for leadership focuses on the whole organization, with a long-term perspective and collection of information across programs.

Owen and Lambert (1998) also note that the roles of evaluator and organizational development consultant become blurred, particularly if the evaluator is asked to provide recommendations.

In fact, Patton (1999) has argued that evaluators have numerous opportunities to act as organizational development consultants. He explains that, because programs are embedded in organizational contexts, improving programs may require attention to that organizational context. Patton found, through a case study, that the attainment of project goals is not necessarily related to the extent to which the program contributes to the fulfillment of the organizational mission. In other words, some programs are very good at doing the wrong thing when viewed from the perspective of the organization's business strategy. Being aware of this potential paradox and addressing it is one way in which evaluators contribute to organizational development.

Evaluators may see themselves, or be seen by others, as agents for organizational change. As change agents, evaluators need to be conscious about how change occurs; a prerequisite for change is organizational readiness. The type of organization being changed is an additional consideration—whether the organization is a small, community-based, nonprofit agency or a large federal bureaucracy. Evaluators can influence organizational decision making by first having affected or changed the perceptions held by those who are decision makers for the program. This can occur through the recommendations made at the end of an evaluation, or through continual rather than sporadic involvement with the program. Education of stakeholders throughout the planning and evaluation cycle, and through personal perseverance with regard to dissemination of information and accessing decision makers, can also contribute to some degree of organizational change. In the end, the evaluator's skills in generating an understanding of what is needed from the organization must be matched to the organizational type, culture, and readiness for change.

An early point of discussion between the planner/evaluator and the hiring organization needs to focus on who owns the data that are gathered and analyzed throughout the planning and evaluation cycle. Generally, planners and evaluators who are hired as consultants have very limited

proprietary rights over the data. The contract with the hiring organization specifies who owns the data and what can and cannot be done with it. An external evaluator needs some leeway and latitude so that potentially unfavorable findings can be seen by those who need to know about them. Getting the information to the decision makers is essential if the evaluation—whether a process monitoring or effect evaluation—is to serve as an effective feedback loop that fosters the development of the health program and the organizational functioning. Evaluators who cannot express concerns to key decision makers will be frustrated and are essentially divested of their proper responsibility. Ownership of data and dissemination of findings can become not only an internal political issue but also an ethical issue in terms of getting the results to decision makers and to making the research findings public, which can influence choices. The underlying issue is that evaluators may have a different perspective and interpretation of the evaluation results than members of the hiring organization; this divergence may lead to conflict and tension.

The organization is also a context for ethical dilemmas. Mathison (1999) has suggested that there are minimal differences in the ethical dilemmas faced by internal and external evaluators. However, the two parties live in different environments, which inevitably influence their actions or responses to ethical situations. Ultimately, evaluators need to be clear about their relationship with the organization and their role vis-à-vis ethical factors.

Health Policy

The interface between program evaluation and health policy can be intricate. Health policy becomes the focus of action in two general situations. One situation arises when an independent program evaluation has been done and the stakeholders are attempting to use the evaluation findings to change health policy. VanLandingham (2006) suggests that two strategies are needed to increase the use of evaluation results for making policy changes. One strategy is to develop communication

networks that allow for exchange of information with legislative stakeholders, including enabling their input into the evaluation design and process. The other strategy is to communicate the evaluation results effectively and in a manner that is readily understandable and actionable, as was discussed earlier. Actionable evidence is a potential sticking point, as Julnes and Rog (2007) note. Each source of evidence, whether from outcome research, literature synthesis, or practice, has imperfections that make action based on that data less than foolproof. These sorts of flaws create a barrier to using evaluation results for policy making or legislation.

The other situation in which health policy becomes the focus or action arises when the evaluation is done under contract to a governmental agency or a government-sponsored program. In this situation, there can be pressures related not only to the way the evaluation question is framed but also to the methodology used in the outcome evaluation (Chelimsky, 2007). At the heart of these pressures is the political agenda for finding supportive or disconfirming evidence for the program. The pressures to uphold current policy or to provide actionable findings for making policy change can be seen most easily with population-level programs that are governed by health policy, such as Medicaid, Medicare, SCHIP, WIC, and other federal programs. Given the longevity and popularity of these programs, the pressures might be relatively small when compared with evaluations of more controversial programs, such as the Just Say No program or abstinence-only programs. Program evaluators who are faced with these pressures may need to make both politically and personally uncomfortable decisions about how to proceed.

► **Responsible for Evaluation Quality**

There are three main approaches to assessing the quality of a program evaluation. One approach is to determine whether the evaluation met the

program evaluation standards of the American Evaluation Association (www.eval.org). The Joint Commission on Standards for Educational Evaluation of the American Evaluation Association has created four standards, or criteria, for evaluation practice—utility, feasibility, propriety, and accuracy—that remain cornerstones for assessing the quality of health program evaluations. The 4 American Evaluation Association criteria together constitute one of the 3 potential approaches described here. To achieve the feasibility standard, the evaluation must be developed and implemented with the same careful planning as was required to develop and implement the health program. Evaluations must be conducted in ethical and legal ways that are unbiased and sensitive to the vulnerability and rights of all program and evaluation participants, which requires attention to ethical, legal, and moral parameters. Evaluations must also collect appropriate and accurate data that are consistent with the purpose of the evaluation and the health program being evaluated. According to this perspective, the scientific rigor of the evaluation is only one of the four criteria (that is, accuracy) by which a health program evaluation can be assessed.

A second approach is to confer with the users of the evaluation for their perspective on its usefulness. Evaluations need to be useful in meeting the needs of the program stakeholders, program audiences, and funding agencies. This requires communication, dialogue, and negotiation. Evaluations must be feasible in terms of cost, political and diplomatic factors, and time lines.

The third approach is to assess the evaluation in terms of its scientific rigor. The elements of a rigorous evaluation are presented throughout this text. One element is that the conceptual foundation of the evaluation should be stated explicitly, including the use of biological, social, psychological, or other theories relevant to the program effect theory and references made to previous evaluations of the same or similar health problems. Another element is that the evaluation question should be stated clearly and should focus on identifying the effects of the program. This emphasis on effects and

outcomes is consistent with the purpose of meta-evaluations and evidence-based reviews to identify the most effective intervention for the stated health problem. Next, the evaluation design must be valid, credible, and feasible. In other words, the scientific rigor of the design must be evident in the published findings or the report. A high-quality evaluation of outcomes also includes attention to multiple health domains of the outcome, using existing reliable measures and multiple modes of data collection when necessary or scientifically justified. The attention to multiple domains stems from the multiple components and multiple facets of health that might be affected by the health program. An optimal design includes random assignment to the program but does not compromise the provision of the health service to those not receiving the program. It also includes careful selection of program sites and careful attention to incentives for participating in the evaluation. The data analysis methods must also be appropriate for the data collected and the evaluation question.

Mohan and Sullivan (2006) summarize key characteristics (explicit or implicit) across the various standards that apply to evaluators: Maintain independence, report conflicts of interest, maintain confidentiality of information, and keep records of the evaluation process. Of course, each of these elements can be addressed to varying degrees, and rarely does any single evaluation address all elements to perfection. These criteria also provide an outline that evaluators can follow to avoid evaluation failure and that should serve as a reminder of the challenges inherent in program evaluation.

The term *meta-evaluation* has been used since at least 1982 (Feather, 1982). One type of meta-evaluation focuses on programs either funded by or provided by a particular agency; such an evaluation seeks to understand the overall effectiveness of programs associated with that agency. Agency-based meta-evaluations are done using existing evaluations conducted by programs related to that agency. Major philanthropic foundations and large public health agencies are the types of agencies likely to engage in meta-evaluation.

► Responsible for Dissemination

Dissemination refers to the intentional and systematic distribution of protocols, findings, and recommendations in a manner that makes that information accessible to a specific audience. Most recommendations and reports are intended for internal program use and are distributed only to key program stakeholders and staff members. In contrast, broader dissemination of lessons both about implementing the program and its effect on recipients can be important for the advancing the science and practice related to the health problem. For programs that were the subject of evaluation research because they were novel or experimental, the findings from those evaluations should be disseminated broadly as a contribution to the science and the literature. Equally important is dissemination of information to ensure program fidelity and to serve as general guidelines for subsequent program development.

Related to dissemination is the diffusion or replication of programs, as well as the sustainability or expansion of existing programs. Fisher et al. (2007), for example, reported that the findings from 14 demonstration projects related to diabetes self-management would be disseminated in hopes of the programs being implemented more broadly across clinical settings. In contrast, the study by Hill, Maucicone, and Hood (2007), which explored adaptations of programs, found that limits on time and resources pose serious threats to the replication of programs in a manner that maintains fidelity to the interventions. As a program moves into a new planning cycle, reassessment of the local capacity to deliver the program may be in order. For example, Alfonso et al. (2008) conducted a capacity assessment to determine the sustainability of a physical activity program for children. Through community involvement, they were able to identify strategic activities necessary for program sustainability. Their work reinforces the cyclic nature of program planning.

Broader dissemination includes making the report more widely accessible—for example, at the program website or the website of the evaluator. It also includes publishing the findings in academic and practice journals so that a broader audience can become aware of the effectiveness of the interventions as well as the issues involved in implementing the program. Having the evaluation published in peer-reviewed and practice journals helps ensure that the evaluation findings become part of the evidence to be used in future program development efforts, especially when the journals are publicly accessible without requiring paid subscriptions or library access. As Gorman and Conde (2007) point out, however, a conflict of interest may arise when the published program evaluation was conducted by the program developer; thus, the consumer of published evaluation reports should be sensitive to this possibility. Besides publication in academic journals, many other modes of dissemination are available to the evaluators and their stakeholders. Those modes, along with the breadth of their reach and intended audiences, are summarized in **TABLE 16-6**.

► Responsible for Current Practice

The discipline of evaluation continues to evolve, with the constant emergence of new concepts, updates to approaches, and development of new models that synthesize key elements. Throughout this text, the intent has been to provide a historical and general overview of program planning and evaluation, without advocating for any particular model. Responsible evaluation practice includes staying current with these developments in the field, both to enhance one's own evaluation practice and to be able to respond to questions about the "latest and greatest" models and trends. The value of models of planning and evaluation is that they allow the user to sort ideas, tasks, and considerations quickly into recognizable patterns that can be easily remembered.

One such model is RE-AIM, which stands for reach, effectiveness, adoption, implementation, and maintenance. The RE-AIM framework can be used to organize a flowchart of program activities (Farris, Will, Khavjou, & Finkelstein, 2007) or to compare two programs on summary indices (Glasgow, Nelson, Strycker, & King, 2006). The RE-AIM framework is intended for individual behavioral change programs, and thus, it is more difficult to apply to programs at the population or infrastructure levels of the public health pyramid.

Other models that have gained recognition focus on the dissemination and uptake of innovations in health programs. One is the Consolidated Framework for Implementation Research (CFIR; Damschroder et al., 2009). The CFIR model specifies a set of factors that may influence the implementation of interventions intended to improve the health of individuals or groups. Greenhalgh, Robert, MacFarlane, Bate, and Kyriakidou (2004) proposed a model that explains the diffusion of innovation in health-care organizations. One question to consider is, What influences the dissemination and uptake of program intervention ideas? Brownson et al. (2007), in seeking to answer this question, found a difference between state health department practitioners and local health department staff members in terms of implementation of recommended programs. Thus, there is no one way to ensure broad dissemination and program uptake.

A health program that begins in one organization would be an innovation to another organization, so the diffusion of the new program becomes a key concept in thinking about dissemination of the program evaluation findings. More important, the growing fields of implementation and dissemination science have relevance to the work of planning, implementing, and evaluating health programs, and program planners need to remain current with developments in these fields.

Overall, the best approaches to staying current are the obvious approaches. Belong to the relevant professional associations, such as the American Evaluation Association. Read the

TABLE 16-6 Examples of Dissemination Modes, Audiences, and Purposes

Mode of Dissemination	Breadth of Audience	Intended Audience	Examples of Purposes
Website	Widest	General public, researchers, program planners, program implementers	To convey the existence of the program and its organization, to recruit
Press release, mass media	Wide	General public	To convey the existence and value of the program and its organization
Newsletters	Wide	Researchers, program planners, program implementers	To provide updates
Professional conferences	Tailored	Researchers, program planners, program implementers	To share findings and insights
Learning collaboratives	Tailored	Researchers, program planners, program implementers	To share problem solving and solutions
Social media	Tailored or wide	Interested public	To provide updates, to recruit, to convey the existence and value of the program
Scholarly journals	Narrowly tailored	Researchers, program planners	To share findings and insights
Trade journals, technical reports	Narrowly tailored	Interested professionals, program implementers	To share findings and procedures
Fact sheets	Narrowest	Persons interested in the program, persons who received the program	To explain the program, to recruit
Small media (e.g., poster, flyer)	Narrowest	Persons interested in the program, persons who received the program	To convey the existence of the program, to recruit

scientific literature on program planning and evaluation regularly. Tap into the social network of health program planners and evaluators, whether through being a listserv member or social media follower. Be willing to delete and replace older ideas and terminology with the more current ideas and language.

► Across the Pyramid

At the direct services level of the public health pyramid, the concern is with how the processes discussed in this chapter ultimately affect the quality, type, and existence of the direct service health programs. Most of the processes discussed occur at the infrastructure level but can subsequently affect decisions regarding what is done at the direct services level of the pyramid. Nonetheless, informed consent from program and evaluation participants is most likely to be obtained at this level of the pyramid because health programs at the direct services level involve individuals who are accessible for obtaining consent. Meta-analyses, along with evidence-based reviews, are likely to focus on direct services and programs because they are widely studied and evaluated.

At the enabling services level, obtaining informed consent for participation in the health program and evaluation may be feasible and therefore necessary.

The interpretations and sensemaking of groups is different from that of individuals or of populations. This may be an important consideration in reporting the evaluation findings. Also, aggregates and groups are often the recipients of enabling services and programs, making their participation both feasible and helpful in generating recommendations.

At the population-based level, the issue of informed consent becomes blurred with the implementation of health policy because consent is almost always implied by the passage of the health policies. Nonetheless, as Davis and Lantos (2000) point out, ethical considerations pervade population-level services. These authors

caution that if sanctions are attached to a failure to comply with required services, such as immunization, harm may be done. Unfortunately, the nature of the harm is likely to be overlooked by policy makers when health policy is formulated and evaluated because policy makers tend to focus on the benefits. Because population-based services are so closely linked to health policy, the processes that occur at the infrastructure level are critical in determining the nature and scope of health programs at this level of the pyramid. However, the growing body of evaluations of population-based health services, often couched as health policy studies, can lead to meta-evaluations and evidence-based practice for population-based services.

At the infrastructure level of the public health pyramid, sensemaking occurs within individuals, but the collective processes by which organizations and policy makers achieve a shared understanding and interpretation of evaluation findings leads to policy decisions regarding health programs. Policy adoption and program implementation are two possible uses for evaluations (McClintock & Colosi, 1998). Recommendations are given to the decision makers, and decisions regarding health program implementation are made. At some point in this process, the organizational, contextual issues become evident. This evolution is reflected in the findings from Brownson et al. (2007) that, although state health officials may be aware of effective programs, they may not have the authority, budget resources, staffing, or legislative support to implement those programs. Also, approaches to improving organizational processes are essentially infrastructure processes that have consequences for programs at the other levels of the public health pyramid. Decision makers, who are part of the infrastructure, may need to reconcile differences between organizational process improvement and evaluation with regard to perspectives and recommendations. Procedures for dealing with ethical issues and IRB and HIPAA procedures must emanate from the infrastructure level and be applied consistently throughout the pyramid levels.

► Discussion Questions and Activities

1. At each level of the public health pyramid, identify at least two factors that can affect the acquisition of informed consent from those involved in providing evaluation data about a health program.
2. What might be some effective strategies to prevent the various misuses of evaluations that were described in this chapter?
3. In what ways might evidence-based practice (of any of the health disciplines) benefit from a meta-evaluation of programs to address a given health problem?
4. Using evaluation findings as part of feedback loops to improve or sustain health programs is predicated on a set of assumptions about humans and decision makers. What might be some of the assumptions that underlie this perspective? What, if anything, can be done to overcome, address, or deal with those assumptions to create more effective feedback loops?

► Internet Resources

Dissemination and Implementation Science

These websites provide resources related to dissemination and implementation science: <http://dissemination-implementation.org/index.aspx> and <http://www.pcori.org/research-results/research-dissemination-and-implementation>.

Health Insurance Portability and Accountability Act

There are several websites specific to the HIPAA regulations, such as federal websites (<https://www.hhs.gov/hipaa/for-professionals/index.html>). And, perhaps not surprisingly, a lot of information can be found at Wikipedia (<http://en.wikipedia.org/wiki/HIPAA>).

Human Studies

A number of websites focus on the protection of human subjects, including websites operated by the Department of Health and Human Services (www.hhs.gov/ohrp/) and various universities, such as the website operated by the University of Minnesota (www.research.umn.edu/irb/index.html#.T1Ejn3ILMpk).

Research Quality

Various resources are available to help check on overall quality of research. For example, this paper by West and colleagues for the Agency for Healthcare Research and Quality (AHRQ) deals with assessing the quality of research: www.thecre.com/pdf/ahrq-system-strength.pdf. The Key Evaluation Checklist by a renowned evaluation center is also helpful throughout the research process: <https://www.wmich.edu/evaluation/checklists>.

Writing Recommendations

This World Health Organization website has information on writing reports, with this sub-link about writing recommendations. <http://colelearning.net/who/module3/page43.html>.

References

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Alfonso, M. L., Nickelson, J., Hogeboom, D. L., French, J., Bryant, C. A., McDermott, R. J., & Baldwin J. A. (2008). Assessing local capacity for health intervention. *Evaluation and Program Planning*, 31(2), 145–159.
- Brownson, R. C., Ballew, P., Dieffenderfer, B., Haire-John, D., Health, G. W., Kreuter, M. W., Meyers, B. A. (2007). Evidence-based interventions to promote physical activity: What contributes to dissemination by state health departments. *American Journal of Preventive Medicine*, 33(1 suppl), S66–S78.
- Centers for Medicaid and Medicare Services. (2005). *The Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. Retrieved from www.cms.hhs.gov/HIPAAGenInfo/Downloads/HIPAAALaw.pdf
- Chapman, G. B., & Elstein, A. S. (2000). Cognitive processes and biases in medical decision making. In G. B. Chapman & F. A. Sonnenberg (Eds.), *Decision making in health care: Theory, psychology, and applications* (pp. 183–210). Cambridge, UK: Cambridge University Press.

- Chelimsky, E. (2007). Factors influencing the choice of methods in federal evaluation practice. *New Directions for Evaluation*, 113, 13–33.
- Damschroder, L. J., Aron, D., Keith, R. E., Kirsch, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation for health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4, 50. Retrieved from <http://www.implementationscience.com/content/4/1/50>
- Davis, M. M., & Lantos, J. D. (2000). Ethical considerations in the public policy laboratory. *Journal of the American Medical Association*, 284(1), 85–87.
- Department of Health and Human Services. (2008). *Office for Civil Rights—HIPAA*. Retrieved from www.hhs.gov/ocr/hipaa
- Dyer, C. B., Goodwin, J. S., Pickens-Pace, S., Burnet, J., & Kelly, P. A. (2007). Self-neglect among the elderly: A model based on more than 500 patients seen by a geriatric medicine team. *American Journal of Public Health*, 97(9), 1671–1676.
- Dynarski, M. (1997). Trade-offs in designing a social program experiment. *Children and Youth Services Review*, 19(7), 525–540.
- Farris, R. P., Will, J. C., Khavjou, O., & Finkelstein, E. A. (2007). Beyond effectiveness: Evaluating the public health impact of the WISEWOMAN program. *American Journal of Public Health*, 97(4), 641–647.
- Feather, J. (1982). Using macro variables in program evaluation. *Evaluation and Program Planning*, 5(3), 209–215.
- Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgments under certainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.
- Fisher, E. B., Brownson, C. A., O'Toole, M. L., Shetty, G., Anwuri, W., Fazzone, P., . . . Bazzarre, T. L. (2007). The Robert Wood Johnson Foundation Diabetes Initiative: Demonstration projects emphasizing self-management. *Diabetes Educator*, 33(1), 83–84, 86–88.
- Gabriel, R. M. (2000). Methodological challenges in evaluating community partnerships and coalitions: Still crazy after all these years. *Journal of Community Psychology*, 28(3), 339–352.
- Glasgow, R. E., Nelson, C. G., Strycker, L. A., & King, D. K. (2006). Using RE-AIM metrics to evaluate diabetes self-management support interventions. *American Journal of Preventive Medicine*, 30(1), 67–73.
- Gondolf, E. W. (2000). Human subject issues in batterer program evaluation. *Journal of Aggression, Maltreatment and Trauma*, 4, 273–297.
- Good, P. I., & Hardin, J. W. (2012). *Common errors in statistics (and how to avoid them)* (4th ed.). New York, NY: John Wiley & Sons.
- Gorman, D. M., & Conde, E. (2007). Conflict of interest in the evaluation and dissemination of “model” school-based drug and violence prevention programs. *Evaluation and Program Planning*, 30(4), 422–429.
- Greenhalgh, T., Robert, G., MacFarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*, 82, 581–629.
- Hastorf, A., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal Psychology*, 49(1), 129–134.
- Hendricks, M., & Papagiannis, M. (1990). Do's and don'ts for offering effective recommendations. *Evaluation Practice*, 11(2), 121–125.
- Hill, L. G., Maucicone, K., & Hood, B. K. (2007). A focused approach to assessing program fidelity. *Prevention Science*, 8(1), 25–34.
- Howell, E. M., Devaney, B., Foot, B., Harrington, M., Schettini, M., McCormick, M., . . . Zimmerman, B., (1997). *The implementation of Healthy Start: Lessons for the future*. Washington, DC: Mathematica Policy Research.
- Hussemann, J. M., Mortimer, J. T., & Zhang, L. (2016). Exploring the correlates of parental consent for children's participation in surveys: An intergenerational longitudinal study. *Public Opinion Quarterly*, 80, 642–665.
- Iriti, J. E., Bickel, W. E., & Nelson, C. A. (2005). Using recommendations in evaluation: A decision-making framework for evaluators. *American Journal of Evaluation*, 26, 464–479.
- Julnes, G., & Rog, D. J. (2007). Current federal policies and controversies over methodology in evaluation. *New Directions for Evaluation*, 113, 1–12.
- Lipshitz, R., Gilad, Z., & Suleiman, R. (2000). The one-of-us effect in decision evaluation. *Acta Psychologica*, 108(1), 53–71.
- Longfield, J. N., Morris, M. J., Moran, K. A., Kragh, J. F., Wolf, R., & Baskin, T. W. (2008). Community meetings for emergency research community consultation. *Critical Care Medicine*, 36(3), 731–736.
- Marvin, G. (1985). Evaluation research: Why a formal ethics review is needed. *Journal of Applied Social Sciences*, 9, 119–135.
- Mathison, S. (1999). Rights, responsibilities, and duties: A comparison of ethics for internal and external evaluators. *New Directions for Evaluation*, 82, 25–34.
- McClintock, C., & Colosi, L. A. (1998). Evaluation of welfare reform: A framework for addressing the urgent and the important. *Evaluation Review*, 22(5), 668–694.
- Mohan, R., & Sullivan, K. (2006). Managing the politics of evaluation to achieve impact. *New Directions for Evaluation*, 112, 7–23.
- Morris, M., & Jacobs, L. R. (2000). You got a problem with that? Exploring evaluators' disagreements about ethics. *Evaluation Review*, 24(4), 384–406.
- National Institutes of Health. (2003). *Impact of the HIPAA privacy rule on NIH processes involving the review, funding, and progress monitoring of grants, cooperative agreements and research contracts*. Notice NOT-OD-03-025. Retrieved

- from <http://grants2.nih.gov/grants/guide/notice-files/NOT-OD-03-025.html>
- Owen, J. M., & Lambert, F. C. (1998). Evaluation and the information needs of organizational leaders. *American Journal of Evaluation, 19*(3), 355–365.
- Patton, M. Q. (1997). *Utilization focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (1999). Organizational development and evaluation. *Canadian Journal of Program Evaluation* (Special Issue), 93–113.
- Patton, M. (2008). *Utilization focused evaluation* (4th ed.). Newbury Park, CA: Sage
- Shaw, T., Cross, D., Thomas, L. T., & Zubrick, S. R. (2015). Bias in student survey findings from active parental consent procedures. *British Educational Research Journal, 41*, 229–243.
- Stevens, C. J., & Dial, M. (1994, Winter). What constitutes misuse? *New Directions for Program Evaluation, 64*, 3–14.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.
- VanLandingham, G. R. (2006). A voice crying in the wilderness: Legislative oversight agencies' efforts to achieve utilization. *New Directions for Evaluation, 112*, 25–39.
- Weick, K. E. (1979). *The social psychology of organizing*. Reading, MA: Addison-Wesley.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

Index

Note: Page numbers followed by *f*, *t* and *e* indicates figures, tables and exhibits respectively.

A

ABCD. *See* asset-based community development

ACA. *See* Affordable Care Act (ACA)

acceptability of intervention, PEARL system, 113–114

accountability, 246–247

accreditation evaluation, 8, 20

acculturation, 40

accuracy

of back-translated

questionnaires, 37

data entry, 216, 223, 281

ensuring testing, 248

increasing credibility with feedback about, 368

linked data issues, 338

standard for evaluation, 8–9, 9*t*, 400

actions, implementation-focused, 257–258

active protection, 15, 131

activity logs, 223, 224*t*

adequacy index, calculation of, 341*t*

administration credibility of

intervention dosage, 128

adolescent pregnancy prevention program, 340

advisory boards, 228

advocacy approach, planning, 66–67

AEA. *See* American Evaluation Association (AEA)

Affordable Care Act (ACA), 77

African Americans

and diversity of health

professionals, 38

tailoring interventions for, 38

age

population characteristics data on, 96

quality of life acronyms and, 97*t*

Agency for Healthcare Research and Quality (AHRQ), 10

funding of evaluations, 10

performance measurement, 261

aggregates

characteristics of, 52

enabling services for, 22

use of public health pyramid, 22

aggregation of data, 344

AHRQ. *See* Agency for Healthcare Research and Quality (AHRQ)

alcohol and substance

abuse, influence on

interventions, 38

alternative plan formats

business plans, 195, 196*t*

logic models, 193–195, 194*t*

altruism, motivating volunteers, 179

ambiguity, defined, 17

American Evaluation

Association (AEA)

evaluation standards, 281

founding of, 8

Internet resource, 27

program evaluation

standards, 400

American Recovery and

Reconstruction Act

(ARRA), 214

American Society for Quality, 260

analysis paralysis, case study, 366

antecedent factors, as variables, 273

apolitical approach, planning, 66

APEXPH, 70

archival data sources, 93

ARRA. *See* American Recovery and Reconstruction Act (ARRA)

articulate statements, 389–390

assessment

community health. *See* community health assessment

of cultural competency, 48–49

data-related and methodological issues, 83–85

of implementation, 219–222

marketing, 76

needs, 62–63, 76–77

organizational, 75–76

perspectives on, 71–75, 72*f*

public health pyramid in, 23

steps in conducting, 77–83

workforce, 77

asset-based community

development (ABCD), 74

asset model of needs, 73*t*

asset perspective model, 74–75

assets assessment, 12

association analyses, 349–350, 350–352*t*

assumptions, check against, 141

assurance, using public health pyramid in, 23

ATLAS-ti software, 377

attitudes data, 96

attrition

comparison group, posttest only

designs, 299–301

nonresponse bias due to, 321

two-group, prospective (cohort) design, 306

attunement, cultural, 47

audience, report writing for, 392–393

audiovisual recordings, 363

autonomy principle, 384

B

backlash, 354

backsliding, 354

Balanced Scorecard (BSC), 251

baseline data, 289

- basic priority rating system (BPRS), 111–113, 112*t*, 113*f*
 - behavioral domain, measurement variables, 277*t*
 - Behavioral Risk Factor Surveillance System (BRFSS), 348, 368
 - behaviors data, related to health problem, 96
 - benchmarking, 256
 - bias
 - in design
 - comparison group, posttest only design, 299–301
 - multiple-group, time-series design advantage, 303–304
 - one-group, posttest design, 298
 - one-group, time-series design, 303
 - two-group, pretest/posttest, with random assignment, 309
 - effect evaluations and, 290–291
 - internal vs. external evaluators, 10
 - as misuse of evaluations, 397
 - nonresponse, in sampling, 321
 - objectivity in qualitative methods, 369
 - observational, 365
 - response, to questionnaire data, 327–328, 328*t*
 - big data, definition of, 329–330
 - bioethics, 383
 - biological diversity, 39
 - birth records, 328
 - blindness, cultural, 46
 - bounded outcomes, health outcomes classifications, 292
 - Bowe County program, 255
 - BPRS. *See* basic priority rating system (BPRS)
 - break-even analysis, 205–207, 206*e*, 207*t*, 208*e*
 - breast cancer, 31, 38
 - BRFSS. *See* Behavioral Risk Factor Surveillance System (BRFSS)
 - broader dissemination, 402
 - BSC. *See* Balanced Scorecard (BSC)
 - budgets, 201–217
 - break-even analysis, 205–207, 206*e*, 207*t*, 208*e*
 - cost analyses, types of, 209, 212, 212*e*
 - cost-benefit analysis, 213
 - cost comparison, 213
 - cost description, 212
 - cost-effectiveness analysis, 213
 - cost minimization, 212–213
 - cost-utility analysis, 213
 - developing, 182, 201–202
 - evaluations, 186
 - fiscal resources, 182
 - Internet resources, 217
 - justifications, 207, 209
 - monitoring variance in, 209, 210–211*e*, 230
 - program costs, 204
 - for random assignments, 308
 - for start-up and evaluation costs, 205
 - terminology, 202–203
 - burden-of-disease measures, 96
 - Bureau of Primary Health Care, 53–54
 - business plans
 - Internet resources, 197–198
 - program theories as, 195
- C**
- CAHPS. *See* Consumer Assessment of Health Plans (CAHPS)
 - Capacity Assessment for Title V (CAST-5), 255*t*
 - capacity objectives, 149
 - capture-recapture sampling, 318
 - CARE. *See* Commission on Accreditation of Rehabilitation Facilities (CARF)
 - case-control design, retrospective, 304–305
 - case files, as data source, 328
 - case studies
 - key feature of, 365
 - organizational assessments using, 76
 - as process evaluation data, 223, 225*t*
 - qualitative research using, 365
 - CAST-5. *See* Capacity Assessment for Title V (CAST-5)
 - categories, qualitative data analysis, 372–374
 - category saturation, 370–371
 - causal factors, 104–107
 - diagramming health problem, 102–104
 - examples across pyramid, 116–117*t*
 - as independent variables, 273
 - overview of, 104–105
 - causal relationship. *See* cause-and-effect
 - causal theory, 390
 - diagramming health problem using, 102–110
 - effect objectives for, 154–155*t*
 - elements of, 104–108
 - examples of, 107–108*f*
 - identifying effect evaluation variables using, 271–272
 - model, 104*f*
 - outcomes and impacts in, 135–136
 - as part of effect theory, 125–126, 126*f*, 135–136, 140*t*
 - writing statement of, 108–110
 - cause-and-effect
 - determining effect evaluation design, 289
 - diagramming of, 251
 - cause diagrams. *See* fishbone (Ishikawa) chart
 - CBA. *See* cost-benefit analysis (CBA)
 - CDC. *See* Centers for Disease Control and Prevention (CDC)
 - CEA. *See* cost-effectiveness analysis (CEA)
 - ceiling effect, 240
 - Center for Evidence-Based Medicine, 312
 - Centers for Disease Control and Prevention (CDC), 53, 68
 - development of logic models, 193–194
 - Internet resource for, 27, 197
 - intervention effects evaluations, 285
 - certification, health professional, 177–178, 177*t*
 - change
 - evaluators as agents of, 10–11, 398–399
 - fallacies in interpreting evaluation data, 353–354, 355*f*
 - Internet resources for, 144–145
 - interpreting implementation data, 255–257
 - managing quality and fidelity in, 258–259
 - program effect evaluations and, 339–343
 - when and what not to, 259
 - changeability, prioritizing based on degree of, 114

- charges, service, 185, 202
- charts, report writing with, 393
- check sheet, 249*f*
- checklists
- collecting process evaluation data, 223, 224*t*
 - to facilitate cultural competency, 51*e*
- children, ethical issues of evaluating, 387–388
- Children, Youth, and Families at Risk (CYFAR), 243
- CHIP. *See* Community Health Improvement Process (CHIP)
- CI. *See* confidence interval (CI)
- claims, importance of, 390
- client records, 223, 224*t*
- clinic records, 231
- clinical significance, statistical vs., 343
- cluster trial, 309
- coalitions
- development of, 50–51
 - govern participation in, 51
 - for health programs, 50
 - measuring human resources through, 228
 - monitoring system for, 221
 - strategies for, 50–51
- coding
- analyzing qualitative data with, 373*t*, 374
 - data collection using observation, 364
- cognitive process domain, measurement variables, 277*t*
- cohort designs, prospective, 305–306
- collaboration, 50, 61, 61*f*, 194
- collaborators, evaluators as, 11
- Commission on Accreditation of Rehabilitation Facilities (CARF), 20–21
- communication
- action approach, planning, 67
 - as management skill, 181
 - process theory forming basis for, 142–143
 - style, 46
- community, 47
- action groups, using evaluation results, 13
 - assessment, program life cycle, 14
 - as context, 60–61
 - health assessment of, 77
 - participatory research approach based on, 66
 - as target or recipient, 60–61
- community based, definition of, 61–62
- community competence process, 63
- community diagnosis, 108
- community driven, definition of, 61–62
- community focused, definition of, 61–62
- Community Guide to Preventive Services, 145
- community health assessment, 59–87, 99, 108–109. *See also* community needs assessment
- analyzing, 100
 - characteristic of, 59–60
 - characterizing health problems. *See* health problems
 - collecting data, 91–94
 - community as context and recipient, 60–61, 61*f*
 - conducting, 77–83
 - as context and target, 60–61
 - data-related issues of, 83–85
 - defining based, focused, and driven, 61–62, 62*f*
 - defining community, 59–60
 - elements of, 60*t*
 - epidemiological model of, 72
 - Internet resources, 86–87
 - methodological issues of, 83–85
 - needs assessment. *See* needs assessment
 - overview of, 77
 - in prioritization stage of planning, 82
 - public health models of, 74
 - sense of, 60
 - social model of, 74
 - across the pyramid, 85
 - types of, 75–77
- Community Health Improvement Process (CHIP), 70
- community health workers, diversity among, 41–42*t*
- community members, using evaluation results, 13
- community needs assessment, as type of evaluation, 19
- community resilience, 63–64
- Community Tool Box, 86, 118
- community trial, 309
- comparative need, 63
- comparison group, outcome documentation design, 299–301
- comparison questions, 348, 349*t*
- comparisons, in qualitative data analysis, 372
- competence, cultural. *See* cultural competency
- comprehensive evaluations, 20
- comprehensive rational approach, planning, 67–68
- confidence interval (CI), 100–101
- confidentiality linked data issues, 338
- confirmability, in qualitative methods, 369
- consortiums, 78, 228
- construction, sample
- calculating response rates, 319–321
 - hard-to-reach populations, 318
 - overview of, 317
 - sample size, 318–319, 320*f*
- consultants
- as external planners and evaluators, 10, 399
 - proprietary rights over data and, 399
- Consumer Assessment of Health Plans (CAHPS), 255*t*
- consumer involvement, 47
- context, community as, 60–61
- continuous quality improvement (CQI)
- internal evaluators trained in, 10
 - overview of, 248
 - quality improvement tools, 248–251
 - relevance to health programs, 251–252
- continuous variables, 276–277
- contracts, responsible, 398–400
- control
- charts, 249*f*, 251
 - definition of, 17
- convenience samples, nonprobability, 322
- correlational analyses, 349–350, 350–352*t*
- cost analyses, 209, 212–213, 212*e*
- cost-benefit analysis, 213
 - cost comparison, 213
 - cost description, 212
 - cost-effectiveness analysis, 213
 - cost minimization, 212–213
 - cost-utility analysis, 213

- cost-benefit analysis (CBA), 213
- cost comparison, 213
- cost description, 212
- cost-effectiveness analysis (CEA), 213
- cost minimization, 212–213
- cost-utility analysis, 213
- costs. *See also* economic analyses
- causality and design, 289–290, 290*f*
 - criteria for good
 - interventions, 134
 - evaluations of, 19–20
 - monetary resources, 182
 - physical resources, 179
 - service charges vs., 185, 202
 - time, 180
 - training, 178
 - transportation, 180
- coverage, 234–237
- accountability, 246, 247*t*
 - calculating efficiency of, 236*e*
 - interpreting implementation data, 255–257
 - maintaining program quality and fidelity, 258
 - measures of, 234–237, 235–236*e*, 235*t*
 - process evaluation of
 - program, 234
 - units of service, 237
- CPT codes. *See* Current Procedural Terminology (CPT) codes
- CQI. *See* continuous quality improvement (CQI)
- credibility
- assessing evaluation, 399
 - of coalition relationships, 50
 - persuasive statements based on, 390
 - of qualitative methods, 368
- critical planning theory, 67
- criticality, 384–385
- cultural appropriateness, 47
- cultural attunement, 47
- cultural blindness, 46
- cultural competency, 44
- checklist for developing, 51*e*
 - in community engagement, 51
 - cultural blindness and, 46
 - cultural destructiveness and, 44, 46
 - cultural incapacity and, 46
 - cultural openness and, 46–47
 - cultural proficiency and, 47–48
 - of data collectors and interpreters, 84
- domains of, 47
- enhancing, 48–49
 - examples of each stage, 45*t*
 - guiding program changes with, 391–392
 - of health programs, 51
 - overviews of, 47
 - and sensitivity, 49
- cultural conflict approach, 39–40
- cultural content approach, 40
- cultural destructiveness, 44, 46
- cultural identity, 34
- relevance to health programs, 32
- cultural incapacity, 46
- cultural integration approach, 40
- cultural leverage, 38
- cultural openness, 46–47
- cultural proficiency, 47–48
- cultural sensitivity, 47
- culture, 32–33
- affecting interpretation of data, 391–392
 - relationship between evaluator, organization and, 398–399
 - relevance and sensitivity of questionnaires to, 325
- culture-bound syndromes, 32
- current practice, responsibility for, 402–404
- Current Procedural Terminology (CPT) codes, 127
- D**
- DALYs. *See* disability-adjusted life-years (DALYs)
- DARE. *See* Drug Abuse Resistance Education (DARE)
- data cleaning, 336
- data collection
- accuracy of, 216, 281
 - community health assessment, 83–85
 - computerized, 335–336
 - demographic, 231
 - descriptive. *See* descriptive data, collecting
 - information system output, 185, 230
 - interpreting implementation, 255–257
 - missing, 279–280
 - precise timing of, 289
 - process evaluation, 223, 224–225*t*
 - program evaluator
 - responsibilities for, 389–392
 - qualitative analysis of, 372–375
 - quantitative analysis of. *See* quantitative data analysis
 - reliability concerns, 280–281
 - setting target values. *See* target values, using data to set
 - training for, 375
 - turning into persuasive information, 389–390, 391*t*
- data collection methods
- big data, 329–330
 - case study methodology, 365
 - focus group method, 363
 - in-depth individual interviews, 361–362
 - observation method, 364
 - overview of, 324–325
 - physical data, 330
 - secondary data, 328–329
 - surveys and questionnaires, 325–328
 - written open-ended questions, 363–364
- data dictionary, 215
- data dredging, 344
- data entry, of quantitative analysis, 335–338
- data management, of quantitative analysis, 335–338
- data sources. *See also* effect evaluations, data sources for
- archival, 93
 - case study methodology using
 - multiple, 365
 - observational, 92–93
 - other types of, 93–94
 - primary, 91
 - prioritizing across pyramid, 116*t*
 - proprietary, 93
 - public, 91–92
 - published literature, 93
 - target-setting decisions and, 157, 159
- de facto manipulation, 293
- decision drift, assessing, 234
- decision making
- business plans facilitating, 195
 - evaluators influencing
 - organizational, 399
 - importance of claims on, 390
 - making recommendations, 397
 - persuasive statements and, 389–390

- phase of planning, 82–83
 phenomena affecting, 391–392
 decision tree, effect evaluation
 design
 for choosing evaluation
 design, 296*f*
 outcome assessment. *See* outcome
 assessment designs
 outcome documentation,
 298–301
 outcome evaluation research,
 307–309
 overview of, 295–297
 website for statistics, 357
 degree of effectiveness, 222
 delivery, intervention, 189–190
 demographic documentation, 231
 denominator, 94
 dentistry, professional diversity
 in, 41*t*
 Department of Health and Human
 Services (DHHS), 10, 29
 dependability, qualitative
 methods, 369
 descriptive data, collecting, 94–99
 attitudes and behaviors, 96
 dynamics leading to problem,
 94–96
 magnitude of problem, 94
 population characteristics, 96
 years and quality of life, 96–99
 descriptive statistics
 odds ratio and relative risk, 100
 overview of, 100, 346–348
 population parameters, 100–101
 tests of significance, 101
 design
 defined, 317
 options identification, 294–295
 desired outcome, at baseline/
 pretest, 340
 destructiveness, cultural, 44, 46
 determinant forces theory, as part of
 effect theory, 135–136, 140*t*
 determinants, 105
 DHHS. *See* Department of Health
 and Human Services
 (DHHS)
*Diagnostic and Statistical Manual
 of Mental Disorders
 (DSM-IV)*, 32
 diagramming
 causal chain of events, 140–141
 health problem, 102–110
 organizational plan outputs,
 148–149
 dietetitian, professional diversity
 among, 41*t*
 difference scores, measuring, 339
 diffusion of innovation curves,
 measuring, 340
 direct costs, budgets and, 202
 direct healthcare services
 public health pyramid, 22
 across pyramid, 25
 direct services, public health
 pyramid
 accountability and quality
 improvement, 259
 community health
 assessments, 85
 disparities, 51–52
 effect evaluation design, 311
 examples of interventions,
 128, 129*t*
 focus of, 4–5
 intervention effects
 evaluation, 284
 objectives and setting targets, 171
 overview of, 22
 process evaluation measures, 241*t*
 program evaluator
 responsibilities, 404
 program implementation, 195
 program theory and
 interventions, 143
 qualitative methods, 376
 quantitative data analysis, 356
 sampling and data sources
 for, 330
 time frame used in objectives, 148
 directed outcome, direction of, 340
 disability-adjusted life-years
 (DALYs), 97–98
 calculating burden of disease, 97
 causes of, 98*t*
 disaster preparedness, 87
 discounting process, 203
 discrimination, 46
 disparities
 causes of, 32
 coalition development and, 50
 community health assessments
 and, 50
 defining, 30–31
 diversity and, 32–33
 Internet resources, 53–54
 public health pyramid, 51–53
 relevance of, 29–30
 target-setting decisions and, 159
 dissemination
 definition of, 401
 report, 398
 responsible for, 401–402, 403*t*
 diversity
 choosing indicators of, 34
 cultural competency and, 44–49
 disparities and, 32–33
 effects of, 30*f*
 health and, 32
 health programs and, 33
 influence on interventions. *See*
 interventions
 within healthcare organizations
 and programs, 43–44
 overview of, 30
 documentation
 of effect evaluation designs for
 outcome, 298–301
 of implementation, 220
 of participants and recipients,
 230–231
 domains
 data sources from health and
 wellbeing, 326*t*
 measurement variables for
 different, 276–277*t*
 domestic violence, and ethics, 388
 dosage, intervention
 evaluation measures of, 238
 fidelity to, 232
 manipulability of, 133
 overview of, 128, 130
 relationship of change to, 342
 Drug Abuse Resistance Education
 (DARE), 132
 duration, of intervention dosage, 128
- ## E
- EBM. *See* evidence-based medicine
 (EBM)
 EBP guidelines. *See* evidence-based
 practice (EBP) guidelines
 ecological correlation, 344
 ecological design, outcome
 assessment design, 306–307
 ecological fallacy, 306–307, 354–355
 ecological model
 program theory and
 interventions, 144
 public health pyramid as,
 23–25, 24*f*
 economic analyses, prioritizing
 health problems using
 PEARL, 113–114

- education
 - first evaluations were in, 6, 293
 - professional diversity among
 - health professions, 41–42*t*
- effect evaluations, 19. *See also*
 - intervention effects evaluation; outcome evaluation
- data sources for
 - across the pyramid, 330
 - big data, 329–330
 - Internet resources, 331
 - overview of, 324–325
 - physical data, 330
 - secondary data, 328–329
 - surveys and questionnaires, 325–328
- design, 287–312, 315–331
 - across the pyramid, 311–312
 - avoiding failures, 309–311
 - bounded outcomes and, 282
 - causality and, 289–290, 290*f*
 - caveats, 288–294
 - decision tree overview, 295–297
 - free of bias, 290–291
 - groups and, 291–292
 - identifying options, 294–295
 - individual vs. population-level interventions, 292
 - Internet resources, 312
 - intervention and observational designs, 292–294
 - outcome assessment. *See* outcome assessment designs
 - outcome documentation, 298–301
 - outcome evaluation
 - research, 307
 - overview of, 287–288
 - practical issues with experimental designs, 307–308
 - retrospective vs. prospective orientation and, 291
 - time span and, 291
 - using designs derived from multiple paradigms, 294
- overview of, 19
- quantitative data analysis in. *See* quantitative data analysis
- sampling
 - across the pyramid, 330
 - calculating response rates, 319–321
 - calculating sample size, 318–319, 320*f*
 - hard-to-reach populations, 318
 - for outcome assessment, 322–324, 323*t*
 - for outcome evaluation, 324
 - overview of, 315
 - probability vs. nonprobability, 317, 317*t*
 - realities of, 315–317
 - tailoring for diversity, 31*t*
- effect objectives, 149–151, 151*f*, 154–155*t*
- target-setting process for, 156
- effect theory
 - avoiding failure, 309–311
 - components of, 135
 - defined, 124
 - establishing target values using, 156
 - generating, 138–140
 - health gain intervention criteria and, 132–133
 - impact accountability relating to, 246, 247*t*
 - outcomes and impacts in, 134–137
 - causal theory, 135–136
 - checking against
 - assumptions, 141
 - comparing with espoused theory and theory-in-use, 139*t*
 - diagramming causal chain of events, 140–141
 - example of, 137*f*
 - generating effect theory, 135
 - impact theory, 136–137
 - intervention theory, 136
 - involving key stakeholders, 138
 - overview of, 134–135
 - using scientific literature, 138–139
- overview of, 125–126, 126*f*
- process evaluation of, 233
- random assignment and, 308
- roots of failure, 143*f*
- variables from program, 271–274
- writing objectives for, 149
- effective and efficient (lower-cost) interventions, 113
- effectiveness, 222–223, 353
- effectiveness index, calculation of, 341*t*
- effects
 - calculating sample size, 319
 - establishing objectives, 156
 - efficacy, of program
 - intervention, 222
 - efficiency
 - accountability for, 246, 247*t*
 - interpreting implementation data, 256
 - of program intervention, 223
 - ratio, interpretation, 354
 - egalitarian approach, 385
 - EHR. *See* electronic health record (EHR)
 - electronic health record (EHR), 214
 - electronic information systems
 - advantages of, 255
 - collecting survey information through, 327
 - for data collected for effect evaluations, 335–336
 - data collection using
 - observation, 364
 - electronic medical record (EMR), 214
 - eligibility screening
 - intervention delivery, 189–190
 - overinclusion and underinclusion, 186–187
 - overview of, 185–186
 - scope, 187–189
 - screening, 189
 - services utilization, 185–186
 - emergency care, ethical issues, 388
 - employees. *See* personnel
 - EMR. *See* electronic medical record (EMR)
 - enabling services, public health pyramid, 22
 - community health assessments, 85
 - disparities, 52
 - effect evaluation design, 311
 - examples of interventions in, 128, 129*t*
 - focus of, 25
 - health problems, 115
 - intervention effects evaluation, 284
 - objectives and setting targets, 171
 - overview of, 22
 - process evaluation measures, 241*t*
 - program evaluator
 - responsibilities, 404
 - program implementation, 195
 - program theory and interventions, 143–144
 - qualitative methods, 376
 - quality improvement, 259–260

- quantitative data analysis, 356
 - sampling and data sources
 - for, 330
 - time frame used in objectives, 148
 - epidemiological designs, application of, 289
 - epidemiological model of needs
 - collecting descriptive data, 94
 - for community assessments, 73*t*
 - comparing with other models, 73*t*
 - epidemiological perspective model, 72
 - epidemiology, 289
 - rates, 102, 103*t*
 - EpiInfo™, statistical shareware, 336
 - espoused theory
 - effect theory, theory-in-use vs., 139*t*
 - overview of, 138
 - process evaluation of, 233
 - ethics
 - planning for effects
 - evaluation, 282
 - in prioritization stage of
 - planning, 82
 - program evaluator
 - responsibilities, 383–389
 - dilemmas, 399
 - evaluation, 387–388
 - evaluation quality and, 400–401
 - HIPAA and, 388–389
 - institutional review board and
 - informed consent, 385–386, 386–387*t*
 - overview of, 383
 - planning, 383–385, 384*t*
 - professional accountability for, 246–247
 - random assignment and, 308
 - ethnic diversity, 33
 - ethnicity, 34
 - health disparities and, 30–33
 - Ethnograph software, 377
 - evaluation
 - accountability standard, 8
 - avoiding failure of, 309–311
 - budgeting for, 186, 205
 - context of
 - basis of merit and effectiveness, 6
 - circumstances not amenable to, 21
 - culture of, 7–8
 - definitions of, 8
 - generation of, 6
 - history of, 6
 - interdependent and cyclic nature of, 11–13
 - Internet resources on, 27
 - judgmental quality of, 8
 - mandated and voluntary, 20–21
 - planning and, 10
 - as profession, 8–11
 - program life cycle, 13–14
 - public health pyramid using in, 23
 - results use as cyclical link, 13
 - standards of, 9*t*
 - types of, 19–21
 - effect of CQI/TQM in, 251
 - ethics and, 387–388
 - evaluator responsibilities. *See* evaluators, responsibilities of
 - first done in field of education, 292–293
 - formative, 20, 259
 - HIPAA and, 388–389
 - implementation of, 221–222
 - Internet resources for, 197
 - intervention effects. *See* intervention effects evaluation
 - misuse of, 397–398
 - physical resources for, 179
 - planning and. *See* planning and evaluation cycle
 - process, 221
 - process-monitoring, 399
 - program theory guiding, 141–142
 - report writing in. *See* report writing
 - research vs., 268–270, 269–270*t*
 - underinclusion affecting, 186–187
 - using goals for, 148
 - Evaluation Exchange, 27
 - evaluation failure, 310–311, 310*t*
 - evaluation phase, qualitative methods, 360
 - evaluation research, 270
 - evaluators, 8–11
 - founding of AEA, 8
 - professionalization of, 8
 - requirements for planners and evaluators, 10
 - responsibilities of, 383–405
 - across the pyramid, 404
 - contracts, 398–400
 - for current practice, 402–404
 - dissemination, responsible for, 401–402
 - ethics, 383–389
 - evaluation quality, 400–401
 - Internet resources, 405
 - reports and presentations, 392–398
 - spin of data and information, 389–392
 - roles of, 10–11
 - selecting statistical tests, 347
 - standards, 9
 - types of professionals as, 10
 - evidence-based interventions, 131–132, 190
 - evidence-based medicine (EBM), 82
 - evidence-based practice (EBP) guidelines, 66, 82
 - executive summary reports, 392
 - exhaustive category, 374
 - exhibits, in reports, 393
 - expenditures, 202
 - expenses, of qualitative data analysis, 375
 - Experimental and Quasi-Experimental Designs for Research* (Campbell and Stanley), 293
 - experimental evaluation designs, 293
 - external validity, 292
 - explanations, program theory enabling, 142
 - exploratory data analysis, 346
 - exposed group, 291
 - expressed need problem, 62
 - external evaluators, 10
 - external validity, 292, 368
 - externalities, 203
- ## F
- face validity, 281
 - facilities, 179
 - failures
 - effect and process theory, 142, 222*f*
 - effect evaluation design, 309–311
 - sources of, program, 309–311, 309*f*
 - family-based programs, 62
 - family-driven programs, 62
 - family-focused programs, 62

feasibility standard, 9
 achieving, 400
 assessing evaluation quality, 400
 for good interventions, 133–134

federal health policy, 6

federal programs
 funding of evaluations, 10
 history of evaluating, 6
 report writing in, 392

feedback loops, and program life cycle, 13–14

fidelity, program, 258–259

findings. *See* presentation of findings

fiscal resources, organizational plan inputs, 182

fishbone (Ishikawa) chart, 249*f*, 251

fixed costs, budgets and, 202

flowcharts diagram, 250*t*, 251

focus group method, qualitative data, 363

formative evaluation, 20, 259

frequency, of intervention dosage, 128

full-coverage programs, 187–189, 353

funding agencies
 developing budget, 202
 and health programs, 10
 indicators mandated by, 152
 presenting findings of qualitative data to, 375
 requiring implementation documentation, 220
 requiring knowledge of satisfaction level, 240
 responsible contracts with, 398
 using evaluation results, 13

fuzzy aspects, of planning, 14–19, 18*t*
 assumptions, 16–17
 paradoxes, 14–16
 uncertainty, ambiguity, risk, and control, 17–19

G

GAO. *See* General Accounting Office (GAO)

Garson, study designs, 312

General Accounting Office (GAO), 6, 261

generalizability of conclusions, 390

geographic information systems (GIS), 214

geographic mapping, 101, 214

GIS. *See* geographic information systems (GIS)

goal-oriented approach, caveats to, 170–171

goals
 choosing indicators for, 151
 defining, 147
 good objectives and, 154–156
 overview of, 147–148

graphics, in report writing, 393

GraphPad.com, intuitive biostatistics, 357

groups
 choosing evaluation design, 291–292
 managing for fidelity and quality, 258–259
 measuring change as difference score, 339

guidance, of program theory, 141–142

H

Haddon model, 95–96, 95*t*

Hanlon method. *See* basic priority rating system (BPRS)

Hanlon's approach planning, public health programs, 111

hard-to-reach populations, 318

health, 3
 concept of, 4
 criterion for good interventions, 132–133
 disparities. *See* disparities
 marketing, 185. *See also* social marketing
 policy changes, 399–400
 politics and right to care, 4

health administration, professional diversity in, 41*t*

health care, outliers in, 337

health education, 288
 professional diversity in, 41*t*

health informatics terminology, 214

health information system, advantages of, 255

health information technology (HIT), 214

Health Insurance Portability and Accountability Act (HIPAA), 405
 effects of, 388–389

health maintenance organizations (HMOs), 69

health problems, 91–118, 153
 collecting data from multiple sources, 91–94
 consequences of, 96
 defined, 4
 defining in community health assessment, 81
 descriptive data, collecting. *See* descriptive data, collecting

diagramming, 102–110
 with indicators, 153*f*
 Internet resources, 118
 mapping of, 101
 model of, 103–104
 planning and evaluation cycle, 11–12, 12*f*
 prioritizing, 110–115
 statement of, 91
 stating, 102
 statistics for. *See* descriptive statistics
 target values for, 160, 164
 across the pyramid, 115–117
 use of term, 4
 writing causal theory of, 108–110

health professionals, using evaluation results, 13

health programs, 81
 community, 62
 cultural competency of, 51
 cultural openness in, 46–47
 development and evaluation, 4–5
 across pyramid, 25–26
 concept of health, 4
 evaluation as profession, 8–11
 fuzzy aspects of planning, 14–19
 health programming planning, 5–8
 health programs, projects, and services, 4–5
 history of planning, 5–8
 implementation of, 12–13
 Internet resources, 27
 Layetteville and Bowe County, 25
 life cycle of, 13–14
 overview of, 3
 planning and evaluation cycle, 11–14, 12*f*
 public health pyramid. *See* public health pyramid types of evaluation, 19–21

- diversity and, 33
 environment for, 32
 evaluation, contribution of disciplines to, 288*t*
 for health problems. *See* health problems
 history of, 5–8
 implementation of. *See* implementation, program
 logistics planning for, 75
 planning for. *See* planning
 relevance of disparities to. *See* disparities
 relevance of diversity to. *See* diversity
- health provider sectors, 43
- Health Resources and Services Administration (HRSA), 47, 53, 153
- health services, 5, 104
 cost vs. charges for, 202
 defined, 4–5
 effect evaluation design in area of research, 289
 enabling, public health pyramid, 22–23
 planning for. *See* planning
- health status indicators, socioeconomic status and, 33
- health status information, measurement of, 231
- Healthcare Effectiveness Data and Information Set (HEDIS), 255*t*
- healthcare organizations, development of, 48*e*
Healthy People 2020, 26*t*, 79, 385
Healthy People publications, evolution of, 5–6
- healthy years equivalent (HYE), 97*t*
 HEDIS. *See* Healthcare Effectiveness Data and Information Set (HEDIS)
- heterogeneity, of qualitative sample, 370
- at high risk, 81
- HIPAA. *See* Health Insurance Portability and Accountability Act (HIPAA)
- histograms, 249*f*, 251
- history threats, 298, 303
- HIT. *See* health information technology (HIT)
- holistic concept, of health, 4
- homogeneity, of qualitative sample, 370
- hospice care, life cycle example in, 14
- human resources, 177–179
 licensure and statutory regulation of professionals, 177–178
 measuring inputs and outputs, 228–229
 organizational assessment of, 75–76
 overview of, 177
 process evaluation of, 228–229
 retaining staff as informational resources, 180
 training for program delivery, 178
 volunteers as, 178–179
 workforce assessment of, 77
- HYE. *See* healthy years equivalent (HYE)
- I**
- ideal coverage, measuring, 235
- ideal design, characteristics of, 289
- identity. *See* cultural identity
- “if-then” statements, organizational theory, 125
- illness, culture and, 32–33
- immunizations
 adult, 107–108
 rates, 102
- impact
 accountability for, 246
 evaluation, 13, 19
 goals as statements about, 147–148
 overview of, 134–135
- impact theory
 effect objectives for, 154–155*t*
 identifying effect evaluation variables using, 271–272
 outcomes and impacts in, 136–137
 as part of effect theory, 125–126, 126*f*, 136–137
- impact variables
 nomenclature convention for, 271
 summary of, 283*t*
 using effect theory to identify, 271–272
- implementation assessment, 221
- implementation documentation, 220
- implementation evaluation, 221–222
- implementation-focused actions, 257–258
- implementation phase of planning, 83
- implementation, program, 175–197
 accountability for, 246
 assessing, 219–222
 formative evaluations for initial stages, 259
 Internet resources, 197
 as longest sustained effort phase, 171
 organizational plan inputs. *See* organizational plan inputs
 organizational plan outputs. *See* organizational plan outputs
 planning cycle, 14
 program theory as business plan, 195
 program theory as logic model, 194
 services utilization plan inputs. *See* services utilization plan inputs
 summary of, 192
 across the pyramid, 195–197
- implied meanings, 372
- importance, prioritizing based on, 114
- improvement
 comparison of methodologies, 252*t*
 fallacies of interpreting evaluation data, 354
 quality, 10, 248, 260
- in-depth individual interviews, 361
- incentives, for participation in research, 321–322
- incidence, health problem, 947
- inclusion, 186–187
- incremental approach, planning, 64–65
- independent variables, causal factors as, 273
- indicators
 applying to outcome objectives, 153
 Bowe County health problems with, 153*t*
 developing objectives by choosing, 151–154
 target-setting decisions and, 157
- indirect costs, budgets and, 202

- individual interventions
 - effect evaluation design for, 292–293
 - two-group, prospective (cohort) design, 305–306
- individual-level effect evaluations, 295
- Industrial Revolution, planning for populations health, 5
- infant mortality rate, 31
- informatics, 253–255
- information
 - program evaluator responsibilities for, 391–392
 - systems, 201, 213–216, 215f, 230
 - technology, 214, 253–255
- informational resources, 180, 227t
- informed consent, 385–386, 386–387t
- infrastructure services, public health pyramid
 - community health assessments, 85
 - disparities, 52–53
 - effect evaluation design, 311–312
 - health problems, 115
 - intervention effects evaluation, 284
 - objectives, 171
 - overview of, 22–23
 - process evaluation measures, 242
 - program evaluator responsibilities, 404
 - program implementation, 195–197
 - program theory and interventions, 144
 - across pyramid, 26
 - qualitative methods, 376
 - quality improvement, 260
 - quantitative data analysis, 356
 - sampling and data sources for, 330
- in-kind donations, budgets, 203
- innovative method, qualitative data, 366
- inputs. *See* organizational plan inputs; services utilization plan inputs
- instant feedback aspect, focus group method, 363
- Institute for Policy Research, 86
- institutional review boards (IRBs), 385–386, 386–387t
- institutionalized program, program life cycle, 14
- instrumentation threat, 299
- insurance claims, as secondary data sources, 328
- intended audience, 71, 80
 - defined, 71
 - full vs. partial-coverage programs, 187–189
 - intended population vs., 185–186
 - overinclusion and underinclusion, 186–187
- intended population, 71
 - criterion for good interventions, 132
 - defined, 71
 - full vs. partial-coverage programs, 187–189
 - sampling for effect evaluations, 315–317
 - use of random assignment in, 308
- interaction, potential of community for, 59
- interesting statements, and persuasion, 390
- internal evaluators, 10
- internal planners and evaluators, 10
- internal validity, bias in design and, 290
- international health programs
 - Internet resource for, 144–145
 - need for global intervention typology, 128
- Internet resources, 86–87
 - budgets, 217
 - community health assessment, 86–87
 - development and evaluation context, 27
 - disparities, 53–54
 - effect evaluations, design, 312
 - effect evaluations, intervention, 285
 - effect evaluations, sampling and data sources, 331
 - health problems, 118
 - objectives and setting targets, 172
 - process evaluation, 243
 - program evaluator responsibilities, 405
 - program implementation, 197–198
 - program theory and interventions, 144–145
 - qualitative methods, 377
 - quality improvement, 260–261
 - quantitative data analysis, 357
- interobserver agreement, 280
- interpersonal information systems, measurement of, 230
- interpretation of evaluation data
 - ecological fallacy, 354–355
 - fallacies of, 353–354, 355f
 - overview of, 353
- interrater agreement, 280
- interval variables, 276
- intervening variables, 273–274
- intervention delivery, 189–190, 231–234
- intervention designs, 293
- intervention effectiveness, 113
- intervention effects evaluation, 265–285
 - ethics and, 282
 - Internet resources, 285
 - measure sensitivity, 278–279
 - measures, types of variables, 275–278
 - measures, using units of observation, 275
 - overview of, 265–266
 - planning for, 281–283
 - questions. *See* questions, intervention effects evaluation
 - stakeholders, 282–283
 - standards, 281
 - across the pyramid, 284
 - threats to data quality, 279–281
 - timing, 278, 279t
 - variables from program effect theory, 271–274
- intervention efficacy, 342, 342t
- intervention fidelity, 231–234
- intervention mechanism theory failure, 310t
- intervention theory
 - effect objectives for, 154–155t
 - identifying effect evaluation variables with, 271–272
 - outcomes and impacts in, 136
 - as part of effect theory, 125–126, 126f, 136
 - process evaluation of, 233
 - using random assignment, 308
- interventional evaluation designs, 292–293
- interventions, 126–134
 - across the pyramid, 143–144
 - administration and dosage, 128, 130
 - change and, 342
 - criteria for good, 131–134
 - diversity and, 30
 - examples of, 129t
 - finding and identifying, 126–127
 - health program vs. medical, 128
 - individual vs. population-level, 292

- influences of biological diversity
on, 39
- influences of sociocultural
diversity on, 38–39
- Internet resources for, 144–145
- in investigation phase, 81–82
- material resources for, 179
- need for, 52
- observational designs and,
292–294
- outcomes and impacts of,
134–137
- overview of, 126
- process evaluation of, 232*t*
- program components and,
130–131
- program failure and, 222*f*
- program theory and, 123
- random assignment and, 308
- rapid assessment for, 75
- specifying, 128, 130
- types of, 127–128
- using volunteers to deliver, 178
- interviews
ability and skill of, 361
- assessing patient satisfaction, 240
- collecting process evaluation
data, 223, 225*t*
- demographic documentation, 231
- in-depth individual, 361
- investigation phase of planning,
81–82
- IRBs. *See* institutional review
boards (IRBs)
- Ishikawa (fishbone) chart, 251
- J**
- job satisfaction, 229
- The Joint Commission (TJC), 54,
255, 255*t*
- Joint Commission on Standards
for Educational
Evaluation, 400
- evaluation standards established
by, 9*t*
- judgmental quality, of evaluation, 8
- K**
- kappa statistic, 369
- kappa values, 280
- Kellogg Foundation Evaluation
manual, 197
- key informant, 361
- knowledge domain, measurement
variables, 276*t*
- L**
- leadership
evaluation for, 398
- of managerial personnel, 181–182
- selecting planning group, 79
- legal accountability, 246, 247*t*
- legal issues
accountability for organizational
plan, 246, 247*t*
- considerations for professionals,
177–178
- prioritizing health problems
using PEARL, 113–114
- legality element, 113–114
- licensure, of health professionals,
177–178
- life cycle, health program, 13–14
- linked data, 337–338
- logic model
defined, 193
- goals and objectives deriving
from, 147, 155
- health program model vs.,
103–104
- Internet resources, 197
- program theory as, 193–194
- program theory vs., 124–125
- logic statements, 160
- logistics, program, 133–134, 240–241
- long-term objectives, value of, 157
- longitudinal design, effect
evaluations, 291, 353
- M**
- magnitude, in persuasive
statements, 389
- magnitude of problem, 94, 112
- management
ethical issues of, 388
- evaluation of, 398
- of program quality. *See* quality,
program
- using process evaluation
results, 13
- management by objectives
(MBO), 170
- management information system
(MIS), 214
- management of quantitative
analysis, data, 335–338
- management resources
organizational plan inputs,
180–182
- tracking service plan outputs, 191*t*
- managerial resources, measuring
inputs and outputs, 227*t*
- mandate evaluations, 20
- manifest meanings, 372
- manipulability, of good
interventions, 133
- MAPP. *See* Mobilizing for Action
through Planning and
Partnership (MAPP)
- mapping, geographic, 101, 214
- marketing assessment, 76
- matching variables, 338
- material resources, 179, 190
- materials produced, process
evaluation, 241
- Maternal and Child Health Bureau
(MCHB)
indicators mandated by, 153
- maturation threats, 298
- MBO. *See* management by
objectives (MBO)
- MCHB. *See* Maternal and Child
Health Bureau (MCHB)
- measure sensitivity, 278–279
- measurement, 33–38
analyzing indicators using, 153
- of change as difference score, 339
- collecting descriptive data. *See*
descriptive data, collecting
- community health assessment
issues, 84
- intervention effects evaluation,
275–279
- performance, 252–253, 254*t*
- mediating factors, 105, 106*t*
- diagramming health problem, 105
- examples across pyramid, 117*t*
- overview of, 105
- as variables, 273–274
- mediating variables, 350
- mediators, evaluators as, 11
- medical anthropology, 54
- medical records, as data
source, 328
- medicine, professional diversity
in, 42*t*
- mental health
culture-bound syndromes of, 32
- mandated evaluation linked to, 20
- using communication action
approach for, 67

mental health domain, measurement variables, 277*t*

meta-analyses, 7, 404

meta-evaluations, 7, 20
overview of, 401

methodological issues, community health assessments, 83–85

methods
in evaluation design, 288
as techniques used to collect data, 324–325

MIS. *See* management information system (MIS)

mission statements, 79

misuse of evaluations, 397–398

mixed methods, notion of, 367–368

mnemonics, 131

Mobilizing for Action through Planning and Partnership (MAPP), 70, 86
public health model of needs assessment, 74

moderating factors, 105, 106*t*
diagramming health problem, 105–108
overview of, 105
as variables, 273–274

moderating variables, 350

monetary donations, 182

monetary resources
measuring inputs and outputs, 227*t*
organizational plan inputs, 182
process evaluation of, 230

monitoring phase, qualitative methods, 360

multiculturalism, 48

multiple-group, time-series design, 303–304

multiplicity sampling, 318

multisite programs, target values for, 167

mutually exclusive category, 374

N

NACCHO. *See* National Association of County and City Health Officials (NACCHO)

NAMI. *See* National alliance on Mental Illness (NAMI)

narrative designs, in qualitative approach, 367

National Alliance on Mental Illness (NAMI), 67

National Association of County and City Health Officials (NACCHO), 118

National Center for Health Statistics (NCHS), 86

National Center on Cultural Competence (NCCC), 53

National Committee on Quality Assurance (NCQA), 260

National Health and Nutrition Examination Survey (NHANES), 329

National Institutes of Health (NIH), 29

National Public Health Performance Standards (NPHPS), 255*t*

National Quality Forum, 260

National Science Foundation, 377

national surveys, as data sources, 329

nationality, measures of, 34

NCCC. *See* National Center on Cultural Competence (NCCC)

NCHS. *See* National Center for Health Statistics (NCHS)

NCQA. *See* National Committee on Quality Assurance (NCQA)

needs assessment. *See also* community health assessment
asset model of, 74–75
comparing models of community, 73*t*
data-related and methodological issues of, 83–85
epidemiological perspective on, 72
goal of, 71
involving community members in, 79–80
overview of, 76–77
perspectives on, 71
program development based on, 12
public health models of, 74
public health perspective on, 74
rapid perspective on, 75
social model of, 74
social perspective on, 74
types of needs in, 62–63
types of strengths in, 63–64

negotiation generation, of evaluations, 6

negotiation skill, of managerial personnel, 181

NHANES. *See* National Health and Nutrition Examination Survey (NHANES)

NICs. *See* nursing interventions (NICs)

No Child Left Behind Act of 2001, 253

NOCs. *See* nursing outcomes (NOCs)

nomenclature convention, of program evaluators, 271

nominal group technique, prioritizing health problems, 111

nominal variables, 275–276, 275*t*

nonparametric statistical tests, 347–348, 348*t*

nonprobability samples
for outcome assessments, 322–324, 323*t*
overview of, 317
probability samples vs., 317, 317*t*

nonreactive measures, 92

nonresponse bias, 321

normative need, 62–63, 112

NPHPS. *See* National Public Health Performance Standards (NPHPS)

numbers, in qualitative findings, 374
numerator, 94
about magnitude of problem, 94
qualitative methods, 369

Nursing Intervention Classification, 127

nursing interventions (NICs), 145

nursing outcomes (NOCs), 145

nutrition education, 294

nutritionist, professional diversity in, 41*t*

NVivo software, 374

O

objectives
across the pyramid, 171
caveats to goal-oriented approach, 170–171
defining, 147
effect objectives, 149–151, 151*f*, 154–155*t*
foci of, 148–151
goals and, 147–148
good goals and, 154–156
indicators, 151–154

- Internet resources, 172
 - process objectives, 148–149, 150*t*
 - relationship of change to, 340–341
 - reports relating findings to, 392
 - setting target values. *See* target values, using data to set
 - TAAPS, 148–149, 150*t*, 256
 - TREW, 149, 266, 273, 278, 341
 - objectivity. *See* confirmability, 369
 - observation
 - collecting process evaluation data through, 223, 225*t*
 - collecting qualitative data through, 364
 - observational data sources, 92–93
 - observational evaluation designs, 292–294
 - odds ratio (OR), 100
 - office equipment, 180
 - Office for Human Research Protections (OHRP), 282
 - Office of Management and Budget (OMB), 253
 - Ohio Department of Health, 261
 - OHRP. *See* Office for Human Research Protections (OHRP)
 - OMB. *See* Office of Management and Budget (OMB)
 - one-group, posttest design, 298–299, 305
 - one-group, pretest/posttest design, 297*t*, 299
 - one-group, time-series design, 302–303
 - online databases, abstracts from, 138
 - online survey software, 336
 - operations manual, 182
 - opportunity costs, in economic analyses, 203
 - oral presentations, 392–398
 - ordinal variables, 275*t*, 276
 - organizational abilities, of management, 181
 - organizational assessment, 75–76
 - organizational charts, 184–185
 - organizational culture, 44
 - organizational plan
 - accountability relating to, 246, 247*t*
 - making recommendations related to, 394–397
 - in process theory, 125, 140*t*
 - organizational plan inputs, 175–182
 - diagram showing components of, 148, 149*f*
 - example of measuring, 227*t*
 - fiscal resources, 182
 - human resources, 177–179
 - informational resources, 180
 - managerial resources, 180–182
 - overview of, 175, 177
 - physical resources, 179–180
 - process evaluation across pyramid, 241–242
 - process evaluation of, 220*f*
 - quantifying human resources, 228–229
 - quantifying, overview of, 223, 226
 - quantifying physical resources, 229
 - time, 180
 - transportation, 180
 - organizational plan outputs, 182–185
 - diagram showing components of, 148
 - information system, 185, 201
 - measuring, 227*t*
 - operations manual, 182
 - organizational chart, 184–185
 - process evaluation of, 220*f*, 241–242
 - quantifying, 230
 - timeline, 182
 - organizational records, 223, 224*t*
 - outcome assessment, 268, 269*t*, 322–324, 323*t*
 - designs, 301–307
 - ecological design, 306–307
 - multiple-group, time-series, 303–304
 - one-group, time-series, 302–303
 - overview of, 301
 - two-group, pretest/posttest, 301–302
 - two-group, prospective (cohort), 305–306
 - two-group, retrospective (case-control), 304–305
 - outcome documentation designs, 298–301
 - comparison group, posttest only, 299–301
 - intervention effects evaluation, 268, 269*t*
 - one-group, posttest only, 298–299
 - one-group, pretest/posttest, 299
 - overview of, 298
 - outcome evaluation, 19
 - intervention effects evaluation, 268, 269*t*
 - overview of, 19
 - planning and evaluation cycle, 13
 - realities of sampling for effect evaluations, 317
 - research designs, 307
 - sampling for, 324
 - outcome-focused evaluations, 7, 7*t*
 - outcome objective target value, relationship of change to, 340–341
 - outcome theory, effect objectives for, 154–155*t*
 - outcome variables
 - in multiple-group, time-series design, 303
 - nomenclature convention for, 271
 - summary of, 283*t*
 - two-group, retrospective (case-control) design, 304–305
 - using effect theory to identify, 271–273
 - outcomes
 - achieving objectives. *See* objectives
 - assessing evaluation for, 401
 - bounded vs. unbounded, 292
 - Bowe County health problems with, 153*t*
 - indicators for, 151–154
 - overinclusion affecting, 186
 - overview of, 134–135
 - reports relating findings to, 392
 - outliers, 337
 - outputs. *See* organizational plan outputs; services utilization plan outputs
 - overcoverage, 236*e*, 256
 - overhead expenses, budgets and, 202
 - overinclusion, 186–187
 - ownership of data, 399
- P**
- PACE-EH. *See* Protocol for Assessing Community Excellence in Environmental Health (PACE-EH)
 - PAHO. *See* Pan American Health Organization (PAHO)
 - Pan American Health Organization (PAHO), 5
 - pandemics, 384

- parametric statistical tests, 347–348, 348*t*
- pared means method, 168
- Pareto charts, 250*t*, 251
- parsimony assumption, 141
- partial-coverage programs, 187–189
- participants
 - choosing evaluation design groups, 292
 - process evaluation of
 - assessing coverage, 234
 - dosage measures, 238
 - overview of, 230
 - satisfaction measures, 239
 - work flow, 240–241
 - sampling for effect evaluations, 315–317
- participation data, linking, 337
- participatory evaluation, 7
- participatory focused evaluations, 7, 7*t*
- passive protection, 15, 131
- PEARL. *See* propriety, economics, acceptability, resources, and legality (PEARL)
- peer debriefing, and credibility, 368
- Pell Institute, The, 377
- perceived need, 63
- percentage agreement, qualitative methods, 368
- performance measurement systems, by healthcare organizations, 255*t*
- performance measures
 - Internet resources, 260–261
 - overview of, 252–253
 - relevance to health programs, 253
 - terminology of, 253
- personnel
 - ethical issues, 388
 - licensure and statutory regulations, 177–178
 - managerial resources, 180–182
 - overview of, 177
 - process evaluation of, 228
 - training for program delivery, 178
 - using internal planners and evaluators, 10
 - using volunteers, 178–179
- persuasive information
 - in report writing, 392
 - turning data into, 389–390, 391*t*
- PERT charts. *See* Program Evaluation Review Technique (PERT) charts
- photovoice methodology, 366–367
- physical data sources, 330
- physical resources
 - measuring inputs and outputs, 227*t*
 - organizational plan inputs, 179–180
 - process evaluation of, 229
- physiological health domain, measurement variables, 276*t*
- pilot programs
 - assessing readability of questionnaires, 327
 - pretesting interventions, 190
 - program life cycle, 13–14
- place, and social marketing, 185, 207
- planning
 - advocacy approach, 66–67
 - apolitical approach, 66
 - approaches to, 64–69, 65*t*
 - budgeting as part of, 204–209
 - communication action approach, 67
 - community health assessment. *See* community health assessment
 - comprehensive rational approach, 67–68
 - decision-making phase, 82–83
 - definition of, 64
 - ethics and, 383–385, 384*t*
 - Hanlon's approach to, public health programs, 111
 - health program
 - and evaluation, 10
 - history of, 5–8
 - interdependent and cyclic nature of, 11–13
 - internal vs. external planners for, 10
 - public health pyramid using in, 23
 - goals and objectives. *See* objectives
 - implementation phase, 83
 - incremental approach, 64–65
 - intervention effects evaluation. *See* intervention effects evaluation
 - investigation phase, 81–82
 - phase, 12
 - prioritization phase, 82
 - public health programs, models for, 69–71, 71*t*
 - strategic planning approach, 68
 - team formation and development, 78–79
 - vision phase, 79
- planning and evaluation cycle, 11–14, 12*f*
 - cultural tailoring for diversity, 31*t*
 - diversity throughout, 30*f*
 - effects evaluation, 266*t*
 - overview of, 11, 12*f*
 - potential points for recommendations, 396, 396*f*
 - program life cycle, 13–14
 - public health pyramid in, 23
 - qualitative methods throughout. *See* qualitative methods using evaluation results as cyclical link, 13
- policy makers, 390, 404
 - using evaluation results, 13
- politics
 - feasibility criterion for good interventions and, 134
 - health policy evaluation and, 13, 400
 - introducing federal social service programs, 6
 - right to health care and, 4
 - social model of needs assessment and, 74
- PolyHeme, 388
- population
 - characteristics data, 96
 - defining for community health assessment, 80–81
 - descriptive statistics, 100
 - hard-to-reach, 318
 - at risk, 81
 - small numbers and small areas statistics, 101–102
- population-based programs, 240
- population-based services, public health pyramid, 22, 25–26
 - community health assessments, 85
 - disparities, 52
 - effect evaluation design for, 292, 311
 - examples of interventions in, 128, 129*t*
 - for health problems, 115
 - intervention effects evaluation, 284
 - maintaining quality and fidelity, 260
 - objectives and setting targets, 171
 - process evaluation measures, 241*t*
 - program evaluator responsibilities, 404
 - program implementation, 195

- program theory and
 - interventions, 143–144
 - qualitative methods, 376
 - quantitative data analysis, 356
 - setting target values for, 167
 - time period for setting objectives, 148
- population-focused programs
 - collecting demographic data in, 231
 - effect evaluation design for, 292
- population-level effect evaluations, 297*t*
- posttests
 - decision tree overview of, 295–297
 - defined, 289
 - measuring difference scores, 339
 - as outcome assessment, 308–309
 - outcome documentation design, 298–299
- power
 - analysis, 319, 331
 - communication action and, 67
- PRECEDE-PROCEED model, 75–76, 86
- precursor data
 - Haddon model for, 94–95
 - to health problem, 94
- prediction analyses, 350–353, 352*t*
- pregnancy prevention program, for adolescent, 340
- presentation of findings
 - evaluator responsibilities for, 393
 - in qualitative data, 375
 - using report writing. *See* report writing
- pretests
 - assessing readability of questionnaires, 327
 - defined, 289
 - desired outcome at, 340
 - measuring difference score, 339
 - outcome assessment, 308–309
 - outcome documentation design, 299
- prevalence, 94
- price, and social marketing, 185, 207
- primary data sources, 92
 - case study methodology using, 365
 - collecting, 92
 - methods for collecting, 325
- primary prevention, 127, 128
- prioritization phase of planning, 82
- prioritizing health problems, 110–115
 - based on importance and changeability, 114–115, 114*t*
 - with BPRS, 111–113, 112*t*, 113*t*
 - with nominal group technique, 111
 - with PEARL component, 113–114
- probability (random) samples
 - nonprobability samples vs., 317, 317*t*
 - overview of, 317
 - sampling for outcome evaluation using, 324
- process evaluation, 19, 219–243
 - across the pyramid, 241–242
 - assessing implementation, 219–222
 - for comprehensive evaluations, 20
 - data collection methods for, 223, 224–225*t*, 229
 - efficacy, effectiveness and efficiency in, 222–223
 - effort across life of program for, 176*f*
 - as implementation evaluation in this text, 221
 - improvement methodologies and, 252*t*
 - Internet resources, 243
 - interpreting implementation data, 255–257
 - of intervention, 233
 - overview of, 219
 - in planning and evaluation cycle, 13
 - quantifying inputs to organizational plan, 223, 226–229
 - quantifying outputs of services utilization plan, 234–241
 - quantifying outputs to organizational plan, 230
 - role of, 222
 - timeline, example of, 183–184*e*
 - using results of, 13
- process implementation, 31*t*
- process monitoring, 221
 - data, 229
 - key aspect of, 233
- process objectives
 - establishing target values for, 156
 - good goals and, 154–156
 - interpreting implementation data, 256
 - overview of, 148–149, 150*t*
 - across the pyramid, 171
- process theory, 19
 - avoiding failure, 310, 310*t*
 - clarifying eligibility guidelines during, 316
 - components of organizational and services utilization plan, 176*f*
 - defined, 124
 - designing process evaluations from, 19
 - elements of, 219, 220*f*
 - focused actions of, 258
 - minimizing overinclusion and underinclusion in, 186–187
 - for neural tube defect prevention program, 193*f*
 - neural tube defects and congenital anomalies health problem, 192*f*
 - overview of, 125
 - process evaluation of, 220*f*
 - for program implementation, 175–197
 - relevance to economic evaluations, 204*f*
 - revising actions based on, 258
 - roots of failure, 143*f*, 222*f*
 - writing process objectives in, 148–149, 150*t*
- processes, defined, 3
- process-monitoring evaluation, 398
- product, and social marketing, 185
- profession diversity, 40–43, 41–42*t*
- professionals
 - accountability of, 246–247
 - evaluators. *See* evaluators
 - legal and regulatory considerations, 177*t*
- program accountability, 246
- program coverage efficiency index, 237
- Program Evaluation Review Technique (PERT) charts, 251
- program evaluation standards, AEA, 400
- program failure
 - roots of, 222*f*
 - sources of, 310, 310*t*
- program life cycle, 13–14
- program logistics, 240–241
- program monitoring, 221
- program participants, using evaluation results, 13
- program reach, 232*t*
- program recipients, using evaluation results, 13
- program staff, using evaluation results, 13

- program theory
 across the pyramid, 143–144
 as basis of communication, 142–143
 as business plan, 195
 effect of CQI/TQM in, 251
 effect theory, 125–126, 126*f*
 enabling explanations, 142
 functions of, 141–143
 goals and objectives deriving from, 147
 Internet resources for, 144–145, 197
 and interventions, 123
 logic model vs., 124–125
 making recommendations on, 394, 395, 396
 making scientific contribution with, 143
 model of, 124*f*
 overview of, 124–125
 in planning and evaluation cycle, 12
 process theory, 125
 providing guidance, 141–142
 tailoring for diversity, 31*t*
 understanding and generating change, 145
- programs. *See* health programs
- project directors, using evaluation results, 13
- project managers, using evaluation results, 13
- projects, defined, 4
- Promoting Health/Preventing Disease* (DHHS), 5
- promotion, and social marketing, 185
- proprietary data sources, 93
- propriety, economics, acceptability, resources, and legality (PEARL), 113–114
- propriety standard of evaluation, 9
 PEARL system, 113
- prospective design, 305–306
 disadvantages of, 291
- Protocol for Assessing Community Excellence in Environmental Health (PACE-EH), 70, 74, 86–87
- provider diversity, 40–43, 41–42*f*
- providers, as data source, 92
- psychology, effect evaluation design in, 288
- public data sources, 91–92
- Public Health Foundation, 261
- public health model of needs, 73*t*
- public health perspective model, 74
- public health pyramid, 21–25, 77, 85
 for accountability and quality, 259–260
 direct services level of, 115, 216–217
 disparities, 51–53
 as ecological model, 23–25, 24*f*
 for effect evaluations, design, 292, 311–312
 for effect evaluations, sampling and data sources, 330
 for effects evaluations, interventions, 284
 enabling services level of, 115, 216
 infrastructure level of, 115
 for objectives and setting targets, 171
 overview of, 21–23, 22*f*
 partial-and full-coverage programs by, 188*t*
 population-based services level of, 115
 population services level of, 216–217
 for process evaluation, 241–242
- program evaluator responsibilities, 404
- for program implementation, 195–197
- for program theory and interventions, 143–144
- for qualitative methods, 376, 376*t*
- for quantitative data analysis, 343–344, 356
- types of interventions in, 129*t*
- use in program planning and evaluation, 23
- published literature, as data sources, 93
- purposive samples, nonprobability, 322
- pyramid, public health. *See* public health pyramid
- qualitative evaluation checklist, 377
- qualitative methods, 359–377
 across the pyramid, 376, 376*t*
 analysis of data, 372–375
 case study, 365
 focus group, 363
 in-depth individual interview, 361
 innovative methods, 366
 Internet resources, 377
 key benefits and challenges using, 362*t*
 multiple methods and triangulation, 367–368
 narrative designs, 367
 observation, 364
 overview of, 360–361, 360–361*t*
 photovoice, 366–367
 presentation of findings, 375
 sampling for, 369–371, 370*t*, 371*t*
 scientific rigor in, 368–369
 throughout planning and evaluation cycle, 359–360
 written open-ended questions, 362–363
- quality assurance, 248
- quality improvement approaches to, 248
 internal evaluators involved with, 10
 Internet resources for, 260–261
 tools, 249–250*t*
- quality of life measures, 96–99
- quality, program, 245–261
 accountability for, 246–247
 creating change in, 255–259
 formative evaluations and, 259
 informatics and information technology, 253–255
 Internet resources, 260–261
 overview of, 245
 performance measurement, 252–253, 254*t*
 quality assurance and, 248
 quality improvement approaches, 248
 across the pyramid, 259–260
- quality, research
 assessing program evaluation, 400
 Internet resources for, 405
 program evaluator responsibilities for, 400
 threats to data, 279–281
- quality-adjusted life-years (QALYs)
 quantitative data analysis, 335–357
 across the pyramid, 356
 change, as difference score, 339

Q

QALYs. *See* quality-adjusted life-years (QALYs)

qualitative data analysis
 issues, 374–375
 overview of, 372–374
 software, 374

- change, issues with quantifying, 339–343
- data entry and management, 335–338
- Internet resources, 357
- interpretation, 353–355
- levels of, 343–344
- linked data, 337–338
- multiple methods and triangulation, 367–368
- outliers, 337
- sample description, 338
- statistical answers to questions, 345–353, 345*t*, 346*f*, 347*t*
- quasi-experimental evaluation designs, 294
- questionnaires
- analysis of. *See* quantitative data analysis
 - assessing patient satisfaction in, 240
 - collecting process evaluation data, 223, 225*t*
 - considerations in constructing, 325, 327
 - gathering demographic documentation through, 231
 - response biases in, 327–328, 328*t*
 - survey considerations, 327
- questions
- collecting qualitative data
 - focus group method, 363
 - for in-depth individual interview, 361
 - overview of, 360–361
 - written open-ended questions, 362–363
 - intervention effects evaluation, 266–271
 - characteristics of right question, 267–268
 - evaluation vs. research, 268–270, 269–270*t*
 - outcome assessment, 268, 269*t*
 - outcome documentation, 268, 269*t*
 - outcome evaluation, 268, 269*t*
 - overview of, 266–267
 - rigor in, 270–271
 - statistical answers to
 - association, 349–350, 350–352*t*
 - comparison, 348, 349*t*
 - description, 346–348
 - overview of, 345–346, 345*t*, 346*f*, 347*t*
 - prediction, 350–353, 352*t*
 - queuing, 189, 232*t*
 - quota samples, nonprobability, 322
- ## R
- race, 34–35, 39
- health disparities and, 30–33
 - indicators used to measure, 35–36*t*
- random assignment, 307
- advantage of, 307
 - overview of, 307
 - two-group pretest/posttest, 308–309
- random selection, 307
- random-digit dialing, 318
- randomized clinical trial, 308–309
- rapid assessment, 75
- rapid assessment and response (RAR), 75
- rapid assessment and response and evaluation (RARE), 75
- rapid BI, 172
- rapid model of needs, 73*t*
- rapid perspective model, 75
- RAR. *See* rapid assessment and response (RAR)
- RARE. *See* rapid assessment and response and evaluation (RARE)
- reach, 234–237. *See also* coverage
- Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) model, 402
- ReCAPP BDI Logic Model course, 197
- recipient audience, 80
- recipients, 390
- community as, 60–61
 - defined, 80, 185–186
 - evaluating intervention effects on, 266
 - implementing program with volunteers who are also, 178
 - measuring participation of, 237
 - process evaluation of, 230–231
- recommendations, report or presentation, 394–397, 394–396*f*
- recordings, audiovisual, advantage of, 364
- recruitment, of volunteers, 179
- referral sampling, 318
- regression analysis, 351–353
- regression to the mean, 302
- relative need, 63
- relative risk (RR), 100
- relevant data, evaluation questions, 267
- reliability
- community health assessments and, 84
 - data collection and, 280–281
 - in qualitative methods, 369
- report writing, 392–398
- involving and negotiating with program partners, 398
 - making recommendations, 394–397, 394–396*f*
 - misuse of evaluations, 397–398
 - overview of, 392–393, 393*t*
- required antecedent factors, 104, 106*t*
- research
- evaluation vs., 269–270, 269–270*t*
 - Internet resources for study design, 312
 - outcome evaluation designs, 307
- research firms, 10
- resource-sensitive approach, 385
- resources. *See also* human resources
- establishing face validity with, 281
 - Internet. *See* Internet resources
 - making recommendations based on, 395
 - in planning and evaluation cycle, 12
 - organizational plan and. *See* organizational plan inputs
 - prioritizing health problems using PEARL, 113–114
- resources domain, measurement variables, 277*t*
- response rates, calculating sample, 319–321
- responsibility, 246. *See also* evaluators, responsibilities of
- responsiveness generation, 6
- retrospective design, 291, 304–305
- revenues, 202–203
- “right” evaluation questions, 267–268
- rigor, in evaluation, 270–271
- risk, defined, 17
- route of administration, 128
- RR. *See* relative risk (RR)

S

- sampling
- calculating response rates, 319–321
 - calculating sample size, 318–319, 320*f*
 - describing for statistical analysis, 338
 - disadvantages of using observation method in, 365
 - hard-to-reach populations, 318
 - for outcome assessment, 322–324, 323*t*
 - for outcome evaluation, 324
 - overview of, 315
 - probability vs. nonprobability, 317, 317*t*
 - qualitative approaches to, 369–371, 370*t*, 371*t*
 - realities of, 315–317
 - samples, defined, 315
 - across the pyramid, 330
- satisfaction, measuring participant, 238–240
- saturation, of qualitative sample, 370–371
- scale, of program evaluations, 327
- scatter diagrams, 250*t*, 251
- SCHIP. *See* State Child Health Insurance Program (SCHIP)
- scholars, using evaluation results, 13
- scientific contribution, process theories, 143
- scientific literature, effect theory from, 138–139
- scientific rigor
- assessing evaluation for, 400
 - presenting findings of qualitative data with, 375
 - for qualitative methods, 368–369
- scope
- health program, 188
 - measuring patient satisfaction, 240
- score, measuring for change, 339
- screening tests, 187
- secondary data sources, 328–329
- case study methodology using, 365
- secondary prevention, 127
- Seeing Statistics webbook, 357
- selection bias, 300, 302
- self-report questionnaires, 231
- self-report responses, validating, 338
- sense of community, concept of, 60
- sensemaking, program evaluator responsibilities for, 391–392
- sensitivity
- cultural competency and, 49
 - measuring effects evaluation, 278–279
 - screening tests, 187, 187*t*
 - testing magnitude of problem, 94
- sequential data, 365
- seriousness, prioritizing health problems, 112–113
- service charges, 185, 202
- service delivery accountability, 246, 247*t*
- service plan completion, 238
- services. *See* health services
- services utilization data, 337
- services utilization plan
- making recommendations related to, 394–397
 - in process theory, 125, 140*t*
 - types of accountability relating to, 246, 247*t*
- services utilization plan inputs, 185–190
- diagram showing components of, 148, 149*f*
 - eligibility screening, 185–186
 - examples of measures, 232*t*
 - examples that can be measured, 231*f*
 - intervention delivery, 189–190
 - process evaluation of, 220*f*, 241–242
 - quantifying, 230–234
 - queuing, 189
 - social marketing, 185
- services utilization plan outputs
- diagram showing components of, 148, 149*f*
 - examples of measures, 232*t*
 - examples that can be measured, 231*f*
 - process evaluation of, 220*f*, 241–242
 - in program implementation, 191–192
 - quantifying, 234–241
 - coverage as program reach, 234–237
 - dosage measurement, 238, 239*e*
 - program logistics, 240–241
 - satisfaction measurement, 238–240
- SES. *See* socioeconomic status (SES)
- Six Sigma process, 251
- size, sample, 318–319, 320*f*
- skip patterns, 337
- sleeper effect, 354
- small areas or numbers, statistical techniques for, 101–102
- SMART acronym, 155–156
- snowball samples, 318, 324
- social desirability, and response bias, 327
- social health domain, measurement variables, 277*t*
- social marketing
- across the pyramid, 195
 - Internet resources for, 197
 - process evaluation of, 232*t*
 - reaching large audience through, 185
- Social Marketing Institute, 197
- social model of needs, 73*t*, 74
- social perspective model, 74
- social sciences, 322
- social service programs, history of, 6
- societal priorities, good interventions, 134
- sociocultural diversity, 38–39
- socioeconomic groups, needs assessment, 74
- socioeconomic status (SES), 33
- sociology, 288
- software
- for linking data, 338
 - outlining oral presentations using, 393
 - for qualitative data analysis, 374, 377
 - for statistics, 336
- specificity
- screening test, 187, 187*t*
 - testing magnitude of problem, 94
- sponsors, 78
- spreadsheets, entering data into, 336
- spurious finding, 353
- staff. *See* human resources; personnel
- stakeholder participation
- effects evaluation, 267, 282–283
 - full vs. partial-coverage programs, 188
 - gaining consensus from, 143
 - goals and objectives, 147
 - making recommendations to, 396
 - minimizing potential for misuse by involving, 397
 - objections to random assignment, 308
 - planning process, 82–83
 - presenting findings of qualitative data to, 375
 - process evaluation, 228
 - process evaluation results, 13
 - program theory, 138, 143

- public health pyramid, 23
 - questionnaires, 327
 - stakeholders
 - and coalitions, 50–51
 - evaluator to educate, 335
 - standardized intervention, 233
 - standards
 - analyzing indicators, 153
 - data collection methods, 223
 - evaluator responsibility for, 401
 - intervention delivery, 233
 - for model programs, 14
 - overview of, 9
 - planning for effects
 - evaluation, 281
 - training for program
 - delivery, 178
 - State Child Health Insurance Program (SCHIP), 20
 - Statistical Education Through Problem Solving (STEPS) webpage, 357
 - statistical significance, clinical vs., 343
 - statistical software, 336
 - statistical tests
 - example of, 351*t*
 - Internet resources for, 118
 - nonparametric, 347–348, 348*t*
 - quantitative data analysis, 345–353
 - statistics, health problem, 99–102
 - descriptive statistics, 100
 - epidemiology rates, 102, 103*t*
 - mapping geographic information systems, 101, 214
 - for small numbers and small areas, 101–102
 - STATS webpage. *See* Steve's Attempt to Teach Statistics (StATS) webpage
 - statutory regulations, 177–178
 - STEPS webpage. *See* Statistical Education Through Problem Solving (STEPS) webpage
 - Steve's Attempt to Teach Statistics (StATS) webpage, 357
 - strategic planning approach, planning, 68
 - stratification, for target setting, 159, 167
 - strength, dosage, 128
 - success, defining program, 157
 - summative evaluations, 20
 - supplies, cost of, 179
 - surprise findings, 353
 - surveys
 - as linked data, 337
 - questionnaires vs., 325
 - response biases in, 327–328, 328*t*
 - as secondary data sources, 329
 - systems theory, 67
- T**
- TAAPS objective, 148–149, 150*t*, 256
 - target audience
 - in community health assessment, 80–81
 - marketing assessment of, 76
 - as primary data source, 92
 - reaching through services
 - utilization plans, 125
 - setting target values, 157
 - target, community as, 60–61
 - target population, 80, 92
 - measuring participation in, 237
 - in program planning and evaluation, 80–81
 - target-setting decisions and disparities in, 157, 159
 - target values, using data to set, 156–170
 - accepting current value, 160
 - across the pyramid, 171
 - change based on results of statistical test, 164
 - decisional framework for, 156–159, 158*t*
 - decrease in health problem, 166
 - logic statements to developing, 160
 - long-term, 167
 - options, overview of, 160–170, 161–163*f*
 - options, summary of, 165–166*t*, 169–170, 169*t*
 - overview of, 156
 - population-based/multisite programs, 167
 - stratification and, 159
 - Taxonomy of Educational Objectives, 172
 - teams
 - management building, 181
 - planning group, 78
 - technical generation, 6
 - technology
 - in apolitical approach to planning, 66
 - feasibility criterion for good interventions, 133–134
 - management skills in, 182
 - templates, causal theory statement, 109–110, 110*t*
 - terminology, qualitative data analysis, 372
 - tertiary prevention, 127
 - testing effect, 299
 - tests
 - effect evaluation design, 289
 - pilot, 190
 - sensitivity, 187
 - of significance, 101
 - specificity, 187
 - test-retest approach, 280
 - theoretical sampling, 371
 - theory. *See also* program theory
 - critical planning, 67
 - in development phase of planning, 123–124
 - causal. *See* causal theory effect. *See* effect theory espoused. *See* espoused theory impact. *See* impact theory intervention. *See* intervention theory program. *See* program theory systems. *See* systems theory
 - theory-in-use (or theory-in-action), 138, 233
 - time frame, used in objectives, 148
 - time resources
 - budgets and cost of, 203
 - measuring inputs and outputs, 227*t*
 - requirements for qualitative methods, 360–361
 - time-series
 - analysis, 353
 - multiple-group designs, 303–304
 - one-group designs, 302–303
 - time span, effect evaluation design, 291
 - timeframe, what portion of
 - recipients experience
 - what extent of which
 - type of change (TREW) objectives, 149
 - evaluation, timing of, 278, 279*t*
 - outcome documentation, 268, 269*t*, 273
 - questions, intervention effects evaluation, 266
 - timeline
 - intervention effects evaluation, 278
 - organizational plan outputs, 182
 - for short-term health program, 183–184*e*

timing, of effect evaluation, 278, 279*t*

Title V programs, 153

TJC. *See* The Joint Commission (TJC)

Tool for Health and Resilience In Vulnerable Environments (THRIVE), 118

tools of change, 197

total quality management (TQM)

- internal evaluators trained in, 10
- glossary, 260
- quality improvement tools, 249–250*t*
- relevance to health programs, 253

TQM. *See* total quality management (TQM)

training

- program delivery, 178
- qualitative research, 364, 375
- volunteers for implementation, 179

transferability (applicability),

- qualitative methods, 368–369

transportation, 180, 227*t*

trend analysis, 351, 356

TREW objectives. *See* timeframe,

- what portion of recipients experience what extent of which type of change (TREW) objectives

triangulation, 367–368

trigger effect, 354

trigger event, 11

Turning Point Program, University of Washington, 261

two-group, pretest/posttest design, 301–302, 308–309

two-group, prospective (cohort) design, 305–306

two-group, retrospective (case-control) design, 304–305

U

unbounded outcomes, health

- outcomes classifications, 292

uncertainty, defined, 17

undercoverage, 11, 13, 235, 237

underinclusion, eligibility screening

- and, 186–187

unexposed group, 291

unintentional injury mortalities, 31

United Nation Children's Fund (UNICEF), 243

units of data, 372

units of observation, 275

units of service (UOS), 191, 237

- infrastructure, 242
- measuring, 242

univariate statistics, 346

University of Iowa, College of Nursing website, 144–145

University of North Carolina website (UNC), 118

unmeasured factors, adding to

- difference, 342–343, 343*f*

unobtrusive measures, 92

usefulness, assessing evaluation, 400

utilitarian approach, 385

utility

- cost-utility analysis, 213
- as standard of evaluation, 9

utilization-focused evaluations, 6–7, 7*t*

V

validity

- assessing evaluation quality, 401
- assumptions, 141
- community health assessment problems, 84
- of measures, establishing, 281

values

- and attitudes, 47
- and beliefs, 48
- shared, 60

variables

- costs, budgets and, 202
- intervention effects evaluation, 271–274
- levels of measurement, 275–278

variance, monitoring budget, 209, 210–211*e*

venue-based sampling, hard-to-reach populations, 318

vision phase of planning, 79

vital records, 328, 337

voluntary evaluations, 20–21

volunteers, training, 178–179

W

web-based surveys, 216

well-being, public, 5

Western Michigan University, Evaluation Center, 27

Wheel of Public Health Interventions, 145

WHO. *See* World Health Organization (WHO)

wishful thinking, and planning, 82

Women, Infants, and Children (WIC) nutrition programs, 151, 188

word processing software, 190*e*

work flow, process evaluation, 240–241

workforce

- assessment, 77
- training in cultural diversity, 178

workgroup teams, 170

working hypotheses, 368

worksites stress-reduction program, 127

World Health Organization (WHO)

- defining health, 4
- developing health goals, 5
- process evaluation handbook, 243

written open-ended questions, 362–363

Y

years of healthy life (YHL), 97

years of life lost (YLL), 99

- causes of, 98–99*t*

years of life measures, 96–99

years of potential life lost (YPLL), 99

YHL. *See* years of healthy life (YHL)

YLL. *See* years of life lost (YLL)

YPLL. *See* years of potential life lost (YPLL)

Z

Z-test, setting target values, 164