

JAMES J. NEUTENS
LAURNA RUBINSON

RESEARCH

منتدى اقرأ الثقافي
www.iqra.ahlamontada.com

TECHNIQUES

for the Health Sciences

FOURTH EDITION

***Research Techniques for
the Health Sciences***

FOURTH EDITION

Research Techniques for the Health Sciences

FOURTH EDITION

James J. Neutens

The University of Tennessee Graduate School of Medicine/Knoxville

Laurna Rubinson

The University of Illinois/Urbana-Champaign

Benjamin Cummings

San Francisco Boston New York
Cape Town Hong Kong London Madrid Mexico City
Montreal Munich Paris Singapore Sydney Tokyo Toronto

Acquisitions Editor: *Sandra Lindelof*
Associate Editor: *Emily Portwood*
Editorial Assistant: *Jacob Evans*
Managing Editor: *Deborah Cogan*
Production Supervisor: *Mary O'Connell*
Senior Marketing Manager: *Neena Bali*
Manufacturing Buyer: *Jeffrey Sargent*
Cover Designer: *Riezebos Holzbaur Design Group*
Text Designer: *Hespenheide Design*
Production Management and Composition: *Katrina Ostler & Sandy Reinhard, Black Dot Group*
Copy Editor: *Black Dot Group*
Proofreader: *Black Dot Group*
Indexer: *Black Dot Group*
Text Printer: *Courier/Westford*
Cover Printer: *Lehigh-Phoenix Color*
Cover Image Credit: *Ali Mazraie Shadi/iStockphoto*

Copyright © 2010 Pearson Education, Inc., publishing as Pearson Benjamin Cummings, 1301 Sansome St., San Francisco, CA 94111. All rights reserved. Manufactured in the United States of America. This publication is protected by copyright and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission(s) to use material from this work, please submit a written request to Pearson Education, Inc., Permissions Department, 1900 E. Lake Ave., Glenview, IL 60025. For information regarding permissions, call (847) 456-2635.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed in initial caps or all caps.

Pearson Benjamin Cummings™ is a trademark, in the U.S. and/or other countries, of Pearson Education, Inc. or its affiliates.

Library of Congress Cataloging-in-Publication Data

Neutens, James J.

Research techniques for the health sciences / James J. Neutens, Lurna Rubinson.—4th ed.
p. ; cm.

Includes bibliographical references and indexes.

ISBN-13: 978-0-321-59641-3

ISBN-10: 0-321-59641-2

1. Public health—Research—Methodology. 2. Health education—Research—Methodology.

3. Health promotion—Research—Methodology. 4. Medicine—Research—Methodology.

I. Rubinson, Lurna. II. Title.

[DNLN: 1. Public Health. 2. Research—methods. 3. Data Collection—methods. 4. Health Education.

5. Health Services Research. 6. Research Design. WA 20.5 N497r 2010]

RA440.85.R83 2010

610.72—dc22

2009000251

Benjamin Cummings
is an imprint of



ISBN-10: 0-321-59641-2
ISBN-13: 978-0-321-59641-3
www.pearsonhighered.com 1 2 3 4 5 6 7 8 9 10—CRW—13 12 11 10 09

This book is dedicated to:

Mary, Lizzi, Emily, and especially my parents
J.J.N.

Susan, for being there
L.R.



615044

**This book is not for sale or
distribution in the U.S.A. or Canada**

Contents

Preface	xiii
CHAPTER 1 What Is Research?	1
Health Science Research	1
Using Science in the Quest for Knowledge	2
Summary	10
Critical Thinking Questions	11
Suggested Activities	12
References	13
CHAPTER 2 Developing the Research Proposal	14
Selection of the Problem	14
Sources of Problems	15
Statement of the Problem and Research Questions	17
Subproblems	18
Components Comprising the Setting of the Problem	19
Formulation of Hypotheses	20
Significance and Justification of the Problem	21
Résumé of Related Literature	22
Proposed Research Procedures	23
Budget Considerations	23
Time Schedule	23
Research Proposal Checklist	24
Summary	26
Critical Thinking Questions	26
Suggested Activities	27
References	27
CHAPTER 3 Critical Review of the Literature and Information Sources	29
Purposes of the Review	29
Steps in the Review Process	31

Conducting a Computer Search: Finding the Evidence	35
Determining the Level of Evidence	37
Finding Shortcuts to Determining the Level of Evidence	45
Writing the Section on Related Literature	45
Summary	47
Critical Thinking Questions	47
Suggested Activities	48
References	48
CHAPTER 4 Considering Ethics in Research	50
Case Study	50
General Ethical Dilemmas in Human Research	51
Justification to Experiment on Humans	54
Vulnerable Target Groups: Children	55
Vulnerable Target Groups: Elderly	57
Informed Consent: Truth-telling and Deception	58
Right to Privacy and Confidentiality	60
Responsibility for Harmful Consequences	62
The Duty to Continue a Successful Research Effort	63
Therapeutic and Nontherapeutic Research	64
Sponsored Research	64
Publication of Unethical Research	65
Research or Just a Look-See?	65
Role of the Institutional Review Board	66
Case Discussion	70
Summary	70
Critical Thinking Questions	71
Suggested Activities	72
References	72
CHAPTER 5 Conducting Experimental and Quasi-Experimental Research	74
Case Study A	74
Case Study B	74
Characteristics of Experimentation	75
Control in Experiments	76
The Hawthorne Effect in Controlling Situations	77
Advantages and Disadvantages of the Experimental Method	78
Internal and External Validity	78
Constructing Experimental Designs to Control Variables	84

	The Randomized Controlled Trial: Clinical Trials	88
	Case Discussion for Case Study A	94
	Case Discussion for Case Study B	95
	Summary	95
	Critical Thinking Questions	96
	Suggested Activities	96
	References	97
CHAPTER 6	Data Collection Through Surveys and Self-Reports	99
	Case Study	99
	Characteristics of Survey Research	99
	Survey Flow Plan	100
	Survey Design	103
	Data Collection Methods	104
	Survey Sampling	111
	Questionnaire Design and Construction	112
	Attitude Scale Construction	117
	Cover Letter	122
	Pretesting and Questionnaire Revision	122
	Focus Groups	124
	The Delphi Technique	124
	Case Discussion	126
	Summary	127
	Critical Thinking Questions	128
	Suggested Activities	128
	References	129
CHAPTER 7	Sampling Designs and Techniques	133
	Case Study	133
	The Purpose of Sampling	134
	The Sampling Frame	134
	Sampling Techniques	135
	Sample Size	141
	Sample Size for Studies with Hypothesis:	
	Analytical Studies	142
	Sample Size for Studies Without Hypotheses:	
	Surveys and Descriptive Studies	143
	Case Discussion	149
	Summary	150

	Critical Thinking Questions	151
	Suggested Activities	151
	References	152
CHAPTER 8	Qualitative Research	153
	Case Study	153
	Characteristics of Qualitative Research	153
	Theoretical Foundations	157
	Methods of Qualitative Research	158
	Techniques of Collecting Qualitative Data	170
	Analyzing Qualitative Data	176
	Using Computers in Data Analysis	178
	Case Discussion	179
	Summary	179
	Critical Thinking Questions	180
	Suggested Activities	180
	References	181
CHAPTER 9	Evaluation Research	182
	Case Study	182
	Introduction	182
	Purposes of Evaluation Research	183
	What Can Be Evaluated?	184
	Steps in Conducting Evaluation	184
	Evaluation Models	188
	Types of Evaluation Research	197
	Methodological Approaches in Evaluation Research	204
	Case Discussion	210
	Summary	210
	Critical Thinking Questions	211
	Suggested Activities	212
	References	212
CHAPTER 10	Analytical Epidemiologic Studies	214
	Case Study A	214
	Case Study B	214
	The Nature of Epidemiology	215
	Analytical Methodologies in Epidemiology	215
	Cohort Investigations	217

Case-Control Studies	218
Establishing Causation	220
Problems of Error	220
Control in Epidemiological Research	222
Analysis of Results in Analytic Epidemiology	224
Case Discussion of Case Study A	232
Case Discussion of Case Study B	232
Summary	233
Critical Thinking Questions	234
Suggested Activities	234
References	235
CHAPTER 11 Analyzing and Interpreting Data: Descriptive Analysis	237
Case Study	237
The Meaning of Statistics	238
Statistical Analysis and Data	238
Descriptive Data Analysis Techniques	240
Personal Computers and Information Delivery Systems	250
Case Discussion	252
Summary	253
Critical Thinking Questions	253
Suggested Activities	253
References	255
CHAPTER 12 Analyzing and Interpreting Data: Inferential Analysis	256
Case Study A	256
Case Study B	257
Inferential Analysis	257
Estimation of Parameters	258
Hypotheses Testing	259
Inferential Data Analysis Techniques for Comparing Mean Scores	263
Measures of Relationship and Predictions	268
Nonparametric Tests of Significance	269
Meta-Analysis	271
Case Discussion of Case Study A	273
Case Discussion of Case Study B	273
Summary	273
Critical Thinking Questions	274

	Suggested Activities	274
	References	275
CHAPTER 13	Techniques for Data Presentation	276
	Case Study	276
	Table Presentations	276
	Table Format	281
	Figure Presentations	286
	Graphics and the Computer	293
	Case Discussion	293
	Summary	293
	Critical Thinking Questions	294
	Suggested Activities	294
	References	294
CHAPTER 14	Writing a Research Report	295
	The Report as a Communication Document	295
	Preliminaries	296
	Text or Main Body of the Report	299
	References	306
	Appendixes	307
	Writing Style	307
	Summary	308
	Critical Thinking Questions	309
	Suggested Activities	309
	References	309
APPENDIX A	Common Statistical Procedures.	311
APPENDIX B	World Wide Web Research.	314
	Name Index.	317
	Subject Index	321

Preface

The impetus for the first edition of this book was to fill a void specific to health education and the health science field. The following editions built upon feedback from faculty and students as did this, the fourth, edition. The trend of the book expanded slightly to include those in medical education as well as health education. That theme continues in this edition with the section on clinical randomized trials.

The target audience, however, remains basically the same. This edition is intended for (1) upper level undergraduate and beginning graduate students in the health sciences and health education; (2) practitioners in the fields of school or community health education, public health, medical education, and allied health; (3) professionals in health-related disciplines.

New to the Fourth Edition

We are delighted to make some significant changes to the fourth edition. Each chapter begins with a list of key terms to let students know the essentials of the chapter. New case studies present realistic research efforts that unfold throughout the chapter, bringing research to life in a pragmatic fashion. Critical Thinking Questions are placed at the end of the chapter to emphasize the principal components. The Suggested Activities section at the very end of each chapter have been updated and offer both instructors and students opportunities to apply key concepts presented in the chapter. The following chapter descriptions highlight the content changes in each chapter and give general chapter overviews. The references have been updated throughout, illustrating the most recent research efforts in the health fields. On a final note, instructors teaching this course can now access PowerPoint® Lecture Outlines for each chapter of the book, available for download through the Instructor Resource Center (aw-bc.com/irc).

Chapter Organization and Descriptions

Chapter 1, *What Is Research?*, provides the basics of research with emphasis on a theoretical foundation so students can understand and appreciate the entire process.

Chapter 2, *Developing the Research Proposal*, offers an overview of the research project, in particular a thesis or dissertation. Discussion of hypothesis development is explained in greater detail than in prior editions.

Chapter 3, *Critical Review of the Literature and Information Sources*, presents information well beyond the usual literature review. Worksheets offer a means to evaluate an article based on the degree of research evidence it provides rather than just on the information it reports. Evidence-based, online sources are described to help students find information quickly.

Chapter 4, *Considering Ethics in Research*, provides both historical and current perspectives on research, including a newly viewed vulnerable population: the elderly. IRB review steps are addressed.

Chapter 5, *Conducting Experimental and Quasi-Experimental Research*, goes beyond the usual comparison of experimental and quasi-experimental designs to discuss clinical trials, particularly the randomized controlled trial. Different clinical trial designs are presented for use.

Chapter 6, *Data Collection Through Surveys and Self-Reports*, offers a step-by-step approach for conducting surveys with updates for email and web-based survey techniques. Attitude scale construction is viewed from three approaches. Focus groups and the Delphi technique are discussed, also.

Chapter 7, *Sampling Designs and Techniques*, contrasts probability and non-probability sampling designs. The section on sample size is completely revised to present information on sample size techniques for studies with hypotheses and those without hypotheses.

Chapter 8, *Qualitative Research*, deals with a subject that has become increasingly important in health science efforts and appears to be of great interest to the medical field, given changing health care delivery systems and questionable outcomes. Qualitative research is contrasted with quantitative research. Methodologies, including ethnology, are pragmatically discussed. Updated techniques of collecting qualitative data are detailed, and methods of analyzing and coding such data are presented.

Chapter 9, *Evaluation Research*, outlines the steps in this process and provides several models to the student. Cost analysis is also discussed here.

Chapter 10, *Analytical Epidemiologic Studies*, discusses cohort and case control research methods, as well as ways to analyze the results in such studies. Practical examples are provided to illustrate the methods.

Chapter 11, *Analyzing and Interpreting Data: Descriptive Analysis*, provides examples for students to understand how to calculate measures of central tendency, variation, and correlation. Nominal, ordinal, interval, and ratio measurements and data are contrasted with each other.

Chapter 12, *Analyzing and Interpreting Data: Inferential Analysis*, expands the discussion from Chapter 11 to discuss two principal types of statistical inference: estimation of parameters and hypothesis testing. The latter is presented in a step-by-step fashion for ease of understanding. Data analysis techniques are addressed in conjunction with Type I and Type II errors. Measures of relationship and prediction are presented as well as the necessary steps for a good systematic review of the literature leading to meta-analysis.

Chapter 13, *Techniques for Data Presentation*, suggests the best ways for students to present their data and findings. Table and figure presentations are discussed, as is the

use of graphs, charts, and photographs, and techniques for making use of new technology available for data presentation.

Chapter 14, *Writing a Research Report*, is instrumental for students who are defending a thesis or dissertation. The underlying emphasis is that their document is a communication piece. The reader is given a step-by-step approach for developing a sound report for acceptance by other health care professionals as well as the public.

Appendix A, *Common Statistical Procedures*, lists 27 procedures, with a brief explanation of their common usage.

Appendix B, *World Wide Web Research*, provides sites that are useful in conducting research in the health education and health care arenas.

Acknowledgments

First and foremost we thank the professors and students who offered suggestions to make this the best edition thus far. Their insight and day-to-day use of the book makes this edition more pragmatic and understandable with more examples and explanations.

We are indebted to Emily Portwood, Associate Editor of Applied Sciences at Benjamin Cummings for her diligence, continued guidance, and encouraging nudges. Also thanks to Mary O'Connell and the Black Dot Group for a superb job of ushering the book through copyediting and production. Finally, we thank our family and friends for their patience and support.

J.J.N. and L.R.

List of Figures

FIGURE 1.1	The Health Belief Model	5
FIGURE 1.2	The Cyclical Stages of the Research Project	9
FIGURE 3.1	Theoretical Core and Subheadings	46
FIGURE 4.1	Some participants in the Tuskegee Syphilis Study	52
FIGURE 4.2	Table from 1969 depicting number of participants in the Tuskegee Syphilis Study showing number of syphilitic patients and number of controlled nonsyphilitic patients.	53
FIGURE 5.1	Internal Validity Design	79
FIGURE 5.2	Clinical Trial Phase Comparisons	89
FIGURE 5.3	Crossover Design for Randomized Clinical Trials (RCT)	90
FIGURE 5.4	Factorial Design with Two Independent Variables	91
FIGURE 6.1	Survey Flow Plan	102
FIGURE 6.2	Cover Letter	123
FIGURE 7.1	The Pathway and Relationship from Population to Sampling Unit	136
FIGURE 9.1	The Evaluation Process	185
FIGURE 10.1	Timing Schematic for Cohort and Case-Control Prospective and Retrospective Investigations	216
FIGURE 10.2	A 2 × 2 Table Illustration of Cohort and Case-Control Methodologies	216
FIGURE 11.1	Correlation Scatterplots	248
FIGURE 12.1	Percent of areas under normal curve with 2-scores	258
FIGURE 12.2	Two-Tailed and One-Tailed Regions	261
FIGURE 13.1	Example of Line Graph	287
FIGURE 13.2	Example of Bar Graph	287
FIGURE 13.3	Example of Circle Graph	288
FIGURE 13.4	Example of Scatter Graph	289
FIGURE 13.5	Example of Organizational Chart	290
FIGURE 13.6	Mean Attitude Scores of Eighth-Grade Students Over 12 Months	291
FIGURE 14.1	Title Page	297
FIGURE 14.2	Table of Contents	298
FIGURE 14.3	Abstract	300

What Is Research?

KEY TERMS

dynamic
heuristic
models

research
research process
scientific approach

static
theory

Human beings possess the ability to think rationally and logically, which in turn leads to curiosity. You may have heard of the Philosophy 101 final examination that contained just one question: “*Why?*” Students wrote up to 20 pages, quoting philosophers ranging from Aristotle and Buber, but the correct answer was “*Because.*” **Research**, or the process thereof, answers the question *Why?* There are multiple reasons for a myriad of questions, but how can we know what the best answer is? Research and the application of the scientific method will enable us to answer such questions.

Health Science Research

The health science profession had its beginnings at the start of the nineteenth century, when improvement of the health of school-age children provided impetus for the new discipline. In addition, we health science professionals have strived to provide information to those populations of high risk. These have been our primary goals, and we have attempted to use research and evaluation to improve our ability to meet these objectives. However, criticism about the research conducted can be found in the health science literature. With these criticisms in mind, you might be asking: *Why study health science research? What can it do? What can we expect?* All professionals, including community and school health educators, nurses, physicians, and other health care providers, are involved in the research process.

These questions will be answered as you progress through the text and get better acquainted with the process and product of research in the health sciences. Even though the discipline has been criticized, research in the health sciences has proved to be

valuable. In the school setting alone, research has shown that children have increased their knowledge and altered their attitudes in such areas as smoking, human sexuality, dental health, cardiovascular diseases, drug and alcohol abuse, and driver education.

Research endeavors in the schools can inculcate good and acceptable health behaviors in youths if the programs being tried allow for decision-making and problem-solving skills, improve self-concept and self-esteem, and provide additional social interactions. Such programs, under the auspices of rigorous research, can lead to a reduction of risk factors associated with well-being.

Research in the community, conducted by nurses and physicians, can yield baseline information regarding needs assessments in relationship to health status studies. In addition, public health policies are developed through research in the community. The determination of effective policy strategies can best be accomplished through rigorous and thorough research studies.

Patient education has been the setting for many research projects, including studies on diabetes, hypertension, postsurgical procedures, nutrition, and weight reduction. Many studies, for example, have led to major advances in diabetes control in which patients are able to self-medicate and be relatively free of hospital regimens. Determination of how overweight patients react to specific weight control programs has had an influence on recommendations for proper nutrition and exercise for both youths and adults.

For students to gain a thorough understanding of research and its place in the health sciences, they must have a working knowledge of science, scientific inquiry, and the importance of theory in research.

Using Science in the Quest for Knowledge

Knowledge may be gained or accumulated in many ways. Cohen and Manion (2000) classified ways of knowing into three broad categories: experience, reasoning, and research. We are concerned with the scientific method, or science, and how this science helps us know.

To satisfy our doubts, . . . it is necessary that a method should be found by which our beliefs may be determined by nothing human, by some external permanency, by something upon which our thinking has no effect. . . . The method must be such that the ultimate conclusion of every man will be the same. Such is the method of science. Its fundamental hypothesis . . . is this: There are real things, whose characters are entirely independent of our opinions about them. (Buchler, 1995, p. 42)

Scientists, in their quest for knowledge and truth, use self-correcting devices that serve as built-in checking methods to assure that the conclusions they may reach are factual. Hypotheses are formulated, but so too are alternate hypotheses to test the objectivity of the experiment and experimenter. In addition, by publishing the experiment and its results, scientists allow for others to replicate and inspect their work.

All scientific fields (physics, engineering, psychology, health, etc.) have a method of arriving at knowledge, which will be discussed later as the scientific approach. Science can be considered a method to solve problems or answer questions that investigators

find to be of interest. Scientists acquire specific attitudes that enable them to think and act in a scientific manner. These attitudes are best described by Ary et al. (2006, pp. 13–14):

1. Scientists are essentially doubters, who maintain a highly skeptical attitude toward the data of science. Findings are regarded as tentative and are not accepted by the scientists unless they can be verified. Verification requires that others must be able to repeat the observation and obtain the same results. Scientists want to test opinions and questions concerning the relationships among natural phenomena. Furthermore, they make their testing procedures known to others in order that they may verify, or fail to verify, their findings.
2. Scientists are objective and impartial. In conducting observations and interpreting data, scientists are not trying to prove a point. They take particular care to collect data in such a way that any personal biases they may have will not influence their observations. They seek truth and accept the facts even when they are contrary to their own opinions. If the accumulated evidence upsets a favorite theory, then they either discard that theory or modify it to agree with the factual data.
3. Scientists deal with facts, not values. They do not indicate any potential moral implications of their findings; they do not make decisions for us about what is good or what is bad. Scientists provide data concerning the relationship that exists between events, but we must go beyond these scientific data if we want a decision about whether or not a certain consequence is desirable. Thus, while the findings of science may be of key importance in the solution of a problem involving a value decision, the data themselves do not furnish that value judgment.
4. Scientists are not satisfied with isolated facts but seek to integrate and systemize their findings. They want to put the things known into an orderly system. Thus scientists aim for theories that attempt to bring together empirical findings into a meaningful pattern. However, they regard these theories as tentative or provisional, subject to revision as new evidence is found.

Science as Static and Dynamic

Conant (1951) describes science in two ways: the **static** and the **dynamic**. The emphasis in the *static view* is on the present state of knowledge and of scientists contributing to that state. In addition, the extent of knowledge and the present theories, hypotheses, and principles are considered. In this view science may be used to explain observations and to discover new facts that contribute systematized information to the existing body of knowledge.

The actions of scientists are considered to be the *dynamic view* of science. In this view, the present state of knowledge serves as a base for further inquiry. The **heuristic** view of science (a subset of the dynamic view) means science that discovers or reveals, including the idea of self-discovery. The emphasis here is on discovery of something new that can add to the present base or body of knowledge. In the heuristic method of scientific inquiry, the actual emphasis is on the investigator being imaginative in his or her approach for answering a question or solving a problem.

Science and Theory

The ultimate goal of scientific inquiry is to formulate theories. Theories provide a way to conceptualize, organize, integrate, and classify the facts that scientists accumulate. A **theory** can describe a tentative explanation of some phenomenon. As scientists, when we ask the question *Why?* and attempt to answer it, we are formulating a theory. It is then verified through evidence obtained by either observation or experimentation.

Several characteristics of a sound theory serve to illustrate the constraints that scientists must work within when formulating a theory:

1. A theory should be able to explain the observed facts relating to a particular problem; it should be able to propose the “why” concerning the phenomena under consideration. This explanation of events should be in the simplest form possible. A theory that has fewer complexities and assumptions is favored over a more complicated one. This statement is known as the principle of parsimony.
2. A theory should be consistent with the observed facts and with the already established body of knowledge. We look for the theory that provides the most probable or the most efficient way of accounting for the accumulated facts.
3. A theory should provide means for its verification. This is achieved for most theories by making deductions in the form of hypotheses stating consequences that one can expect to observe if the theory is true. The scientist can then investigate or test these hypotheses empirically in order to determine whether the data support the theory. It must be emphasized that it is inappropriate to speak of truth or falsity of a theory. The acceptance or rejection of a theory depends primarily on *utility*. A theory is useful or it is not useful, depending on how efficiently it leads to predictions concerning observable consequences, which are then confirmed when the empirical data are collected. Even then, any theory is considered tentative and subject to revision as new evidence accumulates.
4. A theory should stimulate new discoveries and indicate further areas in need of investigation. (Ary et al., 2006, pp. 15–16)

The health sciences have been very slow in achieving theoretical bases, probably because health, along with many other social sciences, is a young science. For the past 50 years, health professionals have been collecting data to gather empirical evidence and build toward theoretical constructs. While several models (such as the Health Belief Model and the PRECEDE Model) have been developed for and by health educators, theories that could be directly attributed to the health sciences have been nonexistent.

It must be understood that there is a difference between a model and a theory. Theories provide an understanding of a phenomenon and offer prediction and control. Theories also can provide a way for conceptualizing the world. However, **models** provide perhaps a better way of conceptualizing. Some models are replicas, such as miniature toys; others are symbolic, such as the diagram of the Health Belief Model (see Figure 1.1). Models also can be used in computers, as scientists have been able to program computers to behave in a humanlike manner when solving problems. Models provide us with a simplistic way of looking at complex problems or phenomena.

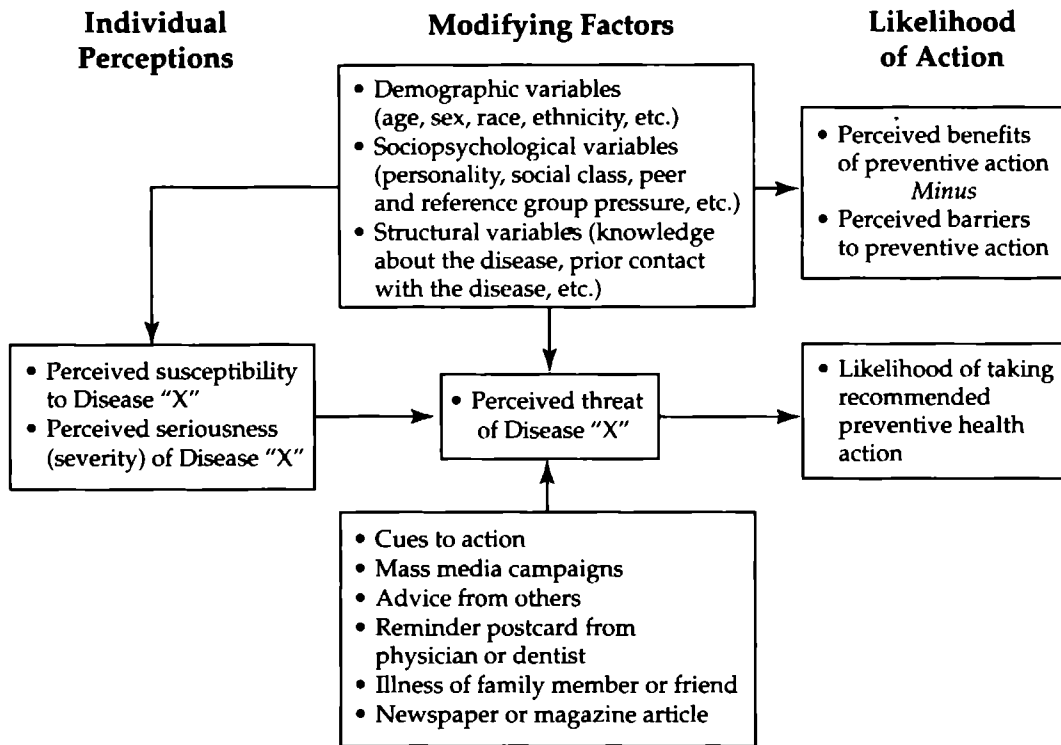


Figure 1.1 The Health Belief Model

Source: From Becker, M. (1974). *H. Ed. Monographs*, 2, 409–419. Copyright © 1974. Reprinted by permission of John Wiley & Sons, Inc.

Basic and Applied Research

When we think about research, our initial image is of a laboratory with animals and scientists in white coats. This image is generally true when basic researchers are at work. *Basic research* aims to expand the knowledge base by formulating, evaluating, or expanding a theory. Research in the medical sciences is usually of this type because biochemistry, biology, and microbiology fall into this pattern. Hence, the primary purpose of basic research is discovering knowledge for the sake of knowledge alone; the practical side of the issue is considered at a later time.

Applied research aims to solve practical problems, although it uses the same characteristics as basic research. Here, theoretical concepts are tested in real situations. Because laboratories cannot be the scene for investigations, the real world (e.g., classrooms, hospitals, clinics) becomes the laboratory for applied health researchers. Most research in the health sciences is considered applied research because it is concerned with testing the processes of health behavior in real-life situations.

Several major health science projects have used the applied research approach. These include the CATCH program, the Healthy Cities program, and the University of Minnesota smoking prevention program. These projects use similar conceptual and theoretical

approaches in schools—that is, in the real world. However, whether basic or applied research is being conducted by health scientists, one type of research must depend on the other for the proper research process to take place. Applications of theories help solve some practical problems, such as when social learning theory is used to attempt to explain why children adopt healthful behaviors. This can work the other way as well, as when theoretical concepts are advanced by the practical use of theory. As in the example just mentioned, new light could be shed on social learning theory if it were found to be useful in classroom situations.

The Scientific Approach

To become familiar with the method of scientific inquiry, future researchers should be attuned to several characteristics of the **scientific approach** (or research process). Rather than reiterate the major traits associated with research, we will use the excellent list provided by Best and Kahn in *Research in Education* (1998, pp. 18–20):

1. Research is directed toward the solution of a problem. The ultimate goal is to discover cause-and-effect relationships between variables, though researchers often have to settle for the useful discovery of a systematic relationship, for lack of enough evidence to establish one of cause and effect.
2. Research emphasizes the development of generalizations, principles, or theories that will be helpful in predicting future occurrences. Research usually goes beyond the specific objects, groups, or situations investigated and infers characteristics of a target population from the observed. Research is more than information retrieval, the simple gathering of information. Although many school research departments gather and tabulate statistical information that may be useful in decision making, these activities are not properly termed research.
3. Research is based upon observable experience or empirical evidence. Certain interesting questions do not lend themselves to research procedures because they cannot be observed. Research rejects revelation and dogma as methods of establishing knowledge and accepts only what can be verified by observation.
4. Research demands accurate observation and description. Researchers use quantitative measuring devices, the most precise form of description. When this is not possible or appropriate, they use qualitative or nonqualitative descriptions of their data-gathering procedures and, when feasible, employ mechanical, electronic or psychometric devices to refine observation, description, and analysis of data.
5. Research involves gathering new data from primary or firsthand sources or using existing data for a new purpose. Teachers frequently assign a so-called research project that involves writing a paper dealing with the life of a prominent person. The students are expected to read a number of encyclopedias, books, or periodical references, and synthesize the information in a written report. This is not research, for the data is not new. Merely reorganizing or restating what is already known and has already been written, valuable as it may be as a learning experience, is not research. It adds nothing to what is known.
6. Although research activity may at times be somewhat random and unsystematic, it is more often characterized by carefully designed procedures, always applying rigorous

analysis. Although trial and error are often involved, research is rarely a blind, shotgun investigation—trying something to see what happens.

7. Research requires expertise. The researcher knows what is already known about the problem and how others have investigated it. He or she has searched the related literature carefully and is also thoroughly grounded in the terminology, the concepts, and the technical skills necessary to understand and analyze the data gathered.
8. Research strives to be objective and logical, applying every possible test to validate the procedures employed, the data collected, and the conclusions reached. The researcher attempts to eliminate personal bias. There is no attempt to persuade or to prove an emotionally held conviction. The emphasis is on testing rather than on proving the hypothesis. Although absolute objectivity is as elusive as pure righteousness, the researcher tries to suppress bias and emotion in his or her analysis.
9. Research involves the quest for answers to unsolved problems. Pushing back the frontiers of ignorance is its goal and originality is frequently the quality of a good research project. However, previous important studies are deliberately repeated, using identical or similar procedures, with different subjects, different settings, and at a different time. This process is replication, a fusion of the words repetition and duplication. Replication is always desirable to confirm or to raise questions about the conclusions of a previous study. Rarely is an important finding made public unless the original study has been replicated.
10. Research is characterized by patient and unhurried activity. It is rarely spectacular and researchers must expect disappointment and discouragement as they pursue the answers to difficult questions.
11. Research is carefully recorded and reported. Each important term is defined, limiting factors are recognized, procedures are described in detail, references are carefully documented, results are objectively recorded, and conclusions are presented with scholarly caution and restraint. The written report and accompanying data are made available to the scrutiny of associates or other scholars. Any competent scholar will have the information necessary to analyze, evaluate, and even replicate a study.
12. Research sometimes requires courage. The history of science reveals that many important discoveries were made in spite of the opposition of political and religious authorities. The Polish scientist Copernicus (1473–1543) was condemned by church authorities when he announced his conclusion concerning the nature of the solar system. His theory that the sun, not the earth, was the center of the solar system, in direct conflict with the older Ptolemaic theory, angered supporters of prevailing religious dogma, who viewed his theory as a denial of the story of creation as described in the book of Genesis. Modern researchers in such fields as genetics, sexual behavior, and even business practices had personal convictions, experiences, or observations that were in direct conflict with some of the research conclusions.

Upon reading the previous list, you may get a view of researchers that is not realistic but ideal; imaginative, honest, hard working, very rigid, and probably boring—because all they know is the subject they are so relentlessly pursuing. We can argue that this description is not accurate, especially for the health science researcher. The health professional is

usually people-oriented and thus conducts research in real-world settings: hospitals, schools, places of worship, community centers, and so on. However, the good health researcher seeks to be rigorous and to adhere to scientific standards at all times.

Because we have set as one of the goals of this text to provide a basis from which students could conduct research, we will list and briefly explain the stages of the **research process**. Consideration here should be given to the nature of the research process (i.e., that one component is integral to all others, and that good research becomes almost cyclical). Figure 1.2 points out the various stages of the research process and attempts to show its cyclical nature.

Selecting a Problem The health science student will probably decide on a subarea of interest to focus his or her research upon. Selecting a problem involves asking good questions and communicating with others who might be familiar with your research topic. Colleagues, other students, supervisors, and faculty are some people you might ask for assistance in formulating your problem. Below are some examples of research problems that are usually stated in the form of a question:

1. What was the extent of the impact of the reduction of federal aid to dependent children and their families?
2. What factors influence eating disorders among various ethnic/racial groups?
3. What are the effects of a cardiac rehabilitation program?

Formulating Hypotheses The hypothesis is the researcher's tentative explanation that will predict the significant results of the research study or process. However, the hypotheses are always supported by theory and/or previous research. Examples of hypotheses that relate to the problems in the previous section are:

1. Dependent children and their families who receive less federal aid will have lowered health status than those dependent children and their families who have not been affected by the reduction program.
2. Disordered eating is more prevalent among Blacks, Asians, and Latinos as compared to Whites.
3. Those patients who complete all phases of a cardiac rehabilitation program will have a better quality of life than those who do not complete the program.

Reviewing the Literature Relevant literature provides the hypotheses and initial problem selection. In addition, a thorough review of material may lead to suggested investigative methods.

Listing the Measures Identifying all the possible measures enables the researcher to tighten the hypotheses by eliminating and rethinking those that have no available measure or none that can be developed for use in the study.

Describing the Subjects The researcher carefully describes and considers the types of subjects necessary for the project. Particular care should be given to the number and availability of the subjects.

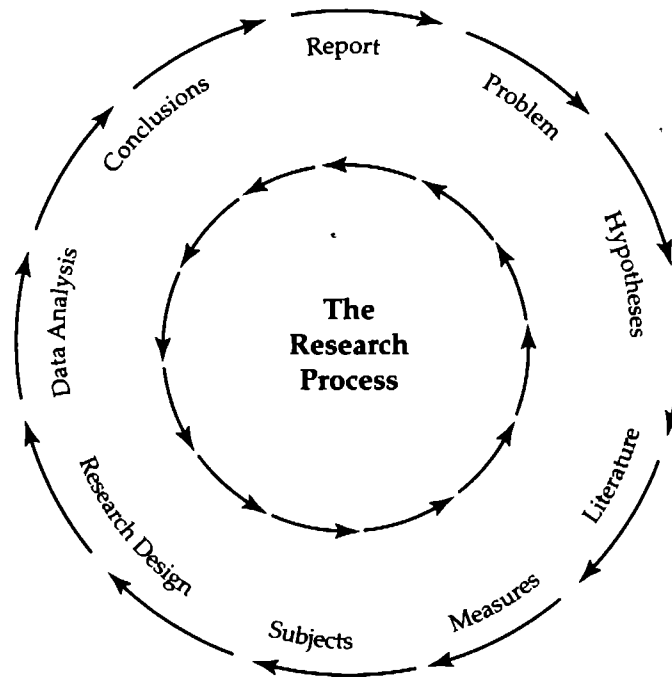


Figure 1.2 The Cyclical Stages of the Research Project

Constructing a Research Design The research design should be fully explained so that the researcher is sure that the design will allow the student to test the stated hypothesis. Chapter 5 discusses the various types of experimental designs.

Constructing and Identifying Measurement Devices The adoption and/or construction of appropriate instruments is used to measure the selected variables. There are some standardized instruments in the health sciences; however, modification of these instruments, the construction and pilot testing of new ones, or the construction of questionnaires and interview schedules may be necessary.

Analysis of the Data A plan to analyze the data should be carefully considered so that the number of subjects, the instruments, and the method of recording the data all coincide to fit the analysis procedure. This is an extremely important part of the stages of the research process because all too often the instrumentation is not geared for appropriate data analysis, therefore rendering results improper or inadequate.

Generating Conclusions The data should reveal several conclusions that are directly related to the hypotheses.

Writing the Report of Research Chapter 14 details report writing, and instruction is offered for each section of the report.

Research Methodologies in the Health Sciences There are several methodologies that are utilized by the health science investigators, each of which will be described in detail in the subsequent chapters. These methods, in brief, are:

1. **Experimental research** is a study in which the investigator controls and manipulates one or more of the variables. The focus of experimental research is on the relationships between the variables. The major purpose of this type of research is to determine what will happen.
2. **Survey (interview and observational) research** are considered descriptive *methodologies* in which the results reveal what is happening in a particular occurrence. This research involves recording, describing, analyzing, and interpreting conditions that presently exist. Comparisons and contrasts are attempted to reveal relationships between the nonmanipulated variables.
3. **Evaluation research** is a method of assessing a process or program in a specific situation (Wiersma, 2005). Evaluation may establish clear and specific criteria for success. Evidence is collected from a sample of the population, translated into quantitative terms, and compared with the previously set criteria. Conclusions are then drawn about the effectiveness, merit, and success of the program that was studied.
4. **Historical research** describes what has occurred in the past. “The process involves investigating, recording, analyzing, and interpreting the events of the past for the purpose of discovering generalizations that are helpful in understanding the past, understanding the present, and to a limited extent, in anticipating the future” (Best & Kahn, 2006, p. 24).

The type of methodology used is most often determined by the questions asked and the kinds of data that will be collected. Too often, inexperienced investigators will attempt an inappropriate methodology for their convenience. A thorough discussion of these methodologies can be found in the remaining chapters of the text.

The health sciences have grown at a very fast pace since the early 1950s, when they became a separate and valued part of the educational process. During this time, the health field has become increasingly sophisticated in its approach to research. From the early days of one-group, no-control studies, to today’s Solomon Four-Group designs, which include correlation of biomedical data, we have seen health science researchers publish and present their work in prestigious journals and meetings.

SUMMARY

Human beings possess the ability to think rationally and logically, which, in turn, leads to curiosity. *Why?*, the most asked question, leads researchers to conduct studies to find answers to questions, sometimes simplistic and other times complex.

Research in the health sciences is still in its early years, because the profession is relatively young. Even though they have their critics, and most of them are on target, health scientists have dramatically advanced in the field’s research and evaluation efforts in the

1970s and 1980s. Using science in the quest for knowledge, health scientists have become imaginative, rigorous, and conscientious in their approach to research.

The method of science has enabled us to accumulate knowledge in many ways. Scientists assure themselves and the public that their conclusions are based on fact by having built-in checking mechanisms to ensure the accuracy, replication, and inspection of their work. Scientists have very specific attitudes about their work that set them apart from the layperson attempting research. Science has two broad views: static and dynamic, in which the latter describes the actions of scientists—how they think and behave to solve intricate problems.

The ultimate goal of scientific inquiry is to formulate theories. Theories enable us to conceptualize the facts that investigators accumulate. Several characteristics are inherent in a sound theory, each depicting the constraints under which the scientists work. The health sciences have been slow in developing theories but rely on those of other social sciences such as psychology, anthropology, and education. Researchers in the health field have developed several models (e.g., the Health Belief Model and the PRECEDE Model), which are different from theories. Models provide a simplistic way of looking at complex problems.

There are distinct differences between basic and applied research. However, each may complement the other. Professionals in the health sciences have concentrated upon applying principles and theories for real-world situations. It is difficult to conduct basic research in the settings that are available to health educators: schools, nursing homes, clinics, hospitals, and so on.

The scientific approach is the process investigators use in the quest for knowledge. Several important characteristics of the process enable standards to be scientific and rigorous. Stages of the scientific process begin with the selection of a problem and proceed through the writing of a research report.

Four general methodologies are used in health science research: experimental, survey (interview and observational), evaluation, and historical. The type of methodology required is dictated by the questions that need to be asked and the data that are collected.

CRITICAL THINKING QUESTIONS

1. Discuss why it is important to use scientific inquiry in the research process. Give at least two examples.
2. Do you believe that the Health Belief Model is relevant, considering our present health issues?
3. What is your definition of *theory*?
4. How would you choose to conduct your research? What type(s) of methodology would you utilize?
5. Your professor has asked you to prepare a presentation to your class that differentiates between theory and practice. Using a specific theoretical tenet, depict how you would get from that theory to placing it into a practical situation.

SUGGESTED ACTIVITIES

1. Revise each of the following research topics so that it would be feasible for a research project. Indicate whether the statement is in the form of a hypothesis or a statement of a problem.
 - a. Smoking among adolescents
 - b. Diabetes control at home
 - c. Hypertension in the elderly
 - d. Fitness among stroke victims
 - e. Cardiovascular rehabilitation for angina patients
 - f. Dental health for pregnant teenagers
 - g. Alcohol abuse among homemakers
2. Describe in one sentence each of the following characteristics of the research process:
 - a. Research is directed toward the solution of a problem.
 - b. Research emphasizes the development of generalizations, principles, or theories that will be helpful in predicting future occurrences.
 - c. Research is based on observable experience or empirical evidence.
 - d. Research demands accurate observation and description.
 - e. Research involves gathering new data from primary or firsthand sources or using existing data for a new purpose.
 - f. Although research activity at times may be somewhat random and unsystematic, it is more often characterized by carefully designed procedures.
 - g. Research requires expertise.
 - h. Research strives to be objective and logical.
 - i. Research involves the quest for answers to unsolved problems.
 - j. Research is characterized by patient and unhurried activity.
 - k. Research is carefully recorded and reported.
 - l. Research requires courage.
3. Devise your own definition of research and defend it in a short paper.
4. Discuss how the health sciences have contributed to the body of knowledge concerning one aspect of health behavior.
5. Using the Web, determine five health topics that would be interesting to you in order to complete a research report.

REFERENCES

- Ary, D., Razavieh, A., Sorensen, L., Jacobs, C. & C. Sorenson (2006). *Introduction to research in education*. Belmont CA: Wadsworth.
- Becker, M. (1974). The health belief model and personal health behavior. *H. Ed. Monographs*, 2, 409–419.
- Best, J., & Kahn, J. V. (1998). *Research in education* (8th ed.). Boston: Allyn & Bacon.
- Best, J., & Kahn, J. V. (2006) *Research in education* (10th ed.). Boston: Allyn & Bacon.
- Buchler, J. (Ed.). (1995). *Philosophical writings of Pierce*. New York: Dover.
- Cohen, L., & Manion, L. (2000). *Research methods in education* (2nd ed.). New York: Routledge.
- Conant, J. (1951). *Science and common sense*. New Haven: Yale University Press.
- Wiersma, W. (2005). *Research methods in education*. Boston: Allyn & Bacon.

Developing the Research Proposal

KEY TERMS

hypothesis

research proposal

statement of the problem

researchable problems

research questions

subproblems

Much research ends in futility because the neophyte rushes into research activity—choosing a sample, collecting data, deriving conclusions—with only a meager plan at best. To be successful, the researcher must have a detailed plan as well as an overall conceptualization. The **research proposal** allows the investigator to specify the problem and related components; elaborate on the significance of the research to the health profession; review related literature; and outline the appropriate methodology within an equitable time frame. The sequence employed throughout this chapter is the format used in most theses and dissertations, although each university, not unlike most requests for proposals in funded research, will have modifications that must be followed by the researcher.

Selection of the Problem

One of the most difficult tasks confronting the beginner is to select a researchable problem. More often than not, the newcomer has a proclivity to tackle an exotic issue, thus making the problem either too broad or too narrow in scope. Some factors that should be involved in the ultimate selection are listed here (Bailey, 1994):

1. **Interest:** The researcher should be interested in pursuing the problem area. It should relate to the background and career interests of the student and help develop useful skills for the future.
2. **Operability:** The nature of the problem should be such that the researcher has both the resources and the time available to complete the subject.
3. **Scope:** While the research problem should not attempt to solve all the health dilemmas of the world, neither should it be so small as to negate the variables necessary for adequate results.

4. **Theoretical and practical values:** The research should contribute to the health field, perhaps through publication, and be of benefit to health practitioners.
5. **Health paradigm:** This is the school of thought or model employed by the researcher. For example, the Health Belief Model (James, Champion, & Strecher, 2003) or the PRECEDE paradigm (Green & Kreuter, 2005) could serve as a source to a problem or as a methodological direction.
6. **Values of the researcher:** The myth of value-free research is just that, a myth. The student of research should be aware that in addition to being unstable, values may prejudice the research effort to the degree that all objectivity is lost. Note that even the selection of a problem is value-laden.
7. **Research methodology:** Every researcher has a philosophy of research that affects procedure. Thus the student must be certain that hypotheses are well written and that appropriate criteria are used to interpret the data to reach conclusions.
8. **Reactivity:** The method of data collection should be scrutinized for reactivity. That is, a reactive technique brings about a reaction on the part of those being studied in a way that affects that data. The reactive effect is commonly labeled the “Hawthorne effect” from the study of the Hawthorne Plant of the Western Electric Company in Chicago, where it was found that worker productivity increased simply because the personnel were being observed.
9. **Unit of analysis:** In health research, the unit of analysis may be an individual (such as the health habits of a single anorexic patient) or an entire population (patterns among the hospital anorexic population). The researcher must ascertain which is most appropriate and whether resources are available to collect data.
10. **Time frame:** This is particularly important to the student because only a limited amount of time is usually available. In a cross-sectional study a particular population is involved at a single point in time; in a longitudinal time frame, data are gathered over an extended period of time (such as months or years).
11. **Budget:** To ensure that your proposal is feasible, write up a budget for expensive items. These items may include duplicating costs, travel, and postage. Some universities provide modest financial support for research projects, and you should inquire about these sources.

The student should apply all of these criteria to the potential problem to determine the feasibility of the research effort.

Sources of Problems

Now that we have developed some criteria for selecting a problem, the next step is to commence the hunt. It should be kept in mind that the problem must be researchable (i.e., it must meet the requirements of the research process characteristics outlined in Chapter 1).

At the outset, the beginner should look around at the immediate environment; it teems with **researchable problems**. Many problems in the clinic, the hospital, or the community

lend themselves to investigation. Which technique is most likely to bring about a change in smoking behavior? How does the community feel about the establishment of a wellness clinic at the hospital? Does presurgical education reduce the use of analgesics and the number of days of hospitalization?

Technological advances in medicine require continual revision in patient education, as do studies to measure their effectiveness. Similarly, in school health education, the advent of specialized curricula demands research into presentation format, teacher usage, cost benefits, and evaluation. The community health educator can turn in almost any direction to find new drugs, industrial hazards, environmental pollutants, and health fads that need investigation.

The academic experience of college juniors and seniors and of graduate students should serve as a catalyst for a research project. Textbooks, periodicals, seminar reports, and conference proceedings can inaugurate the mind into the research world. Indices and abstracts such as the *Cumulative Index to Nursing and Allied Health Literature*, *Social Sciences Index*, *Hospital Literature Index*, and *Dissertation Abstracts* provide valuable sources for research ideas. Chapter 3 discusses review of literature and offers suggestions for additional library sources.

If possible, the student should attend workshops, national and state conventions, and government-sponsored programs to gather ideas, and more importantly, to meet current researchers in the field. Closer to home, university faculty can be the impetus for health research. Although topics themselves may be provided, consultation with experienced faculty is desirable to check operability, significance, and value.

To stimulate your thinking in the direction of health research, consult the following list from which problems may be defined. It is important to realize that this is simply a list of ideas, not of properly expressed research problems.

1. AIDS education for schoolchildren
2. Patient education and reduction of health care costs
3. Competency-based education
4. Marketing of health education
5. Patient adherence to drug regimens
6. Media effectiveness in community health education
7. Evaluation of health care programs
8. Autonomy of health educators
9. Health policy and health organizations (e.g., American School Health Association, American Public Health Association, Society for Public Health Education, American Hospital Association, American Nurses Association)
10. Patient education and ethics (e.g., informed consent, confidentiality)
11. Health career objectives of students
12. Internship experiences of community health education students
13. Health promotion

14. Content areas (e.g., sex, drugs, nutrition, AIDS)
15. Health locus of control
16. Behavioral change techniques
17. Safe transportation of toxic wastes
18. Health advocacy
19. Employee assistance programs
20. Health education concerns for rural populations
21. Motivation in health-conscious individuals
22. Mental health education in a clinical setting
23. Computer-assisted instruction in health care
24. Diabetes control

Statement of the Problem and Research Questions

The **statement of the problem** offers focus and direction in the research proposal. The problem statement can be written either as a question or as a declarative statement. In either case, it must be written clearly and concisely. Each word of the statement should be definitive, indispensable, and expressive. On completion, the statement of the problem should be such that it can be read and understood by anyone without the researcher's presence.

Listed here are some examples of poorly written statements that only imply the actual problem:

- Drugs and schoolchildren
- Hypertensive drugs and patients
- The fear of toxic wastes

This indicates to the student that the researcher does not have the problem clearly in mind or at least has not expressed it completely. Needless to say, this would be an inappropriate way to commence a research report.

These three meaningless statements could be refined to show a complete statement of the problem.

- What drug is most frequently abused by those students enrolled in junior high schools in Chatham?
- What factors play a role in the low compliance rate among males and females at an East Tennessee hypertension clinic?
- What are the health fears of residents living near a proposed New Jersey toxic waste dump site?

Research questions are generally used in lieu of hypotheses (discussed below). Sometimes the use of research questions indicates that the research project is not

experimental and does not lend itself to the formulation of hypotheses. Some examples of research questions are:

- If African American and Mexican American women have greater body satisfaction and less concerns about diet than their White counterparts, what type of body presentation or image is culturally valued by each group?
- Is an individual's level of self-efficacy predictive of exercise compliance following completion of a Phase II program? (Vidmar & Rubinson, 1994)
- Will self-efficacy scores allow prediction of the likelihood of attrition for subjects of low socioeconomic status who enter an intensive program for poly-drug abuse treatment? (Steinhoff-Thornton, 1994)

It should be realized that these statements are specific as to topic and population. In other words, the parameters have been established within the statement of the problem. The ideas of the researcher must be clearly stated. Clichés, colloquialisms, slang, and professional jargon obscure thought and should be avoided when research is edited.

Subproblems

Frequently, the main problem has inherent components that, if extracted, would serve as minor, related research projects. These are called **subproblems** and as such could be investigated separately; however, the subproblems must add up to the totality of the principal problem. Further, each subproblem must be written in such a manner as to show how the data will be interpreted. Employing these two characteristics—totality and interpretation of data—the research may distinguish between subproblems and apparent subproblems.

To identify subproblems, first examine the problem statement itself for the components it contains. For example, inspect the following problem statement:

The purpose of this study is to analyze the wellness practices of Kent County Hospital nurses in contrast to the wellness practices they teach to their patients.

The next step is to demarcate the subproblem areas within the problem statement. This can be accomplished by underlining or bracketing the appropriate sections of the statement. Keep in mind that each subproblem must contain a word or words that imply data interpretation by the researcher:

The purpose of this study is [to analyze the wellness practices of Kent County Hospital nurses] [in contrast] to the [wellness practices they teach to their patients].

Now the subproblems may be written out thus:

1. What are the personal wellness practices of Kent County Hospital nurses?
2. What wellness practices are taught by Kent County Hospital nurses?
3. What will analysis of the wellness practices of these nurses indicate when contrasted with the wellness practices taught in the hospital?

It should be noted that each subproblem implies interpretation of the data and that the subproblems add up to the totality of the principal problem.

Components Comprising the Setting of the Problem

Though the problem statement offers focus and direction and the subproblems provide a means to stay on course, further delineation is necessary. It is important to indicate what limitations, delimitations, and assumptions surround the problem, as well as to define terms that may be new to the reader. Also, if the researcher is making any assumptions, they must be specified.

Limitations

Limitations are the boundaries of the problem established by factors or people other than the researcher. For example, in the preceding problem statement, the researcher may have wished to investigate five separate counties. However, permission may have been granted by only three of the five counties and subsequently the data limited to those participating counties. Other limitations could be available resources, time, number of survey forms completed, and honesty of the respondents.

Delimitations

Delimitations deal with the boundaries also, but they are set by the researcher. Though the problem statement indicates what the researcher will investigate, it is important to know what will *not* be included. In other words, the delimitations are an answer to the inquiry, *What are the precise limits of the problems?* This is particularly salient to the novice researcher, who is most likely to attempt to solve every problem imaginable. Delimitations rule out the peripheral considerations, allowing the researcher to concentrate on the central effort. The study mentioned may require the researcher to delimit the population to nurses who have a bachelor's degree, not just registered nurses (RNs). The researcher may delimit the study by geographical location, the size of the population, a central issue, or similar considerations.

Assumptions

An assumption is a condition that is taken for granted and without which the research effort would be impossible. An assumption is believed to be a fact but cannot be verified as one. In the Kent County Hospital study on wellness, the researcher may make the assumption that the teachers will answer the questionnaire honestly and thereby submit appropriate data.

Definition of Terms

Many research studies employ terms that may have special meaning to the study itself. To understand the usage of these terms, the researcher must define each term as it relates to the project at hand. Dictionary definitions are usually not adequate or helpful because they fail to provide the true meaning intended by the researcher. The meaning of "wellness practices" would have to be defined from the Kent County Hospital problem statement. It is recommended that the reader review theses and dissertations to observe the role of the section on definition of terms.

Formulation of Hypotheses

While hypotheses may be included in components that comprise the setting of the problem, they are considered separately because of their significance to the research problem. Simply put, a **hypothesis** is a logical supposition, a reasonable guess, or a suggested answer to a problem or subproblem. A hypothesis provides further direction for the research effort by setting forth a possible explanation for an occurrence. For example, when the monitor of a personal computer fails to work, the following tentative reasons may be posited:

1. The monitor is not plugged in.
2. The interface cable is not connected.
3. The monitor is not turned on.
4. The picture tube is malfunctioning.

Each of these “guesses” can be tested by checking the plug, the interface cable, the on-off switch, and the picture tube.

Students may have a difficult time deciding how to go about formulating hypothesis statements. The easiest way to think about this problem is to know about two types of approaches that are appropriate for developing hypothesis statements: inductive or deductive reasoning. When using *inductive reasoning*, a generalization is made based on relationships that have been observed. You will discern trends and patterns, and then use these as a basis for your explanation and/or the predictive nature of your hypothesis. An example of an observation in a community setting might be that those who want to stop smoking rarely attend the required number of clinic sessions necessary to complete the smoking cessation program. Can you now formulate a hypothesis based on this observation?

With the other type of approach used for stating hypotheses, called *deductive reasoning*, the researcher begins with a theoretical tenet and then makes a prediction as to how it can be applied to a specific situation. For example, you might begin by considering what you know about a theory, such as self-efficacy theory, and then make a prediction about how it will effect the behavior or the participants in a study.

Research Hypotheses

Hypotheses are derived from subproblems, and often a one-to-one correspondence is found. However, on other occasions, just one hypothesis may be developed from the problem statement itself or from a single subproblem. Generally, a hypothesis should: (1) be stated clearly and concisely; (2) express the relationship between two or more variables; and (3) be testable. Hypotheses are neither proved nor disproved. The purpose of testing a hypothesis is to ascertain the probability that it is suggested by fact. In other words, the acceptance or rejection of a hypothesis is based on fact rather than a preconceived bias.

In early stages of study, researchers state a scientific or research hypothesis as a prediction of the outcome of the test. For example, in the medical community it was predicted that as the number of cigarettes smoked increased, so would the incidence of lung cancer. This concise statement expresses the relationship between smoking and lung cancer. Of course, a linear relationship could also be expressed to state that as one variable increases, the other will decrease. The public health educator would predict that as the

usage of contraceptives increases among teenagers, the incidence of teen pregnancy would decrease. In some studies there may be a nonlinear relationship between the variables (e.g., as one variable increases, the other increases and then levels off). It might be predicted that as anxiety increases, the ability to perform increases and then plateaus. Hypothesis statements from the previous discussions would be written as follows:

1. As the number of cigarettes smoked increases, so will the incidence of lung cancer.
2. Teenagers who utilize contraceptives are less likely to become pregnant.
3. As anxiety increases, the ability to perform on an examination increases and then plateaus.

Null Hypotheses

While research hypotheses demarcate the observations to be made, it is difficult to obtain unequivocal support for them. Subsequently, they are usually rephrased into a negative or null form. This negative or no-difference format is called a *null hypothesis*, symbolized as H_0 . The null hypothesis asserts that minor differences between the variables can occur because of chance errors, and thus are not significant differences. In other words, the testing of a null hypothesis reveals either that some force or factor has resulted in a statistical difference, or that it has not resulted in such a difference. When the null hypothesis is rejected, indicating that a statistical difference does in fact exist between the variables, the competent researcher sees this as a red flag and will probe deeper into the problem to discover what has caused the difference and how. For example, a health scientist may find that a particular program alters the attitudes of those exposed to it, thereby rejecting the null hypothesis that the effect of the program would make no difference. This finding leads to another research question: *What caused the program to bring about the change, and could this factor or factors be employed in other programs?* Note that if the researcher rejects a null hypothesis, then the research hypothesis is accepted.

Returning to the problem of wellness practices of Kent County Hospital nurses at both a personal and teaching level, the null hypothesis may be written as:

There are no differences in those wellness practices personally employed or taught in the hospital.

Examination of this null hypothesis shows that it is derived from the third subproblem, and that it expresses the relationship between two variables—practices personally employed and practices taught in the hospital. It is stated concisely and is testable. If the research hypothesis—that there are differences between personal practices and what is being taught—is accepted, then the next step is to explore the dynamics underlying the differences. The research effort should not stop with rejection of the null hypothesis.

Significance and Justification of the Problem

In this section of a research proposal, the researcher has an opportunity to explain why the research effort is so important. The fledgling researcher frequently believes that the significance of the study is self-explanatory or that, because it is of personal interest, it

must be of interest to everyone. Needless to say, that is not usually the case. The relevance of the undertaking to one's peers or community, or to the patients, needs to be stated in a way that the average citizen will comprehend. Although some health researchers may be concerned solely with theory, most health scientists will demand pragmatic value from the research endeavor. Further, with so many areas of health care requiring research, there is no justification for the expenditure of efforts that fail to contribute to the profession.

The researcher must also be able to justify the study by explaining how the project will further knowledge and extend theory. In order to accomplish this, the researcher should be very familiar with and able to articulate any opposing viewpoints through a thorough review and critical analysis of the literature.

Résumé of Related Literature

Those who conduct an initial research project frequently regard the review of literature as wasted time because they believe they could be collecting data more appropriately. Skilled researchers, however, realize that the more one knows about similar research, the more likely the study can be conducted in an intelligent, comprehensible fashion.

As a general guide, keep in mind that the problem statement is central and that everything to be reviewed should serve as an aid in confronting the problem. Similar studies should be checked for population and sampling techniques; study design, including data-gathering instruments; variables measured; extraneous variables that influenced findings; recommendations for future research; and of course, the findings and conclusions. Though the related literature section of the report follows several other sections, it is important to commence a literature review early so that it can help to define the problem statement, develop components that comprise the setting of the problem, justify the study, and plan the design. Research should be conducted with deliberate speed; otherwise, the only thing accomplished is proof of the adage, "Haste makes waste."

Table 2.1 Time Schedule: Wellness Research Project

Dec. 1	Dec. 10	Feb. 5
1. Study approval	1. Obtain permission and develop working procedures with Kent County nurses	1. Develop instrument 2. Select pilot sample 3. Select study sample 4. Mail introductory letters to all participants
March 20	April 15	April 30
1. Mail instrument to nurses in study 2. Revise/complete Chapters I, II, and III	1. Questionnaires returned 2. Data keypunched	1. Fund program 2. Begin data analysis 3. Telephone contact with random sample of participants failing to return questionnaire

Proposed Research Procedures

Up to the point at which the research procedures are discussed, the report has dealt with the nature of the problem, the significance of the problem, and an explanation of what related studies have found. Now a detailed research plan must be outlined to include sampling techniques, methodological steps, instruments employed, administration of instruments, data required, and method of analyzing data.

Budget Considerations

As mentioned in this chapter, the researcher must carefully consider any financial expenditures that might occur as a result of the proposal. Along with an experienced researcher, you should review such expenditures as subject payments, duplication of materials, postage, travel, and software. If these expenditures are not able to be met by you, or with the help from university funds, you should abandon the project and discuss with your advisor a more financially viable study.

Time Schedule

Although a time schedule may not be a requirement of an advisor or funding agency, it is an invaluable device to assist in the budgeting of time and energy. Students are advised to develop a time schedule because their time is limited and academic deadlines are rarely negotiable. Moreover, dividing the research effort into operable portions with realistic dates also helps organization and reduces procrastination. Table 2.1 demonstrates how a student may develop a time schedule for the Kent County problem statement:

The purpose of this study is to analyze the wellness practices of Kent County Hospital nurses in contrast to the wellness practices they teach to their patients.

Feb. 15	March 10
<ol style="list-style-type: none"> 1. Committee review instrument 2. Mail to nurses in pilot study 	<ol style="list-style-type: none"> 1. Revise instrument from pilot study
May 15	June 15
<ol style="list-style-type: none"> 1. Write section on presentation and analysis of data 2. Write summary, conclusions, recommendations 	<ol style="list-style-type: none"> 1. Write, edit final report

Research Proposal Checklist

The research proposal is the initial step in developing the research project; and, as such, the investigator should check each area. The checklist in Table 2.2 offers a series of questions and statements that may be employed for this purpose. It is to serve only as a guide and not as an absolute formula for every research proposal.

Table 2.2 Research Proposal Checklist

A. The Problem

1. The research problem should be able to meet the following criteria:

University	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Replication	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Control	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Measurement	<input type="checkbox"/> Yes	<input type="checkbox"/> No

2. In addition, the following factors affecting problem selection must be considered and should be checked off once contemplated:

Interest	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Operability	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Scope	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Values	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Paradigm	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Methodology	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Reactivity	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Unit of analysis	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Time frame	<input type="checkbox"/> Yes	<input type="checkbox"/> No

B. Statement of the Problem

Write out the problem statement.

- | | | |
|--|------------------------------|-----------------------------|
| 1. Is the problem statement clear and concise? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 2. Does it focus on one research goal? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 3. Does the problem statement set parameters? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 4. Is the interpretation of the data implied in the problem statement? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |

C. Subproblems

Underline or box off the problem statement, and then construct and write out the subproblem(s).

- | | | |
|---|------------------------------|-----------------------------|
| 1. Is each subproblem written in question form? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 2. Is the writing clear and concise? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 3. Can each subproblem be investigated separately? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 4. Does each subproblem show that interpretation of the data will take place? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 5. Do the subproblems add up to the totality of the principal problem? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |

D. Delineations

Determine the precise boundaries of the problem, and write out each.

- 1. Are the peripheral considerations ruled out? Yes No
- 2. Is each delimitation established by the researcher? Yes No
- 3. Is the delimitation written clearly and concisely? Yes No

E. Assumptions

Consider all the assumptions necessary to conduct the study:

- 1. Is each assumption appropriate to the project? Yes No
- 2. Are you assuming too much for the study to be done? Yes No
- 3. Is each assumption really necessary to the study? Yes No

F. Definition of Terms

Ascertain which terms require special interpretation:

- 1. Is each term defined as it relates to the research project? Yes No
- 2. Are the definitions clear and concise, and do they avoid an abundance of professional jargon? Yes No

G. Hypotheses

Apply the following questions to each hypothesis:

- 1. Are the hypotheses written in an understandable fashion? Yes No
- 2. Are they derived from the problem or subproblem(s)? Yes No
- 3. Are they written in null form? Yes No
- 4. Does each hypothesis express a relationship between two or more variables? Yes No
- 5. Is the hypothesis testable? How? Yes No

H. Significance of the Problem

This section is very important to justify approval for collecting data or to obtain research funding.

- 1. List the ways in which the research project will contribute to health science.

- 2. Do others concur that this would be a worthwhile project?
 - Colleagues Yes No
 - Faculty Yes No
 - Major advisor Yes No
 - Related personnel in the field Yes No
 - Related literature Yes No
- 3. Is this section written in a manner that shows why the study should be conducted without the reader having to search for an answer? Yes No

(Continued)

Table 2.2 *Continued*

I. Related Literature		
Whether this segment be brief or lengthy, it should meet the demands listed below.		
1. Have all the resources been reviewed?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
2. Does each section relate to the problem statement?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
3. Is this segment well organized?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
4. Is the related literature current?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
J. Research Procedures		
All the steps of the research plan should be included in this segment. Where applicable, are the following included?		
1. Sample technique	<input type="checkbox"/> Yes	<input type="checkbox"/> No
2. Methodological steps	<input type="checkbox"/> Yes	<input type="checkbox"/> No
3. Instruments employed	<input type="checkbox"/> Yes	<input type="checkbox"/> No
4. Administration	<input type="checkbox"/> Yes	<input type="checkbox"/> No
5. Analysis techniques	<input type="checkbox"/> Yes	<input type="checkbox"/> No
K. Time Schedule		
This section is suggested for better organization.		
1. Is time available to complete the project?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
2. Do all segments or sections completely address the problem?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
3. Is the proposal readable, concise, and cohesive?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
4. Does the proposal represent a best effort?	<input type="checkbox"/> Yes	<input type="checkbox"/> No

SUMMARY

An organized and concise research proposal shows that the researcher has a well-developed plan and that the project is likely to be worked through to completion. This chapter explains the factors that affect problem selection and suggests sources of problems suitable to research. Each section of the proposal is overviewed to include the statement of the problem, subproblems, components comprising the setting of the problem, hypotheses, significance of the problem, résumé of related literature, research procedures, and time schedule. Review of the literature and information sources is a step in the process that assists all aspects of the research proposal.

CRITICAL THINKING QUESTIONS

1. Explain at least two of your research interests. Which one is more pragmatic?
2. Considering your answer to the previous question, develop an hypothesis for that problem. In addition, state your research problem.

3. Utilizing a time line, approximate how long it will take you to complete your study. Is this feasible? If not, alter your schedule.
4. How would you raise the necessary funds for your project? Be specific. What would you include in a letter requesting those funds?
5. Differentiate between research and null hypotheses.

SUGGESTED ACTIVITIES

1. Explain the errors in each of the following problem statements, and rewrite each to meet the demands of a good problem statement:
 - a. The purpose of this study was to examine the relationship between sexual experience and sexual and contraceptive attitudinal responses to a birth control film.
 - b. The purpose of this study was to examine which of three different approaches aimed at helping to curb smoking among teenagers enrolled in public schools was most effective: the scare approach, the fact approach, or the attitude approach.
 - c. The purpose of this study was to investigate the relationship between emotional maturity and accident involvement of male motorcycle operators in Michigan.
 - d. What health education techniques could be used to reduce anxiety in pregnant women who face a cesarean section?
2. Once you have rewritten the problem statements from the first activity, demarcate and write out the subproblems of each.
3. Rewrite each of the following research hypotheses in the form of a null hypothesis:
 - a. As patients' involvement in their education increases, so will their knowledge of their health maintenance.
 - b. As nutrition knowledge among fifth graders increases, their selection of junk food will decrease.
 - c. As more research information about AIDS is imparted to the community, there will be less anxiety within the community.
4. Read the following problem statement and then develop appropriate definitions, limitations, delimitations, and assumptions for such a study:

What are the preabortion and postabortion attitudes of women experiencing problem pregnancies toward self, contraception, intercourse, and abortion?

REFERENCES

- Bailey, K. D. (1994). *Methods of social research*. New York: Free Press.
- Champion, V. & Skinner, C. (2008). The health belief model. In K. Glanz, B. Rimer, and K. Viswanath, (Eds.), *Health Behavior and Health Education: Theory, Research and Practice* (pp. 45–63). San Francisco: Jossey-Bass.

- Green, L., & Kreuter, M. (2005). *Health promotion planning: An educational and ecological approach*. Palo Alto, CA: Mayfield.
- Steinhoff-Thornton, L. (1994). *Self-efficacy as a predictor of attrition for African-American and White client populations in treatment for poly-drug abuse*. Unpublished doctoral dissertation, University of Illinois.
- Vidmar, P., & Rubinson, L. (1994). The relationship between self-efficacy and exercise compliance in a cardiac population. *Journal of Cardiopulmonary Rehabilitation*, 19(4), 272-284.

Critical Review of the Literature and Information Sources

KEY TERMS

bibliography index
computer search
level of evidence

primary source
pyramid of evidence

relevant literature
secondary source

A review of **relevant literature** provides a framework for the hypothesis and statement of the problem. It is usually required in the beginning chapters of a thesis or dissertation. This exercise in reviewing the literature will enable the researcher to formulate ideas and concepts from previous work. We can learn what other investigators have accomplished and have failed to do so that we can make a contribution to the knowledge of health sciences. Many first-time investigators find reviewing the literature a two-sided experience. On one hand, it can be a challenging, interesting, motivating exercise. On the other, it can be tedious and painstakingly slow, especially if one allows curiosity to take over: It is easy to get sidetracked into interesting areas that are peripheral to the project at hand.

Purposes of the Review

While the general purpose of reviewing the relevant literature is to gain an understanding of previous work and to generate new ideas and concepts, the process can additionally help the investigator to:

1. Develop an understanding and grounding in theory.
2. Define the problem.
3. Review the procedures and instruments used.
4. Originate new ideas rather than repeat work already accomplished.
5. Use the recommendations for further research.
6. Critique relevant studies.

Understanding Relevant Theory

Too often investigators in the health sciences approach problems from an atheoretical perspective and, therefore, do not develop a well-defined set of hypotheses. By using the review of literature to search for relevant theoretical perspectives, investigators will gain additional knowledge and confirm hypotheses. This leads to an enrichment of the field in general and builds toward well-founded studies for the future.

Defining the Problem

While reviewing the health science literature, the investigator will be able to develop a concise plan for the study. Often we begin with lofty ideas that are sometimes not workable in the real world of research. The review will enable the investigator to state quite narrowly the relevant hypotheses and research problem.

Reviewing Procedures and Instruments

The review of the literature provides information on, and insight into, proven and unproven methodologies and procedures previously used. Knowing that some research designs are inappropriate, that some sampling frames are inadequate, and that some approaches are unreasonable enables the investigator to improve on his or her research design prior to beginning the testing. In the behavioral sciences, it is especially important to have reliable and valid instruments. The review provides insight into which measures are available and which of those will be useful in the present research study.

Originating New Ideas

Many experimenters have begun the literature search and deduced that their idea would not add any new knowledge or insight to their field of interest. A review of existing research can illuminate interest areas that need subsequent study and indicate useful applications—without reinventing the wheel. While some of us can formulate original ideas and concepts, they also can be manifested by a thorough review of the literature.

Using Recommendations for Further Research

Authors of research studies usually include very specific recommendations for additional research. This is quite helpful to the investigator because the suggestions provide the valuable insights of an experienced investigator in the same, or similar, areas of research. The list that follows might provide impetus for you to formulate your own specific area of study:

Sample Research Ideas

1. Effects of the menstrual cycle on abstinence in a quit-smoking program
2. The outcome and cost of alcohol and drug treatment in a health maintenance organization (HMO)
3. The intervention trial of a substance abuse program for women of childbearing age

4. The relationship of developmental theories to health education curricula in grades K–8
5. The effect of pregnancy on chronic hepatitis C
6. The role of states in ensuring appropriate public health practices
7. The use of telemedicine as a health education tool
8. Means of altering the attitudes of preschoolers toward family life education
9. Parental knowledge and behaviors regarding immunization of children
10. The use of self-efficacy in behavior change

Criticizing Relevant Studies

In order to adequately understand why there are contradictory results in your specific area of study, you must critically review and analyze other areas of study. Usually, contradictory results among studies arise from differing definitions of important terms, varying instruments and methodologies of science, and utilizing different data analyses. As you review and write the literature section of your proposal with a critical eye, you will certainly be challenged but will also make a very important contribution to your field of study.

Steps in the Review Process

You have decided to embark on a research study, and now the time has arrived to begin searching the literature. A tentative problem statement, centered on a theory, has been determined. Now to the library, the Internet, and beyond!

The following is an outline of the steps to consider when beginning a review of the literature:

1. Reading background information
2. Gathering the necessary tools
3. Listing key words
4. Checking preliminary sources, including databases
5. Conducting a computer search
6. Determining what to read
7. Determining the level of evidence
8. Finding shortcuts to determining the level of evidence

Reading Background Information

At this stage of your search **secondary sources** are generally used. These are usually textbooks or encyclopedias written by someone who has not directly observed the described event. A good textbook is written by an author who has searched the literature exhaustively

and compiled a text based on her or his interpretation of other experiments or events. Of course, the same author may also report on experiments he or she has participated in or witnessed. This would be considered a **primary source** because it was written by someone who has observed or participated in an event. The importance of secondary sources is that they usually have a bibliography, which provides the reader with relevant primary sources.

After reviewing the few textbooks devoted to your research problem, you realize that primary sources must be read. These include journals, final reports, or books that contain original research. In addition, government publications are good primary sources for many health topics.

Gathering the Necessary Tools

Systematically gathering data will keep you organized and will prevent you from having to redo what you have already done. Depending upon your computer access and skills, recordkeeping methods can range from notes written on index cards to a computer database. At a minimum, both approaches should contain the name of the author, the title of the reference, and a complete source listing. You should check the format (e.g., American Psychological Association) required for theses and dissertations at your institution to avoid having to recopy bibliographic entries. Additional information, if your approach allows, could be the principal findings and the level of evidence. As discussed next, determining the level of evidence requires critically appraising a source and then categorizing it by the degree to which it supports (i.e., provides evidence for) the topic at hand.

Another necessary tool is a filing system for arranging your **bibliography index**. Filing systems may be arranged by (1) the authors' names, in alphabetical order; (2) date, with the most recent work first; (3) subheading; or (4) level of evidence. Of course a combination of these techniques could also be used. For example, you may organize by level of evidence and, within each level, by author or topic. Using level of evidence allows you to discern the usefulness of the reference.

Listing Key Words

After you have the background information, which is generally gathered from secondary source materials, you will have an idea of the topic area and be able to generate key words or phrases. Use of a thesaurus has proved valuable in this process. Consult the key word listings in each computer database and contact colleagues and professors who might have related interests.

Key words and phrases are necessary because almost all health science sources are organized by subject, and you should have a list of key words to begin looking in the computer databases. As an example, your topic area may be patient education involving diabetes in outpatient settings. When you complete the general review, your first key word list might include patient education programs, outpatients, hospitals, diabetes, and nursing education. Such a list, although quite incomplete at this stage, will provide a starting point when you begin the actual search through the various computer databases.

Checking Preliminary Sources

The next step in the search process is to check the *preliminary sources*. These include reference books, indices, abstracts, guidebooks, and periodicals that help the investigator locate primary sources. Most of the sources in the health sciences are available by computer search.

General Indexing and Abstracting Services

Listed are several databases available through online computer services at most libraries. A growing number of these databases contain full text or direct you to websites with full text of the document.

- **BIOETHICSLINE** provides bibliographic citations to the literature covering the ethical, legal, and public policy issues of health care and biomedical research. Included are citations to journal articles, monographs, chapters in monographs, newspaper articles, court decisions, bills, laws, audiovisual materials, and unpublished documents derived from the literature of many fields, including the health sciences, law, religion, and philosophy.
- **Biological Abstracts** provides indexing for over 9,000 journals in all areas of life science. Subject areas represented include biology, botany, zoology, microbiology, clinical and experimental medicine, biochemistry, biophysics, instrumentation, and methods.
- **CINAHL** (Cumulative Index to Nursing and Allied Health Literature) provides comprehensive coverage of English-language nursing journals as well as journal titles from 17 allied health disciplines, books, book chapters, nursing dissertations, patient education documents, audiovisual materials, and software. OVID has evidence-based filters that can be used with this database.
- **ERIC** consists of two files: the Resources in Education (RIE) file of document citations and the Current Index to Journals in Education (CIJE) file of journal article citations from over 750 professional journals.
- **HAPI** (Health and Psychosocial Instruments) assists in the identification of measurement and evaluation instruments (e.g., questionnaires, checklists, tests) found in the health and psychosocial literature. The database does not include copies of the instruments.
- **MEDLINE** is the National Library of Medicine's (NLM) premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. OVID has evidence-based filters for MEDLINE.
- **MEDLINEplus** is the NLM's website for up-to-date, private consumer health information.
- **National Library of Medicine** includes journal citations updated weekly.
- **PsycINFO** contains summaries of the world's serial literature in psychology and related disciplines.

- **Social Work Abstracts** contains more than 35,000 records from social work and other related journals, spanning 1977 to the present, on topics such as homelessness, AIDS, child and family welfare, aging, substance abuse, legislation, and community organization. Abstracts of dissertations since 1996 are included.
- **TOXNET** is a collection of both bibliographic and factual databases focused on toxicology and the health risks posed by hazardous chemicals. Included are data related to chemical carcinogenesis, genetic toxicology, and the release of toxic chemicals in the environment (Toxic Chemical Release Inventory).

Evidence-Based Full-Text and Abstracting Services

These database services can save the health researcher an immense amount of time when seeking high-quality evidence-based information. (A full discussion of evidence-based data is provided in the section titled *Determining the Level of Evidence*.) Appendix B contains the website addresses of these databases.

- **Bandolier** is a print and Internet health care journal that uses evidence-based medicine techniques. The content is “tertiary,” which means that it distills the information from (secondary) reviews of (primary) trials and makes it comprehensible.
- **Cochrane Library** is an electronic publication available on CD-ROM and the Internet. Published by the National Health Service Centre, it is considered the premier site for evidence-based searches.
- **Database of Abstracts of Reviews of Effectiveness (DARE)** is part of the National Health Service Centre’s Cochrane Library and might be searched before MEDLINE for high-quality reviews.
- **National Guideline Clearinghouse (NGC)** is a public resource for evidence-based clinical practice guidelines. It is sponsored by the Agency for Healthcare Research and Quality (AHRQ) (formerly the Agency for Health Care Policy and Research [AHCPR]) in partnership with the American Medical Association and the American Association of Health Plans.
- **PedsCCM Evidence-Based Journal Club** is a regular publication of critical reviews of clinical trials pertinent to the practice of pediatric critical care.
- **PubMed**, developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine, is based at the National Institutes of Health (NIH). It provides access to bibliographic information drawn primarily from MEDLINE and publisher-supplied citations. It can be searched by level of evidence.

Government Documents

These listings can be accessed as government documents using the website of the Centers for Disease Control for the National Center for Health Statistics (www.cdc.gov/nchs):

- **National Center for Health Statistics (NCHS)** data systems include data on vital events as well as information on health status, lifestyle, exposure to

unhealthy influences, the onset and diagnosis of illness and disability, and the use of health care.

- **National Health and Nutrition Examination Survey (NHANES)** has been designed to collect information about the health and diet of people in the United States. It is unique in that it combines home interviews with health tests.
- **National Health Care Survey (NHCS)** provides data on alternative health care settings, such as ambulatory surgical centers, hospital outpatient departments, emergency rooms, hospices, and home health agencies.
- **National Health Interview Survey (NHIS)** provides information on health limitations, behaviors, insurance, health care access, and injuries. In addition, instrumentation that could be very important to survey researchers is available for review.
- **National Immunization Survey (NIS)** is combined with the Survey of Families with Young Children (SFYC) and the Survey of Children with Special Health Care Needs (CSHCN), which collect information on the immunization coverage and health care of children across the United States.
- **National Survey of Family Growth (NSFG)** is based on personal interviews of a national sample of women aged 15 to 44 in the noninstitutionalized civilian population. The last complete survey was conducted in 2002.
- **National Vital Statistics System (NVSS)** is responsible for the official vital statistics of the United States. State-collected information about such vital events as births, deaths, marriages, divorces, and fetal deaths is included.

Conducting a Computer Search: Finding the Evidence

After you realize just how many sources are available to you when conducting a literature search, you will want to conduct a computerized literature search. Most major university and college libraries are equipped with the hardware and software that will enable you to conduct that search. Using the computer search will enhance your ability to check the preliminary sources.

Why Conduct a Computer Search?

A **computer search** enables you explore available information by using an online literature search engine with access to several databases. Similar to other information online, these databases are usually broader and more frequently updated than books or periodicals. As when conducting any other type of online search, a literature search can be run on phrases or words that appear in the titles of written materials. It is similar to searching Google or even using Google Scholar.

Full-text electronic journals and abstracts are available using a literature search engine. Some online resources are available strictly through university libraries and public libraries. The printouts of the list of citations will include a full bibliographic entry, and this later can be converted to a bibliographic database of your own.

How to Conduct a Computer Search

In most colleges and/or universities, you will be able to conduct the search yourself. The following hints may aid you in conducting your search:

1. **Specify the research problem:** The more precisely your problem statement is written, the more beneficial your computer search will be. A generalized statement will garner far too many descriptors that will lead to too many citations. A statement such as “self-efficacy in predicting smoking behavior of junior high school students” will provide a focus for the search because the interest is in *self-efficacy*, *prediction*, *smoking behavior*, and *junior high school*. These descriptors will limit the number of citations and be precise enough to hone in on the necessary information.
2. **Select the databases:** As we discussed previously, each university or college will have the software it believes necessary to help its users. With the librarian’s help, you can decide which database or combinations of databases would be beneficial for your literature search. In the previous example, PsycINFO and SSCI would be appropriate databases.
3. **Select the descriptors:** With the advice and consultation of the librarian and the procedure provided by the database, you should select the descriptors that best describe your research problem. Return to the example used in Number 1 of this list; the descriptors might be *self-efficacy*, *smoking*, *junior high school students*, and *prediction*. Depending on how the particular database is set up, you would combine these descriptors with *or* or *and* to limit the number of citations. You will also be asked to set the language limits (“English only” is most often requested), and you may also be asked to state a year from which to begin the search.

Many journals now have an electronic format so that the entire article can be retrieved online. While this is convenient, it will save you time only if you know what you should be reading. This is discussed in the next part of this section.

4. **Conduct the search:** You will be asked to enter the descriptors that coincide with those in the databases and determine how many citations are available in each descriptor. At this time, you will probably want to have a printout of approximately 10 references to see if you have used appropriate descriptors or their combinations. Once you have decided that you are on the right track, then you can tell the computer to print anywhere from 20 to thousands (if available). Some databases also provide abstracts or full text, and you can set parameters for which abstracts you wish to have printed.
5. **Increase sensitivity and specificity:** If the search renders too many or too few references, it is important to review your descriptors and redo the search. Two terms used in trying to get the right sources and avoid getting the wrong ones are sensitivity and specificity. *Sensitivity* is the likelihood of retrieving relevant items, whereas *specificity* is the likelihood of excluding irrelevant items (Center for Evidence Based Medicine, 2008).

If you get an unmanageably large number of references, you need to increase the specificity of your search. This can be accomplished by:

- Narrowing your question
- If it is a free text search, using more specific terms

- Using a thesaurus/subject search
- Selecting specific subheadings with thesaurus/subject MeSH (medical subject) headings
- Using *and* to represent other aspects of the question
- Limiting by publication type, year, or some other delimiter

On the other hand, if you retrieve too few references, you need to increase sensitivity by:

- Broadening your question
 - Getting more search terms from relevant records
 - Trying different combinations of terms
 - Using wildcard (?) or truncation (*) features in either free text/text word or thesaurus/subject searches
 - Using *or* to add words of importance
 - Using the explosion feature of thesaurus searches
 - Selecting all subheadings with thesaurus/subject headings
 - Expanding the time frame or publication type
6. **Review the citation list:** After you have received the printout, carefully review it and select those published works that you wish to read. You will probably find additional references in the bibliographies of these citations, which may lead to another computer search.

Determining What to Read: The Information Jungle

The amount of information available today is astonishing, and it is continually growing. The National Library of Medicine database, MEDLINE, contains about 11 million references and abstracts from 4,300 journals. PubMed's retrieval engine links over 700 journals for full text of articles. Needless to say, your goal, like that of other investigators, is to spend the least amount of time finding the best information. As a general rule, useful information must have three attributes: (1) it must be relevant to the research effort; (2) it must be correct; and (3) it must require little effort to procure (Slawson, Shaughnessy, & Bennett, 1994). The formula is:

$$\text{level of evidence} = \frac{(\text{relevance} \times \text{validity})}{\text{work}}$$

Determining the Level of Evidence

Relevance

The relevance component begins with the applicability of the evidence to your problem but goes much further. The information must be critically appraised or evaluated for its validity and research usefulness. This is a crucial step if you are relying on the information to give useful guidance (Rosenberg & Donald, 1995).

Validity

Unfortunately, a large proportion of published health research lacks sufficient methodological rigor or relevance to answer research questions. To overcome this problem in medical research and practice, several investigators at McMaster University in Hamilton, Ontario, Canada developed the concept of *evidence-based medicine* (EBM), a process to systematically find, appraise, and apply research findings to clinical decisions (Oxman, Sackett, & Guyatt, 1993; Rosenberg & Donald, 1995). Specifically, it was designed to provide the best available evidence for patient care involving therapy, diagnosis, harm, and prognosis. Since its origin, it has expanded to include *evidence-based health care* (EBHC), addressing prevention/therapy, health care recommendations, outcomes in health services, decision analysis, economic analysis, and overview studies such as meta-analysis (Lohr, Eleazer, & Mauskopf, 1998). Much has been written about EBHC, and a multitude of websites exist to assist health professionals (see Appendix B).

For our purposes, it is important to understand that the health literature can be grouped into a **pyramid of evidence**. The pyramid categories range from expert opinion to double-blind, randomized, controlled studies. Like clinicians, health educators, researchers, and policy makers want to base their decisions on the best evidence available. For example, if you were considering a communitywide program on smoking prevention for young people, what is the best evidence available to demonstrate effectiveness? The principle of EBHC has been used to establish categories of evidence as well as approaches to the literature to determine the level of evidence for an article under review. The levels of evidence, beginning with the highest, are:

- I Controlled and randomized
- II-1 Controlled but not randomized
- II-2 Cohort or case control
- II-3 Multiple time series
- III Expert opinion or case study

High levels of evidence will not exist for all research or clinical questions because of the nature of the problems and research and ethical limitations.

To achieve evidence-informed decisions, the health educator or researcher should:

- Develop a focused question concerning the problem(s).
- Search secondary databases and the primary literature for relevant articles.
- Access the validity and usefulness of those articles (determine the level of evidence).
- Judge the relevance of the evidence to the problem.
- Implement the findings in the study or educational program.

The Working Group from McMaster University has published a series of “User’s Guides to the Medical Literature” (Oxman et al., 1993). As part of the series, Giacomini and Cook (2000a, b) have applied the questions to qualitative research for interpretation by the clinician. The series offers a step-by-step guide on how to interpret

the medical literature in terms of accuracy or validity. The following information is based on this work but has been modified to fit the needs of the health educator and researcher rather than the clinician.

For all articles reviewed, three basic questions must be asked:

1. Are the results of the study valid?
2. What are the results?
3. Will the results help me in conducting my study or educational endeavor?

The first question addresses the accuracy of the results. That is, are the results accurate and correct, or are they incorrect due to bias or chance? As discussed in Chapter 5, on experimental research, study design attempts to decrease bias as much as possible. Dolan (1998) ordered study designs on the basis of increasing susceptibility to bias. From least biased to most, they are: controlled, randomized trials, cohort studies, case control studies, case series, case reports, and expert opinions. Chance, of course, involves the choice of statistical tests.

The second question is addressed only if you have determined that the results are unbiased or not a result of chance. If the results are valid, you should next consider the results themselves. This includes the precision of those results.

The third question considers the relevance of the results to your presenting problem. Using our previous example of a communitywide smoking prevention program, you would want to know if the article under review really answers the question of effectiveness.

These three questions should be expanded or modified when reviewing an article on specific health issues to help you determine its level of evidence or validity. The five principal areas of concern to health educators and researchers are:

1. prevention, education, or therapy;
2. overview studies (such as meta-analysis);
3. health or educational service outcomes;
4. clinical utilization; and
5. health care recommendations.

Worksheets 1 to 5 on pages 40–44 offer questions that can be used to determine the relevance and validity of articles in each of these areas.

Relevance and Validity Worksheet 1

Articles on Prevention, Education, or Therapy

For articles on controlled studies, cohort and case-control investigations, and related research methodologies.*

Relevance (Is it worth taking the time to read this article?)

1. Does this information pertain to your central problem or question of prevention, education, or therapy?
2. If this information is true, will it change your way of approaching prevention, education, or therapy methodology and content?

Validity (Study design, flaws, and accuracy of information)

1. Are the results of the study valid?
 - 1.1. Was the assignment of subjects randomized?
 - 1.2. How were the cases and controls chosen in case-control studies?
 - 1.3. Were all subjects accounted for at the conclusion of the study?
 - 1.4. Were all subjects analyzed in the groups to which they were assigned?
 - 1.5. Was the sample size large enough to detect a meaningful difference in outcome?
 - 1.6. Were the subjects and/or investigators "blind" to the prevention, education, or therapy technique under investigation?
 - 1.7. Were the groups similar at the beginning of the study?
 - 1.8. Were the groups treated equally except for the prevention, education, or therapy treatment?
 - 1.9. In cohort and case-control studies, was exposure status clearly defined?
 - 1.10. Was the follow-up time adequate to assess the outcome of interest?
 - 1.11. What did the investigators do to control for bias?
2. What are the results?
 - 2.1. How large was the effect from the prevention, education, or therapy treatment?
 - 2.2. How precise were the authors in estimating the treatment effect?
3. Will the results help you conduct your research or carry out your educational endeavor?
 - 3.1. Can the results be applied to your research or education question?
 - 3.2. Were all the important outcomes considered?
 - 3.3. Are the likely benefits demonstrated in this article worth their potential harm, efforts, and costs?

*Dolan, 1998; Guyatt, Sackett, & Cook, 1993, 1994

Relevance and Validity Worksheet 2

Articles on Overview Studies

For articles that summarize the literature and meta-analysis articles that use quantitative methods to summarize the results.*

Relevance (Is it worth taking the time to read this article?)

1. Does the article propose to answer a specific question? (The question addressed by the summary must be very focused, or you will be forced to guess at whether it is pertinent to your investigation or endeavor.)
2. Will this information, if true, change your way of doing things (research or educational methodology, community intervention, and the like)?

Validity (Study design, flaws, and accuracy of information)

1. Are the results of the study valid?
 - 1.1. Were the methods used to locate relevant studies comprehensive and clearly stated?
 - 1.2. Were the criteria used to select articles for inclusion appropriate?
 - 1.3. What is the likelihood that important studies were missed?
 - 1.4. Did the authors appraise the validity of the included studies?
 - 1.5. Did more than one reviewer decide (a) which studies to include; (b) the validity of each study; and (c) which data to extract from the study? (Each of these is a judgment decision, and having two or more reviewers involved in the decision decreases the possibility of bias or random errors. There should be agreement among the reviewers.)
 - 1.6. Was variation between the results of the relevant studies analyzed? (This is a test of homogeneity.)
2. What are the results?
 - 2.1. What are the results of the review as they pertain to your question or problem?
 - 2.2. How precise were the results?
3. Will the results help you conduct your study or educational endeavor?
 - 3.1. Can the results be generalized to your investigation or population?
 - 3.2. Were all the research or educationally important outcomes considered?

* Oxman, Cook, & Guyatt, 1995

Relevance and Validity Worksheet 3

Articles on Health or Educational Service Outcomes

For articles on the outcomes of health or educational services, which have become principal “markers” for politicians, administrators, health care professionals, and researchers.*

Relevance (Is it worth taking the time to read this article?)

1. Does this article include or focus on the health or educational service outcomes that you are investigating?
2. What is the base perspective of the article? In other words, is it from the point of view of an administrator, health care provider, health care deliverer, politician, researcher, or educator?

Validity (Study design, flaws, and accuracy of information)

1. Are the results of the study valid?
 - 1.1. Are the outcome measures accurate and comprehensive?
 - 1.2. Were the comparison groups clearly identified and logically chosen?
 - 1.3. How similar are the comparison groups in regard to important determinants of outcome other than the one under investigation?
 - 1.4. How did the authors handle factors that could affect outcomes?
 - 1.4.1. What was the exact health or educational service provided?
 - 1.4.2. Who provided the service under question?
 - 1.4.3. Where was the service provided?
 - 1.4.4. When was the service provided?
 - 1.5. Is the difference in outcome attributable to differences in prognosis rather than intervention?
 - 1.5.1. Were all important prognostic issues measured?
 - 1.5.2. How accurate or reproducible were measures of patients’ or learners’ prognostic factors?
 - 1.5.3. Did the authors employ multivariate analysis to adjust for differences in prognostic factors?
2. What are the results?
 - 2.1. Do the results have a logical basis?
3. Will the results help you conduct your research or carry out your educational endeavor?
 - 3.1. Can the results be applied to your research or education question?

*Naylor & Guyatt, 1996

Relevance and Validity Worksheet 4

Articles on Clinical Utilization

For articles that discuss decisions related to clinical utilization.*

Relevance (Is it worth taking the time to read this article?)

1. Is the clinical utilization review directly related to the one you are studying?
2. Will the information provided, if true, support a change in clinical utilization at your institution?

Criteria Validity (Study design, flaws, and accuracy of information)

1. Are the criteria valid?
 - 1.1. Did the authors employ a sensible, detailed, and rigorous process to identify, choose, and combine evidence for the criteria?
 - 1.2. What is the quality of the evidence used in framing the criteria? Are the criteria based on evidence from controlled studies, observations, or expert opinions?
 - 1.3. If expert opinions were used, did the authors have an explicit, systematic, and reliable process for choosing panelists and collating their opinions?
 - 1.4. Did the authors address the role of values in influencing the panelists' opinions? (Not all clinicians—for example, generalists versus subspecialists—value the same things.)
 - 1.5. How well are the criteria correlated with patient outcomes? Expect criteria from controlled studies to be highly correlated. When using weaker evidence, the authors should have checked for this correlation.
2. Were the criteria appropriately applied?
 - 2.1. Were the criteria applied in a reliable, unbiased fashion? Was inter-rater reliability used?
 - 2.2. How did the authors handle the uncertainty associated with evidence and values on the criteria-based ratings of process of care? For example, when panelists disagreed, did the authors present alternative results based on harsher or more lenient raters, or did they see uncertainty as either adequate or inadequate care?
3. Can the criteria be used in your own practice or institutional setting?
 - 3.1. Do the criteria in the review really fit your setting? How do the medical culture, values, and circumstances compare? There is less to worry about if the criteria are based on strong evidence such as controlled studies.
 - 3.2. Have the criteria been field-tested for diverse settings, including one or more like your own?

*Naylor & Guyatt, 1996

Relevance and Validity Worksheet 5

Articles on Health Care Educational Recommendations*

Relevance (Is it worth taking the time to read this article?)

1. Do the recommendations directly apply to the health care or educational intervention you are considering?
2. Will implementation of the recommendations improve your health care or educational outcomes?

Validity (Study design, flaws, and accuracy of information)

1. What is the strength of the evidence? In other words, are the recommendations based on strong evidence such as randomized, controlled trials (RCTs) or more on observational studies?
 - 1.1. If there was a systematic review of RCTs, were the results consistent from study to study?
 - 1.2. Did the overview of studies reveal a differing treatment effect? If so, is it due to differences in students/patients, administration of the intervention, outcome measurement, study methodology, or chance? Check for homogeneity of the intervention effect.
 - 1.3. Is the difference between the confidence interval (CI) boundaries for the two most disparate studies greater than 5%? If so, heterogeneity may exist. However, such heterogeneity should be reviewed for both educational and clinical importance as well as statistical significance before claiming that heterogeneity had a bearing on the recommendations.
 - 1.4. Was the evidence based on cohort, case-control, or other observational studies and thus weakened?
2. How large an impact is needed to warrant use of the educational or health care intervention?
 - 2.1. Does the educational/clinical intervention have a great enough effect to warrant its financial, administrative, and perhaps student-teacher or patient-provider burdens?
 - 2.2. What is the incidence of an unwanted outcome for the group if not treated? If treated?
 - 2.3. What is the threshold number of students or patients that needs to be treated (NTT)? (A full explanation of this issue is in Guyatt et al., 1995.)
3. How well does the intervention or treatment work?
 - 3.1. If meta-analysis was used, what is the effect from pooling all the results from the various studies? What is the CI around this point estimate? Is the CI range large or small?
4. Where would you grade the recommendations using the evidence scale below?
5.
 - A1 RCTs with homogeneity and CIs all on one side of the NTT
 - A2 RCTs with homogeneity and CIs overlap the NTT
 - B1 RCTs with heterogeneity and CIs all on one side of the NTT
 - B2 RCTs with heterogeneity and CIs overlap the NTT
 - C1 Observational studies with CIs all on one side of the NTT
 - C2 Observational studies and CIs overlap the NTT

*Guyatt et al., 1995

Finding Shortcuts to Determining the Level of Evidence

The third part of the formula for determining the usefulness of information (presented in the *Determining What to Read* section) is work. While the worksheets are invaluable tools for assessing the relevance and validity of an article, they do take time to complete. In the field of evidence-based medicine or health care, some databases do allow the researcher to search by level of evidence. Databases that conduct systematic reviews are preferred because they locate, appraise, and synthesize evidence from scientific studies in order to provide informative empirical answers to scientific research questions. Systematic reviews differ from other types of reviews in that they adhere to a strict scientific design in order to be more comprehensive, to minimize the chance of bias, and so to ensure their reliability.

Several evidence-based databases are listed in the *Evidence-Based Full-Text and Abstracting Services* section of this chapter. The Cochrane Library is now the premier resource for information on the effectiveness of health care interventions. It is an electronic publication designed to supply high-quality evidence. It is published quarterly on CD-ROM and the Internet and is distributed on a subscription basis. Explicit criteria are used to include or exclude articles, and data are often combined statistically, using meta-analysis, to increase the power of numerous studies, each too small to produce reliable results individually. The complete reviews are exceedingly detailed (some as long as 25 pages) and include background material on the subject, criteria used, computer search strategies, methods of review, description of studies, methodological quality, detailed results, outcome measurements, discussion, implications for both clinicians and researchers, and an extensive reference list.

The Internet has several excellent sites that address evidence-based health care in detail. They range from simple descriptions to resources that benefit health science investigators and practitioners. Appendix B contains a section listing these websites.

Writing the Section on Related Literature

You now have gathered all, or most, of the information necessary to begin writing the review of literature section of your paper, research report, thesis, or dissertation. It is important to note here that at this point you will have already written the introduction (discussed in detail in Chapter 2), which must be related to the review of literature. In addition, you should develop a plan for the review, be sure to have the proper theoretical orientation, and summarize the entire section.

Relating the Review

As we discussed, the purpose of doing the review of literature is to develop an understanding of the background for the study; to very clearly delineate the problem; and to provide an empirical basis for the hypothesis or research questions. To present a clear and concise rationale for attempting the study, the information in the literature review should always relate to the introductory material. This will enable the introduction to flow coherently and present an organized approach to theory and research related to

your topic. The literature should be related to the purpose of the study, the generated hypotheses, and the population in question. Recall that the literature-search is done with a critical eye toward reviewing previous, similar studies and using that information to distinguish between the various kinds of problems your area of interest might encounter.

Developing a Plan

We have stressed that being organized is of paramount importance in preparing for the review and in gathering the materials. That organization will help you in writing the review, as you have already established subheadings. Subheadings are usually based on the variables and their relationship to the problem of your study. To make it easier for the reader, we suggest that each subtopic begin with an introductory sentence to explain the relevance of the section and end with a summarizing section that depicts the conclusions or insights gleaned from this subsection.

Deriving a Theoretical Orientation

Each literature review in the health sciences must have a theoretical orientation, as discussed in the beginning of this chapter. The theory is usually derived from any of the social sciences, and tends to build on the theoretical perspectives of the health-related literature. The central theme to the review is a theoretical core; from it emanates the subheadings, topics, and even subtheories. During the writing of the review, the theoretical orientation becomes part of the organizing framework, as shown in Figure 3.1.

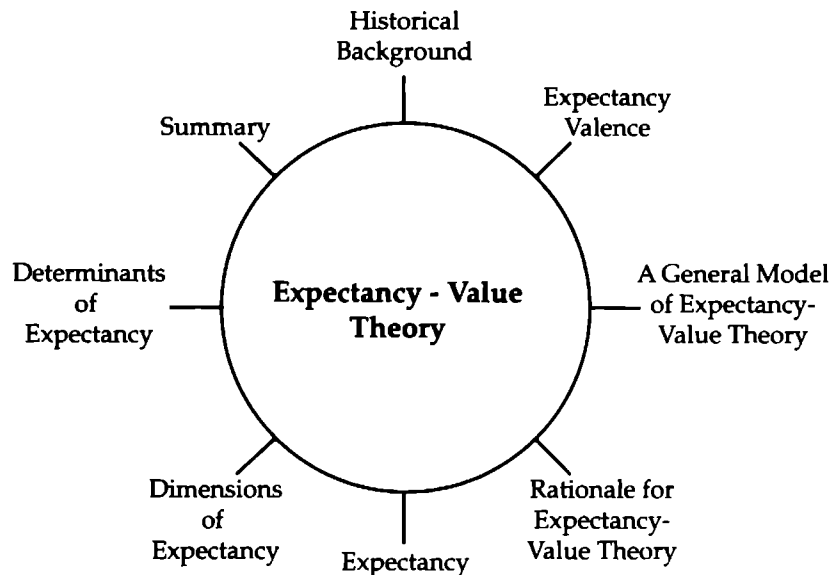


Figure 3.1 Theoretical Core and Subheadings

Summarizing

At the conclusion of the review of literature, a separate subheading titled *Summary* should be included. This section recaps the relevant information relating to theory, previous research, new insights, and the stated hypotheses. Generally, one or two paragraphs should suffice, if presented cogently.

SUMMARY

This chapter told you about the tools and information necessary for a review of literature for a paper, research report, proposal, thesis, and/or dissertation. The benefits of the review include being able to limit the problem, develop an understanding and grounding in appropriate theory, review previously used procedures and instruments, originate new ideas, use the recommendations for further research, and critically review the material. A discussion of the steps in the research process included: (1) reading the background information; (2) gathering the necessary tools; (3) listing key words; (4) checking preliminary sources; (5) conducting a computer search; (6) determining what to read; (7) determining the level of evidence; and (8) finding shortcuts to determining the level of evidence.

The list of indexing and abstracting sources presented in this chapter included evidence-based sources, which allow you to select related studies that offer varying degrees of support to your research effort. Further, those databases can be searched in an evidence-based manner. Several worksheets were provided so that you can ascertain the level of evidence of an article when such sources are not available. Finally, a strategy for actually writing the review was devised so that you can integrate the literature with the introductory materials described in Chapter 2.

CRITICAL THINKING QUESTIONS

1. Why is it important to review the literature regarding your research proposal? Outline the steps you would utilize to undertake the review.
2. Using examples, differentiate between primary and secondary sources.
3. Explain the formula used for attaining the level of evidence when conducting your review of relevant literature.
4. Discuss how you can determine the validity of an article.
5. Why is it important that your review of the literature have a theoretical orientation?

SUGGESTED ACTIVITIES

1. Go the website titled *Library of Congress* (<http://loc.gov/index.html>). Select at least two links, and read their evaluations of various search tools. Do you agree with their evaluations and believe that this site can help researchers?
2. Go to the website of the Centers for Disease Control (<http://www.cdc.gov>), and evaluate the content. Be thorough in your review. Would you recommend this site to other health researchers? Justify your answer.
3. Develop a problem statement, and list the key words you would use in a computer database search. Select at least two databases available at your university and conduct a search. Check your results for specificity and sensitivity. Using the directions in this chapter, increase or decrease your search yield as needed. Compare the results of your second search with those of your first.
4. From the list obtained in Number 3, select two articles that can be reviewed using two of the evidence-based worksheets presented in this chapter. Apply the criteria for relevance and validity. Where would you place the articles on the level of evidence chart?
5. Access the PubMed database at your university. In the search line, enter the key words *interventions* and *cervical cancer* and *sex*. What is the number of items (articles) returned? Next, click on *Limits*, immediately below the search line. In the boxes, limit the publication type to *Randomized Controlled Trials*, the language to *English*, the research to *Human Studies*, and the gender to *Female*. Click on *Search*. What is the number of items returned?

REFERENCES

- Center for Evidence-Based Medicine. (2008). Searching for the best evidence in clinical journals. Available at: <http://cebm.net>.
- Dolan, M. S. (1998). Interpretation of the literature. *Clinical Obstetrics and Gynecology*, 41(2), 307-314.
- Giacomini, M. K., & Cook, D. J. (2000a). Users' guides to the medical literature, XXIII: Qualitative research in health care. A. Are the results of the study valid? *JAMA*, 284(3), 357-362.
- Giacomini, M. K., & Cook, D. J. (2000b). Users' guides to the medical literature, XXIII: Qualitative research in health care. B. What are the results and how do they help me care for my patients? *JAMA*, 284(4), 478-482.
- Guyatt, G. H., Sackett, D. L., & Cook, D. J. (1993). User's guides to the medical literature, II: How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA*, 270(21), 2598-2601.
- Guyatt, G. H., Sackett, D. L., & Cook, D. J. (1994). User's guides to the medical literature, II: How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA*, 271(1), 59-63.

- Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., & Cook, R. J. (1995). User's guides to the medical literature, VIII: How to use clinical practice guidelines. A method for grading health care recommendations. *JAMA*, 274(22), 1800-1804.
- Iohr, K., Eleazer, K., & Mauskopf, J. (1998). Health policy issues and applications for evidence-based medicine and clinical practice guidelines. *Health Policy*, 46(1), 1-19.
- Naylor, C. D., & Guyatt, G. H. (1996). User's guides to the medical literature, XI: How to use an article about a clinical utilization review. *JAMA*, 275(18), 1435-1439.
- Oxman, A. D., Cook, D. J., & Guyatt, G. H. (1995). User's guides to the medical literature, VI: How to use an overview. *JAMA*, 272(17), 1367-1371.
- Oxman, A. D., Sackett, D. L., & Guyatt, G. H. (1993). User's guides to the medical literature, I: How to get started. *JAMA*, 270(17), 2093-2095.
- Rosenberg, W., & Donald, A. (1995). Evidence based medicine: An approach to clinical problem solving. *British Medical Journal*, 310, 1122-1126.
- Slawson, D. S., Shaughnessy, A. F., & Bennett, J. H. (1994). Becoming a medical information master: Feeling good about not knowing everything. *Journal of Family Practice*, 38(5), 505-513.

Considering Ethics in Research

KEY TERMS

Belmont Report	Nuremberg Code	unforeseeable risks
double-blind	risks versus benefits	voluntary consent
Helsinki Declaration	Title 45, Code of Federal Regulations, Part 46,	
informed consent	Protection of Human Subjects	
Institutional Review Board (IRB)		

Case Study

Emily, a nurse on Six East, was hoping that she could combine her clinic work with a research project needed for her master's degree. A large part of her job is to take complete medical and sexual histories as well as physical exams on patients referred to the colposcopy clinic. In this clinic, a colposcope is used to evaluate patients who have an abnormal Papanicolaou (Pap) smear, and a punch biopsy is taken together with an endocervical curettage. Emily follows up on many patients, explaining the results to them—normal or varying degrees of dysplasia. In cases of severe dysplasia or invasive cervical cancer, she helps the patient schedule surgery.

In collecting her information and following the patients, Emily has observed that several things in their history appear to be related to cervical cancer. Some of the more significant events appeared to be early first intercourse, multiple sexual partners, young at marriage, young or early pregnancy, smoking, and having a partner who has had multiple sexual partners. She wondered if her perceptions were accurate and if there may be other precursors to cervical cancer that she was missing.

As a research project, Emily decided to review patient records for the last six months and to collect data for another six months. At that time, she would look at all of her data and do a multiple regression analysis of those risk factors. Since this was part of her work, Emily knew this project would save herself time and would not require informed

consent. The patients would be simply following their usual course of action, and no harm could befall them. While this wasn't a perfect project, it came very close.

In speaking with her project advisor, it was pointed out that several breaches of ethics were contained in her proposal. What ethical problems do you see in her proposed research? If you were Emily's advisor, what would you suggest she do to remedy the problem?

General Ethical Dilemmas in Human Research

Research on human subjects has been conducted since the time of the ancient Greeks. However, not until the atrocities of Nazi research became known was an effort made to protect research subjects. The Nuremberg medical trials documented several charges in which prisoners were treated harshly under the name of medical research, including experiments with temperature extremes, viral studies, and studies conducted after inflicting severe wounds on research subjects. The **Nuremberg Code**, established in 1947, attempted to provide guidelines to prevent future atrocities in human research.

In the United States, even as the Nuremberg trials of 23 physicians were being conducted overseas, the U.S. Public Health Service (USPHS) supported a research project in the rural south with complete disregard for the rights of human subjects.

The study, known as the Tuskegee Study of Untreated Syphilis in the Negro Male (Brandt, 1978), commenced in 1929 in Macon County, Alabama, where Tuskegee is located. This county was found to have the highest syphilis rate in the United States, and it was believed that it merited special attention. The project was regarded as a study in nature rather than an experiment because the purpose was to follow the natural course of the disease. The researchers at the time felt that because so many Blacks had syphilis anyway, it was simply a matter of taking advantage of a natural situation. No formal protocol was written, but letters between Dr. Taliaferro Clark, Chief of the USPHS Venereal Disease Division, and his colleagues revealed that, in addition to observing the natural course of the disease, the researchers desired to show that antisypilitic treatment was unnecessary. This was speculated because many Blacks experienced a spontaneous cure and because 70% of the remainder were not inconvenienced by the disease. It was admitted that 30% of the subjects were highly contagious and seriously affected. Nevertheless, the USPHS chose not to treat the disease with arsenic and bismuth, which was recognized as a treatment at the time.

The male subjects, between the ages of 25 and 60, were not told about the true nature of the study but believed they were being treated for "bad blood," a local term that described several problems (including syphilis, anemia, and fatigue). Figure 4.1 shows some of the participants of the study. The study continued indefinitely so subjects could be watched until they died and autopsies could be performed. Incentives were used throughout the 40-year period to keep everyone participating. Moreover, the USPHS gave the U.S. Army a list of 256 men who were in the study and subsequently drafted, requesting that they not be treated for syphilis. The Army complied.

Although articles about the Tuskegee Syphilis Study appeared in the medical press as early as 1936, news about the study did not reach the national public press until 1972—when the study was still ongoing. The U.S. Department of Health, Education, and Welfare (DHEW; now called the Department of Health and Human Services,



Figure 4.1 Some participants in the Tuskegee Syphilis Study; names unknown. *Center for Disease Control. Venereal Disease Branch/National Archives.*

DHHS) formed a committee to investigate criticisms. Three basic issues arose: (1) should the study have been conducted, and should the men have been informed; (2) should the men have been treated when penicillin became available; and (3) should the study be terminated? Needless to say, many still believe that the incident was handled too casually and that a myriad of ethical issues were not addressed. Figure 4.2 depicts additional statistics about the study.

More recently, information has come forth about radiation research conducted in the 1940s and 1950s. In one investigation at Fernald State School in Waltham, Massachusetts, from 1946 to 1956, mentally disabled boys were given radioactive milk as part of a research project on the digestive system. The boys believed they were in a science club and were administered low-level radioactive forms of calcium and iron in their breakfast milk. Although consent forms were sent to parents and guardians, they failed to disclose information about the radiation. The former Atomic Energy Commission helped sponsor the research ("Retarded kids fed radiation," 1993). In a related study, women attending a free prenatal clinic at Vanderbilt University in the 1940s were given a mildly radioactive isotope to determine how iron was absorbed. This was done under the guise of a nutritional experiment. The follow-up revealed a small, yet significant, increase in cancer in the children born of these women. No documentation of informed consent has been found (Gribben, Norvell, & Van Vorst, 1994). In other experiments, prisoners in Oregon had their testicles irradiated without their consent. In September 1995, President Clinton,

DATA PRESENTED BY DR. B. C. BROWN
Classification of Cases in Tuskegee Study.

	Controls	Syphilitic	Total
Classification at initial examination	200	411	611
Cases added in 1938-1939	-	14	14
Total - Original classification	200	425	625
Controls infected during observation	-9	+9	-
Controls reclassified as syphilitic on basis of additional history	-1	+1	-
on basis of treponemal tests	-8	+8	-
Total - Final classification	182	443	625
Known dead - Number	97	276	373
Percent	53.3	62.3	59.7
Remainder -	85	167	252
Examined in 1968			
Number	36	53	89
Percent	42.4	31.7	35.3

2/4/69:az

Figure 4.2 Table from 1969 depicting number of participants in the Tuskegee Syphilis Study showing number of syphilitic patients and number of controlled non-syphilitic patients. *Center for Disease Control. Venereal Disease Branch/National Archives.*

reacting to a 925-page government report documenting the extent of experimental radiation treatments since WWII commented, "The United States of America offers a sincere apology to those of our citizens who were subjected to these experiments" (Powelson, 1995). He believed that many persons were due compensation from the government.

The beginning researcher should realize that the Tuskegee study and many of the radiation studies were carried out by supposedly forthright Americans through branches of the federal or state governments. Oftentimes, they were supported through tax dollars and generally accepted by portions of the medical community. In the case of the Tuskegee study, only with the hue and cry from the public did it stop. As a health scientist, like Emily, you should scrutinize the objectives, justification, and methodology of your study with ethical eyes. Any researcher may mean well, but failure to consider ethical dilemmas is inexcusable.

Some of the major issues are (1) justification to experiment on humans, especially children, the handicapped, the elderly, and prisoners; (2) informed consent of the subject; (3) confidentiality through the right to privacy; (4) truth-telling and deception; (5) the degree of organization that qualifies a procedure as experimental; (6) the researcher's responsibility for harmful consequences; (7) the duty to continue a successful experiment or research effort; (8) the relationship of therapeutic to nontherapeutic research; (9) sponsored research; and (10) the publication of unethical research. As to be expected, these 10 issues hold inherent ethical problems that tend to compound the research endeavor. This chapter addresses these major issues as well as the institutional review process.

Justification to Experiment on Humans

In any research effort, there must be substantial justification for the need to experiment with humans, including the implications for usage of results. The *Nuremberg Code*, mentioned previously, suggests 10 principles to be addressed in justification and methodology. Simply, they are:

1. **Voluntary consent** of the participant is absolutely essential. The subject must be capable of giving consent without coercion, and full responsibility for obtaining consent rests with the principal investigator.
2. The experiment must be designed to bring forth results that will benefit society and that cannot be obtained in any other manner.
3. Human experimentation should be based on animal research results as well as knowledge of the natural course of events, disease, or problems.
4. All unnecessary mental or physical harm should be avoided.
5. When there is reason to believe that death or disabling injury may occur, no experiment should be conducted except, perhaps, when the experimenting physicians also serve as subjects.
6. The degree of risk should never exceed the humanitarian importance of the problem to be solved.
7. All precaution should be taken to protect subjects from even remote possibilities of injury or death.
8. Only qualified personnel should be allowed to conduct experiments.
9. The subject must be able to withdraw from the experiment at any time if a point is reached that may bring about physical or mental harm.
10. The principal investigator must be ready to terminate the experiment at any stage if it appears that injury or death will result.

On the surface the Nuremberg Code appears to embrace all the necessary components. However, at least three flaws are evident. First, too much onus is given to the principal investigator, especially in regard to informed consent. Concomitantly, the overall tone is that as long as the investigator possesses positive intentions, no harm will come to the subject. It may be asked, *who knows what is good for society, and who knows how much risk is worth that good?* Finally, no one monitors the principal investigator to determine whether his or her actions and decisions are in fact ethical ones. Nevertheless, the Nuremberg Code provided a start in protecting the rights of human subjects, and it has served as a foundation for other documents, such as the **Helsinki Declaration**, which is used worldwide.

Frequently, argument about justification revolves around the interests of (1) the health sciences; (2) the subjects or patients; and (3) the community (Beecher, 1970). While it may be noted in the first issue that acquisition of knowledge and full understanding of any truths are not morally objectionable, it must be realized that not every method is allowable simply because it potentially increases knowledge and understanding. Health sciences, like other sciences, must be placed in line with other life values. When this is

done, it can be readily witnessed that the interests of the health sciences are not the highest values to which all others must be subordinated.

Regarding the interests of the subject or patient as justification, the health science researcher must be cognizant of the myriad of questions raised through consent. For example, what limits should a competent adult be allowed to take? Should the subject or the researcher set the limits? These inquiries become more complex when a researcher deals with special target groups such as children, the ill, the elderly, or prisoners.

The third issue, the interests of the community (i.e., human society, the common good) as justification, introduces more questions. Can public authority endow the researcher with the power to experiment on the individual in the interests of the community when such experimentation may transgress individual rights? Does the person exist for the community, or does community exist for the person? Keep in mind the experiments in Germany in World War II as well as those conducted in Tuskegee. Was there a well-meaning community in both instances?

In Emily's study on risk factors for cervical cancer, what would be an acceptable justification? Is the acquisition of knowledge about such risk factors adequate? Is there benefit for the patients? Is potential benefit for others appropriate justification?

Vulnerable Target Groups: Children

Justification of research on human beings is always demanding but is particularly so for those target groups who are especially vulnerable. Children are often selected for studies in the health science field because they are a captive audience in the school system as well as in pediatric wards across the country. Neither the 1947 Nuremberg Code nor the 1949 International Code of Medical Ethics mentions the use of children or other "incompetents" in nontherapeutic research. The 1964 Helsinki Declaration requires parental or guardian consent for nontherapeutic research on children. This was endorsed by the American Medical Association in 1966. Although it appears plausible on the surface, giving parents or guardians total freedom to submit children to experiments is somewhat frightening. Further, there may be direct or indirect coercion on the parents to "volunteer" their children.

In 1974 Congress mandated that the National Commission for the Protection of Human Subjects and Behavioral Research (hereinafter the Commission) establish guidelines to protect vulnerable populations, including children, from exploitation as research subjects (McCartney, 1978). In brief, the Commission's recommendations to the Secretary of the DHEW in 1977 were as follows:

1. Research involving children is important and should be conducted according to these recommendations.
2. Research may be conducted providing that the Institutional Review Board (IRB) determines that the research is scientifically sound, has been conducted on animals or adult humans first (where appropriate), has minimal risks in design and procedure, provides for privacy of children and parents, and makes selection in an equitable manner.

3. Research that does not involve greater than minimal risk to children may be conducted if the risk is justified by the anticipated benefit for the subjects; if the risk is no greater than alternative approaches; and if consent is given by parents and, when possible, by the children themselves.
4. Research that involves more than minimal risk and holds a prospect of direct subject benefit may be conducted only if such risk is justified by the anticipated results; the risk is at least as favorable to the subjects as that presented by alternative approaches; and consent is given.
5. Research that involves more than minimal risk and fails to hold out the prospect for direct benefit for individual subjects may be supported if the IRB determines that such a risk is only a minor increase over minimal risk; that generalizable knowledge about the condition will be obtained; that the anticipated knowledge is of vital importance for understanding or ameliorating the condition; and of course, that consent is given.
6. When research cannot be approved under the preceding conditions, it can only be conducted provided that it presents an opportunity to understand, prevent, or alleviate a serious problem affecting children; that a national ethical advisory board has reviewed the proposal and determined that it would not violate respect for persons or the principles of beneficence and justice; and that consent is given.
7. In addition to these recommendations, the IRB should solicit the assent of both children and parents or guardians when appropriate; involve at least one parent or guardian in the conduct of the research; and accept a child's objection as binding unless the intervention via research provides direct benefit to the health or well-being of the subject.
8. Parental consent may be waived if it is not reasonably required to protect the subjects. However, there must be an adequate alternative mechanism for protecting the children, depending upon the nature of the research protocol.
9. Children who are wards of the state should be included in research only if it is related to their status as orphans, abandoned children, and the like, or conducted in a setting wherein the majority of the children are not wards of the state. An advocate for each child must be appointed and given the same opportunity to intervene as would a parent.
10. Children who reside in institutions for the mentally disabled or correctional facilities should participate in research only if the conditions regarding research are fulfilled in addition to the aforementioned conditions.

In 1979, the Commission's report titled **The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subject Research** was released, addressing three ethical issues: justice, respect for persons, and beneficence. Since then, it has been revised with an updated version approved by DHHS in 1981 and called **Title 45, Code of Federal Regulations, Part 46, Protection of Human Subjects** (45 CFR Part 46). This policy now governs federally supported research (Cotrell & McKenzie, 2005).

The general rule of thumb set by the DHHS is that anyone below the age of 18 is considered a child. If a researcher wishes to treat subjects under 18 years of age as adults in a research project, complete rationale has to be provided, including any laws, legal precedents, agency rules and regulations, and the like. If the children are wards of the state, an advocate with appropriate background should be appointed for each child. The section in this chapter titled *Exempt and Nonexempt Review Status* provides further information about research with children (see page 68).

In regard to Emily's investigation of cervical cancer, if the patient were 16 years old, is informed consent required for either the medical procedure or the research effort? Does the age of the patient make any difference in collecting this type of data?

Vulnerable Target Groups: Elderly

The elderly population is growing rapidly in the United States with the projection that the number of older Americans will have more than doubled to 70 million by 2030. That would be one in every five people. Critical knowledge gaps exist on several fronts when it comes to the elderly: chronic diseases, economics and health care, access to health care, risk factors, nutrition, pharmaceutical interventions, healthy lifestyles, and so on. To reduce this gap much research has to be done. This was recognized by the Centers for Disease Control with the creation of the Healthy Aging Research Network.

The participation of elderly subjects in research raises the question of special protections. Currently, there are no special regulations since the aging population is seen as a heterogeneous group except under two circumstances: cognitive impairment and institutionalization. In those instances, the elderly are treated as others in similar circumstances. With increased research efforts in dementia, including Alzheimer's disease, researchers need to be particularly careful in the informed consent process (Beattie, 2007; Slaughter, et al. 2007). Fortunately most elderly are neither institutionalized or cognitively impaired. However, research challenges exist when the older person has hearing or vision problems requiring greater efforts for informed consent (Jacelon, 2007). Also, the elderly population is less likely to engage in research if it is disruptive to their routine or fails to provide any benefit to them. Nonetheless, it is important that the elderly be included in research efforts and that they not be the object of paternalism or stereotyping. Researchers cannot use age as a criterion of ability to consent.

When conducting research with the elderly, the investigator and the IRB should consider the following points (DHHS, Office of Human Research Protections, 2008):

1. Does the proposed consent process provide mechanisms for determining the adequacy of prospective subjects' comprehension and recall?
2. How will subjects' competence to consent be determined?
3. Will the research take place in an institutional setting? Has the possibility of coercion and undue influence been sufficiently minimized?

Informed Consent: Truthtelling and Deception

Informed consent essentially entails making the subject fully aware of the research project (a detailed explanation is presented later in this section) and obtaining permission from the subject to go ahead with the project. Over time, informed consent has also come to mean the written document signed by the potential research subject, although oral permission is granted in some cases. The requirement of informed consent is designed to protect the inviolability of the subject. Specifically, Capron (1974) views the functions of informed consent as (1) to promote individual autonomy; (2) to protect the patient-subject's status as a human being; (3) to avoid fraud and duress; (4) to encourage self-scrutiny by the researcher; and (5) to foster rational decision making.

As a result of a Canadian study (Pappworth, 1969) in which a University of Saskatchewan student suffered cardiac arrest and subsequent decrease in memory and concentration during an experiment to test a new anesthetic, the DHEW formulated the guidelines for informed consent. According to the DHHS, the basic requirements of a written informed consent for adults should include:

1. Fair explanation of the research effort—its purpose, expected duration of participation by the subject, and experimental procedures, including exactly what the participant will do
2. Description of any attendant discomforts and risks reasonably to be expected
3. Description of any benefits reasonably to be expected
4. Research projects involving treatment, therapy, or a service must disclose alternative procedures or courses of treatment that might be advantageous to the subject
5. Explanation of how confidentiality will be maintained
6. When more than minimal risk is anticipated, informed consent must include an explanation of compensation (if any) and a statement about whether medical treatments are available in case of injury. (Note that “minimal risk” means a risk that is not greater than the risk encountered in daily life or during the performance of routine physical or psychological examinations.)
7. An offer to answer any inquiries concerning the procedures or whom to contact if problems should arise
8. Instruction that the person is free to withdraw consent and to discontinue participation in the project or activity at any time without prejudice to the subject

Some types of research in the health sciences require additional elements of consent. These elements are:

1. **Unforeseeable risks:** If any exist, these need to be pointed out to the potential research subject. For example, pharmaceutical experiments may have unforeseeable risks to a fetus that must be explained to a subject who is, or is likely to become, pregnant.
2. **Additional costs:** If the subject should require medical assistance, therapy, or some other service, who will pay for it and exactly how much it will cost must be stated.

3. **Investigator termination:** The circumstances under which the investigator would terminate the subject's participation without the subject's consent must be explained.
4. **Termination procedures:** The subject must be told how to withdraw from the research project and what consequences, if any, may result.
5. **New findings:** If new findings are likely during the course of the research project, subjects should be informed of those findings that relate to their willingness to continue participation.
6. **Number of subjects:** When appropriate, subjects should know the number of participants involved in the whole study.

When research involves children (under 18 years of age), written parental permission should be obtained, unless exemption is given by the IRB. Beginning in about junior high or middle school, the child's written consent is needed in addition to parental consent. Children in lower grades should provide oral consent—a positive statement of willingness to participate—in addition to parental consent. Age-appropriate explanation should be given to preschool children in addition to obtaining parental consent. In all cases, but especially for preschoolers, if children experience undue stress, their participation should be discontinued.

While the guidelines provide a basis for informed consent, the researcher still faces several research and ethical dilemmas. For example, how much does the subject need to know before consent can be given? If too much detail is given, the subject may not understand or perhaps may skew the data by acting the way the researcher hopes. In some instances the researcher may not be aware of the potential discomforts that could occur even a year after the research.

Should Emily have informed consent in her study? What are the pros and cons of informed consent in her research effort? If informed consent is to be present in her study, how should she go about it? Would informed consent be different if her patient were 15 years old? 18 years old? What forms and procedures are required by your university or IRB to determine that informed consent has been fairly applied in studies dealing with human subjects?

Informed Consent and Double-Blind Studies

The design of double-blind studies is simple and logical. One-half of the subjects are randomly selected to receive the experimental product while the other half is given a placebo, and the results are compared. Neither the researcher nor the subjects know who obtains the active substance or the placebo; hence the term **double-blind**.

Although the researcher values this methodology to earn accurate results, it possesses many ethical dilemmas. A study conducted by Goldzieher of the Southwest Foundation for Research and Education displays many of the ethical problems encompassed in double-blind research (Veatch, 1977). The purpose of the experiment was to discover whether some of the reported side effects of the contraceptive pill were physiological or psychological. The subjects were primarily poor, multiparous, Mexican American women who had come to a San Antonio clinic for contraception to prevent further pregnancies. Seventy-six of the women were given placebos, and another group

got various hormone contraceptives. None were told that they were involved in a research project or that they were receiving placebos. All were instructed to employ vaginal cream because the contraceptive pill might not be “completely effective.”

The results of the experiment showed that the women taking placebos had many of the same side effects—depression, breast tenderness, and headaches—as those on the contraceptive pill. However, 13% (10) of the 76 women taking placebos became pregnant. Needless to say, these women were deceived; yet, a request for full informed consent would have made the study impossible, because the women came to the clinic for pregnancy prevention. Could some information have been given without ruining the research? If deception is part of the research process, should the experiment be cancelled? In other words, what justification is required to approve deception? Were the results of the Goldzicher study worth 10 women getting pregnant? Who is to decide whether the results of an experimental procedure are potentially worthwhile? Moreover, if such studies are even contemplated, how should the population for the study be chosen? In this case, why were poor, Mexican American women selected, particularly individuals who could not afford medical care as clinic patients?

In short, the double-blind methodology is excellent for some research objectives, particularly if a placebo is employed; however, it is fraught with ethical dilemmas that should be addressed by the principal investigator. Of course, similar concerns apply to single-blind run-in phase clinical trials, too (Mann, 2007).

Overall, recent trends indicate that the requirements for informed consent are becoming more and more rigorous. There are many special circumstances regarding persons who speak foreign languages and other minority groups. It is believed that a long consent form attached to a mail survey will reduce the number of returns. Just reading the form will take more time and thereby reduce returns, and the usual ominous tone is likely to decrease responses even further. Singer (1978) has shown that informed consent procedures lower response rates for interviews. Thorne (1980), in speaking about sociological research, has complained that federal regulations are based on the biomedical model of research and as such are not workable in research that is observational field research. He states that

... the requirement that one obtain signed consent forms from everyone one studies may violate anonymity and actually increase risks for some groups of subjects. In the end the procedures may result in meaningless ritual rather than improving the ethics of field research. (p. 285)

The problem of informed consent for the researcher is difficult, with the major problem being that of application—how much information, how much consent.

Right to Privacy and Confidentiality

All participants in human research have the right to privacy in that they have the right to request that their individual identities remain concealed. While the charge of invasion of privacy can be made in all methodologies, it is most likely to occur with survey research, audiotaping, and videotaping. The question of what constitutes invasion of privacy is quite subjective and may imply something different to the subject than the intent of the

researcher. For example, in a drug survey the participant may have used cocaine but felt very guilty about it and may have not admitted it to anyone. Such a person may understandably feel that his or her privacy is being invaded when asked if cocaine had been taken at any time. This feeling of “invasion” may be underscored if the interview was either audiotaped or videotaped.

To ensure anonymity it is important to explain to the subject that most researchers are interested in group data and that individual scores are compiled with others. Further, individuals are identified by number rather than by name. Perhaps most importantly, the subject needs to understand the importance of the data being gathered; if the project is deemed important enough, the subject may be willing to sacrifice some privacy.

The principle of confidentiality is related to the right to privacy. Who will be able to see the data? In school systems, both teachers and students are concerned that research data may be used to evaluate performance. The health science researcher should treat all data confidentially and ensure that (1) all data is returned anonymously and directly to the research office; (2) all data is rostered by number; and (3) unneeded material is destroyed upon completion of the project.

Of course, there are several other ways in which confidentiality could be broken. It may occur if a subject is a relative or friend of a member of the research team who has even limited access to the data. Research records can be stolen. A questionnaire may be found by a spouse, friend, parent, or colleague, or a telephone message could be taken that identifies the participant. Even the participant may break confidentiality by writing a name on the questionnaire that reveals the illicit use of drugs. These breaks may occur no matter how careful the investigator.

Review the case study at the beginning of the chapter. What has Emily done to ensure the right to privacy and to maintain confidentiality? If her study were conducted as presented, might she violate the right to privacy? What suggestions could be given to Emily?

Audio and Video Recordings

Audio and video recordings of subjects precludes anonymity. This does not mean that they should not be employed in research; however, it does mean that:

1. Complete justification of use must be given by the researcher to the IRB.
2. If confidentiality is threatened more by use of recordings than nonuse, the researcher must have more reason than simple convenience.
3. The researcher must be very specific about how the recordings will be used and information analyzed; who will have access to them; where and for how long they will be stored; and the method of disposal.
4. If the recordings are to be used for purposes other than the research (e.g., an oral history interview used for teaching), complete explanation must be given, and the informed consent must reflect this additional use, including the length of time (e.g., 2 years) that it will be used. Note that a separate release form should be given to subjects regarding the use of the recordings outside of the research

project. This separate form gives the rights of ownership to another individual, group, or institution.

5. If a panel of judges or the like are to review the recordings, the IRB should be informed as to their names (if available), credentials, and functions in regard to the research and recordings. The subjects need to receive this same information.
6. Subjects should be permitted to listen or view their recording upon completion to affirm their permission for use of the material. In the case of children (or other vulnerable groups), parents should be permitted to review the recording. They may not want a video of their child's antisocial behavior or an audio of family "secrets."
7. In some instances, consent to participate may be given on the recording. This is especially true when the investigator uses the subject's voice without name or identification. Generally, this approach requires approval from the IRB because it is a departure from the written form.

Responsibility for Harmful Consequences

Subjects have a right to expect the health science researcher to prevent harm from befalling them and to be sensitive to their need for human dignity. Further, as discussed under the *Informed Consent* section, DHHS guidelines require that each subject be told of any "attendant discomforts and risks reasonably to be expected." Once again, however, how much need be explained so as not to frighten the subject? What one subject finds disquieting, another might not. The researcher must find a common ground for explanation and answer all inquires honestly.

Dava Sobel (1981), a science reporter for the *New York Times*, described her experience as a subject in an experiment at Montefiore Hospital in the Bronx. The purpose of the study was to observe how certain bodily functions change in the absence of timing devices—clocks, calendars, natural light cues, and social regimentation. As the 17th volunteer and the first female, she was placed in a special environment (room) and allowed to set her own schedule according to the dictates of her body for 25 days. Her account of informed consent from the subject's viewpoint is quite noteworthy but perhaps more so is her view of harm to the subject:

Within hours, I realized that I had not quite understood what subjecthood entailed. First came the insertion of the catheter which was extremely painful. The procedure had to be done in both arms, since the doctor mistook the lack of blood flow from the right side for a defective or improperly inserted needle. [Blood samples were part of the experiment.]

. . . The frequency of those samples was my second shock. . . I feel I should have been warned that the samples would be taken "very" frequently [every 20 minutes], interfering constantly with my work, my meals, and the time I expected to be alone in the bathroom. . . .

Feeling tense, I wrote a letter to my husband date "Day 1," stamped it, and gave it to the white-coated technician who came in for a blood sample. He waved it almost

tauntingly and said, "This will go out, but I won't say *when* it will go out. Maybe in a few days." I panicked. (Sobel, 1981, pp. 5–6)

There is no doubt that research of this nature will cause some discomfort, but how much? From Sobel's vantage point, there was too much.

To complicate matters further, she was given very little help in reorienting herself to the "real" world. After 25 days in which she would sleep and eat at any desired interval, Sobel found her normal patterns so disrupted that she "might as well have been living inside a stranger" (p. 7). It took her approximately 2 weeks to readjust to her former schedule, and she missed work most of that time. Should it be the responsibility of the researcher to reimburse a subject who misses that much work? Overall, her account offers the researcher insight into harms, or at least risks, that the subject should be told of. As a point of note, these researchers did change their protocol somewhat.

Kolodny (1977) cautions of potential long-term consequences of study participation. In sexual research, a subject who is observed in some type of sexual activity may experience no problems while the study is underway but may discover feelings of guilt years later. Perhaps, a potential spouse may refuse marriage when he or she learns of the participation.

In the case of Emily, could harmful consequences occur to any of her subjects? What would be the reaction of the patients if they discovered their medical and sexual histories were being used as part of a research project? What would happen to the trust level among the patients going to the colposcopy clinic? Would future research efforts be placed in jeopardy? Could Emily's proposed research have negative connotations for nursing?

Though the health science researcher cannot hope to predict all risks and consequences, he or she should make an effort to communicate all known ones.

The Duty to Continue a Successful Research Effort

Researchers in the health science discipline frequently embark upon research efforts to improve the health of people suffering chronic conditions such as obesity, smoking, hypertension, and others. The underlying question is whether contemporary health scientists have an obligation to continue successful programs.

For example, if Emily successfully demonstrates that five variables serve as risk factors for cervical cancer, is she obligated to plan, implement, and evaluate clinic educational programs warning women of the dangers? What if she found that only two variables contributed? Who would be responsible for the costs of the program?

In another example, if a health scientist finds that a stress reduction program does in fact lower physiological tension, improves work satisfaction, and augments productivity when compared to a control group, should the program be offered to all workers? What if only two of the three variables are positive? Who would be responsible for continuing costs? The obligation of the researcher to the subjects subsequent to the research effort are debatable but should be addressed before the project commences.

Therapeutic and Nontherapeutic Research

If the research objective is to acquire information, should the justification be different than if the objective were to develop a cure or behavior pattern to improve health? In other words, is a different justification required for nontherapeutic research than for therapeutic research? In nontherapeutic research there may be no apparent benefit for the human subjects, while in therapeutic research at least the experimental group may benefit. Of course, in therapeutic research the question arises as to the right of the participant to request placement in the experimental group so that potential benefits may be obtained. While both types of research have ethical obligations, clinical research efforts may have a greater responsibility to be certain the subjects understand the difference between research and clinical care (Horng & Grady, 2003) and to share research results (Fernandez, Kodish, & Weijer, 2003).

Is the nature of Emily's proposed study more nontherapeutic or therapeutic? If she is only attempting to gain knowledge about risk factors for cervical cancer, should her justification be different than if she were in fact planning a therapeutic study? Should health scientists who research in the behavioral field be held accountable for subjects' health behavior? All in all, therapeutic and nontherapeutic research efforts differ in objectives, but differences in justification are a moot point.

Sponsored Research

More often than not health science researchers cannot afford to pay for a project out of pocket. The myriad of expenses—drawing a large sample; training interviewers; paying for postage, offices, and overhead; and computer time—must be met by a sponsor. Usually a government agency such as the U.S. Department of Public Health or a charitable foundation may provide funds and allow the researcher to conduct the project with no strings attached. However, previous examples throughout this chapter have illustrated how some agencies attempt to inflict their views on the project.

Bailey (1994) presents three major areas in which ethical conflict arises between the sponsor and the researcher. First, the sponsor may tell the researcher how to conduct the study or what findings are to be expected. Second, the sponsor may request suppression of findings, which could range from total falsification to manipulation of statistics. Third, the actual sponsor may be concealed or the true purpose of the study hidden. The latter has occurred in several research efforts for which the Central Intelligence Agency served as sponsor (Sjoberg, 1959).

The source of sponsorship in Emily's study is very subtle in that it is her employer. Would it make any difference if a pharmaceutical company supported her efforts by paying for such things as data analysis, computer time, and the like? How much support is too much?

In writing grants or procuring funding from outside sources, the health science researcher must be cognizant of these potential ethical breaks. The researcher must be prepared for compromise in some instances but, in such cases, should examine all ethical questions so as not to end the project with results that have been questionably attained.

Publication of Unethical Research

A research project does not become ethical because it produces valuable data; it is ethical or unethical from its inception. Subsequently, researchers, editors, and editorial boards must look beyond the results of an investigation into all the ethical aspects involved in research. Only then can a fair and just decision be made about possible publication.

One option that may be chosen is the decision not to publish. Ingelfinger (1978), former editor of the *New England Journal of Medicine*, stated that “reports of investigations performed unethically are not accepted for publication” (p. 791). His belief is that researchers should not be involved in unethical acts, directly or indirectly. Subjects are not to be used as a means to an end. Moreover, failure to publish such research should serve as a caveat to other researchers. While some very worthwhile information will be lost to the profession, proponents of this position feel that more good consequences than bad will occur in the long run.

Beecher (1970) suggested a modification of this view by stating that “such material *ordinarily* should not be published” (p. 31). If unethical circumstances exist, Beecher believes that the researcher should report, in the text, where the dilemmas existed. In a parallel position, Levine (1973), former editor of *Clinical Research* and professor of medicine at Yale University, advocates that “manuscripts describing research conducted unethically but which satisfy the usual scientific criteria for acceptability should be published along with editorials on which the ethical deficiencies are exposed and criticized” (p. 763). This plan raises the ethical issues to a level of debate and still allows the results to be received by other professionals. Nevertheless it raises questions, too. Does such publication indicate that the editor or journal approves of such actions or at best frowns upon them? Will it function as a deterrent to unethical research?

In essence, no matter what position is taken, the majority of the responsibility rests with the researcher. As a professional, the health scientist must act in a fashion that is conducive to subject protection and growth of the professional. Publication is both a responsibility and a privilege.

Research or Just a Look-See?

A major question is, when does research become research or an experiment an experiment? Most researchers have “pet” ideas or theories they would like to test a “little bit” before taking on a full-blown study. However, when human subjects are involved, is it fair to include them in such tryouts without informed consent? What if data are collected when the health scientist is functioning as a clinical person and the data are later incorporated into research? For example, if a master’s candidate serves as a counselor in local clinic for patients seeking an abortion and then records the anxiety and psychological trauma experienced by each patient for his or her own growth as a counselor, is that ethical? If, at a later date, the same candidate wishes to incorporate the patients’ reactions into a master’s thesis, is that ethical? The patients were not informed about the use of the data because at the time a research project was not planned. Should consent be obtained even at this late date?

“The borderline between being a human being with whom we work, play, and exchange information and being a human subject of research is not a line at all. It is a misty frontier” (Committee on Research Participation, 1995). These statements reflect the difficulty in identifying boundaries on some occasions. If you were Emily’s research advisor, would you consider her proposed effort as research or a “look-see”? What criteria did you use to arrive at your answer?

From a broad perspective, whenever a person plans a systematic inquiry to gain generalizable knowledge, the effort must be considered research. To clarify this statement, if you can answer “yes” to any of the following questions, you are doing research.

1. Are you planning to procure subjects? (This would be in contrast to people seeking you out for normal, professional services.)
2. Will the data collected be analyzed, interpreted, and disseminated?
3. Do you think the knowledge you will gain can be generalized to similar situations or perhaps lead to new processes or procedures?

In contrast, if the data gathered are to be used in the classroom only, for administrative purposes alone, or for a contractor’s project in which there is no dissemination of data, then the effort might be considered nonresearch. Once you determine that your project is indeed research involving human subjects, the next step is to decide whether your project is exempt from review by the full IRB.

Role of the Institutional Review Board

Two federal agencies, the DHHS and the Food and Drug Administration (FDA), have complete sets of regulations about the review of research involving human subjects. To ensure compliance with the regulations, the government requires each institution conducting research to establish an **Institutional Review Board (IRB)**. In addition, investigators using disease registries usually have to obtain permission from private physicians before contacting patients at home. Researchers using deceased cases generally have to get approval from the state health department and the vital statistics bureau to procure death certificates. As a side note, regulations exist for animal research and research with recombinant DNA, too.

Although IRBs may vary slightly from institution to institution, some functions common to all are:

- To review all research efforts involving human subjects except those projects exempted from review according to regulations
- To develop policies for research with human subjects
- To provide education to investigators and departments regarding policies, procedures, and related issues
- To maintain records in accordance with federal regulations. These are usually kept for a minimum of 3 years after the termination of a project.

The review process may vary slightly from institution to institution. Box 4.1 highlights the general review process for graduate work. In reviewing the submission, the IRB employs the following criteria:

1. **Risks:** These must be kept minimal by design.
2. **Risks versus benefits:** Any risks must be reasonable in relation to potential benefits received and to the potential knowledge gained via the research. Keep in mind that only those risks and benefits involved in research are evaluated, not those obtained from services or therapies subjects would receive even if they were not in the research.
3. **Subject selection:** Selection methodology must be considered in light of research purposes and settings as well as the nature of the population from which the sample is to be drawn. As to be expected, children, pregnant women, the mentally disabled, prisoners, and other vulnerable groups are given special attention.
4. **Informed consent:** Complete documentation must be given to ensure that each participant (or the legal representative) has been sought out. The IRB has the right to observe (or have observed) the consent process.

Box 4.1 THE REVIEW PROCESS

Be certain to review the guidelines particular to your own institution. Generally the review process includes the following steps:

1. **Research investigator:** Completion of the proposal and forms for IRB review is the responsibility of the investigator. The investigator must monitor the progress of the proposal and forms as they go through each level of review.
2. **Graduate students:** Thesis and doctoral committees usually have to approve the research project before the appropriate forms are forwarded for IRB review at the departmental level.
3. **Departmental review:** Although a department head holds the responsibility for signing off on IRB forms, an advisory departmental review committee may be in place. Students should inquire as to how long this process takes.
4. **University review:** If the research project has been approved at the departmental level, it is then forwarded to the university (or institution) for review. If the project is deemed to be "exempt from full review," then the full committee does not have to review it. This takes much less time than a full review. If a full review is required, the investigator should inquire as to the next meeting date and the date for submission of materials. Depending upon the institution, this could easily be a 1-month wait.
5. **Schedule:** It is important to plan time for the review process when designing your study. Upon review, the IRB may approve, disapprove, request minor modifications, or request an external review.

5. *Safety and privacy*: The proposal should show how the data will be monitored during collection to ensure subjects' safety and to maintain both privacy and confidentiality of the data.
6. *Additional considerations*: These are tailored to the specific research project and can include standards of professional conduct, local laws and regulations, the mission of the institution, and so forth.

After its review, the IRB may approve, disapprove, request minor modifications, or request an external review. Accepted submissions require an annual review. The intent of the annual review is to determine whether changes requested in the past review have taken place. The IRB has the right to terminate or suspend the research based upon the annual review.

Exempt and Nonexempt Review Status

When preparing a research proposal, you should get information from your IRB or committee on research participation (CRP). Following their guidelines will increase your chances for approval. In reading the guidelines, you will find that some studies require review by the entire IRB, while other studies are exempt from full review. "Exempt from review" means that you must notify your IRB and complete forms to exempt your project from a full IRB review. It does not mean you can ignore the IRB and simply do what you want. Five areas of research that are often, but not always, exempt from review are:

1. Research about normal educational practices—instruction strategies, effectiveness, classroom management, curricula
2. Educational tests with information recorded in a way to protect the identity of the children
3. Observation of public behavior as long as subjects cannot be identified and the researcher is not involved in the activities being observed
4. Use of existing data (documents, pathological specimens, diagnostic specimens) if they are publicly available or if the information is used in such a way that the subjects cannot be identified
5. Surveys and interviews wherein the data are recorded so that subjects cannot be identified, directly or indirectly. If there is any likelihood that the subject's responses could place him or her at risk of criminal or civil liability or be detrimental to his or her employability or financial standing, a full review is necessary

Educational, especially classroom, research often raises questions as to the need for informed consent and review as noted by DuBois (2002), and the answers are often confounding. The Indian Health Service (IHS) government website (<http://www.ihs.gov/MedicalPrograms/Research/docs/OHRPdecisionFlowWcitations.pdf>) provides a decision-making chart that guides researchers through the thought process of the IRB. Although

surveys and interviews are often exempt for adults, they need to be reviewed when used with children.

When a review by the IRB is required, two routes are available. One route is an expedited review, which is only available to projects involving no more than minimal risk to subjects. If it is determined that greater than minimal risk is present, then a full committee review is required. Box 4.2 illustrates an example checklist for completing forms for IRB review.

Box 4.2 EXAMPLE CHECKLIST FOR IRB REVIEW FORMS

Project director(s): Full name, mailing address, phone number, department or affiliation

Project: Title, external funding agency and identification number, grant submission deadline, starting date, estimated completion date

Exempt research: Category of exempt research (usually on the form)

Objectives of project: This is as they apply to the exact procedures involving human subjects.

Subjects: Description of subjects, criteria for inclusion and exclusion, population from which sample will be taken, duration of subject participation, special components if vulnerable populations are being used, age range of children (if applicable), source and selection of the control group (if any)

Methods or procedures: Research methods involving human subjects, voluntary participation, explanation of no penalty for refusal or early withdrawal

Specific risks and protective measures: List of potential risks to subjects; list of protective measures; explanation of risks (stresses, drugs, experimental manipulations, photos, recordings, and so on) and, if none, an explanation of why; full explanation of

anonymity of identities and confidentiality of data; location of where materials having subjects' names will be stored and who (by name) will have access to them; how materials will be destroyed at the appropriate time

Risk versus benefits: Reasonableness of risks compared to benefits for current subjects, future populations, and for knowledge gained. If risks are minimal, be certain to indicate

Informed consent: Detailed explanation of your method to obtain informed consent, written consent form to be signed by subject (when necessary)

Investigator qualifications: Education and special training (if appropriate), past research efforts of a similar nature regarding human subjects, training of personnel

Adequacy of facilities: In some instances, information about the facility which will support the research must be given. This is especially true when children or other vulnerable groups are involved.

Signatures: Principal investigator, co-investigators, advisors (for students), department head and chair of departmental review committee

Case Discussion

This section highlights some of the issues raised throughout the chapter regarding Emily's research as discussed at the opening of this chapter. The following points apply:

1. Emily's proposed project constitutes "research." She will collect data to gain generalizable knowledge and will be using human subjects.
2. Using her work setting as an opportunity for research is feasible; however, she must be extra careful in her planning, especially with regard to informed consent.
3. A review of literature supports her selection of variables: range of dysplasia, early first intercourse, multiple sexual partners, young age at marriage, young or early pregnancy, smoking, and a partner who has had multiple sexual partners. These variables can all be continuous in nature, allowing for multiple regression.
4. The results of the study may be of great benefit in reducing severe dysplasia through education.
5. If Emily decides to engage patients 18 years of age or younger, she would need special consent from a parent, guardian, or other person who has the authority to give permission.
6. Informed consent is a must for her research for all the reasons outlined in this chapter.
7. Information gained combines behavioral and medical data. As such, confidentiality is of the utmost importance, and Emily should be able to demonstrate appropriate storage of data as well as destruction of such when her research is completed.
8. Emily may have to go through two IRBs—university and hospital—for her project.

SUMMARY

This chapter offered viewpoints on ethical dilemmas that confront the research effort. It was seen that unethical research has been conducted in the past with and without approval of recognized governmental bodies. One of the initial issues to be reviewed by any investigator is the justification to experiment on human subjects rather than on animals or by computer simulation. Frequently, justification revolves around (1) the interests of health sciences; (2) the interests of the subjects or patients; and (3) the interests of the community. All can be questioned.

Justification for research on vulnerable target groups, particularly children, was seen to be more complex. DHHS guidelines were reviewed because they are employed by IRBs.

Another major issue was informed consent, involving truth-telling and deception. This area is of particular importance to both the researcher and the subject and is well scrutinized by professionals. Double-blind methodology presents unique problems to the informed consent issue.

Privacy and confidentiality are two issues highly regarded by subjects. Concomitantly, fear of harm, albeit subjective, is a point of concern to subjects and researchers alike.

Another issue was the degree of organization required to designate an experiment as an experiment. When an experiment or research effort, such as the reduction of smoking, has been proven to be successful, the question of whether it should be continued was discussed. This is particularly so if the nature of the study is therapeutic rather than nontherapeutic.

Sponsorship and publication of results were also seen to be fraught with ethical decisions. Overall, the major responsibility rests with the researcher.

Obtaining permission for research with human subjects requires approval from an IRB. While review boards may vary somewhat from institution to institution, general guidelines were presented.

CRITICAL THINKING QUESTIONS

1. Paco is planning his informed consent form to use in his study of patients who come to the Family Medicine Clinic in his community. His thesis deals with adults who have diabetes mellitus, and he will be randomly assigning them to either an experimental education group or a control group. His list of points to include in the consent form is below. What, if anything, is he missing from his list?
 - a. Fair explanation of the research effort
 - b. Description of any discomforts and risks
 - c. Explanation of confidentiality of test scores and attitude scales
 - d. His name and phone number for subjects to call if questions or concerns arise
 - e. Instructions on how to withdraw from the study and a notation that no penalty will be incurred from leaving the study before it is finished
2. Sarah, an instructor at the local community college, teaches three classes in health education. She is very curious as to the knowledge and attitudes of her students regarding HIV/AIDS. The only way to obtain this information, she believes, is by using a questionnaire—one that has both a cognitive test and an attitude scale. She plans to use the information for planning her lessons on the same subject. If you were her department chair, would you classify her plans as research? Justify your answer.
3. Jo, a nurse practitioner in obstetrics, proposes to research the efficacy of fetal monitoring. Her research would require the use of a fetal scalp monitor during delivery. Having done this hundreds of times in the labor and delivery section of the hospital, she knew it was not too intrusive a procedure. Therefore, she requested exempt status from the IRB. If you were a member of the IRB for her institution, would you grant it? Explain your answer.
4. What ethical considerations should a professor make when distributing a research questionnaire to students in his or her class?
5. What role does The Belmont Report play in conducting research on humans?

SUGGESTED ACTIVITIES

1. Go to your IRB office and get the forms for both exempt and nonexempt status. Compare the two forms. Also, get information about an expedited review versus a full review. Apply this information to research you are planning.
2. Search the Internet for five university websites to ascertain where the IRB office is located in each university. Download a current IRB form and compare it with your university's form.
3. Go to this website and apply your research project to the algorithm.
<http://www.ihp.gov/MedicalPrograms/Research/docs/OHRPdecisionFlowWcitations.pdf>
4. Make a checklist of ethical issues that apply to your intended research project.

REFERENCES

- Bailey, K. D. (1994). *Methods of social research* (2nd ed.). New York: Free Press.
- Beattie, B. L. (2007). Consent in Alzheimer's disease research: Risk / benefit factors. *Canadian Journal of Neurological Science*, 34 (Suppl. 1), S27-S31.
- Beecher, H. K. (1970). *Research and the individual: Human studies*. Boston: Little, Brown.
- Brant, A. (1978). Racism and research: The case of the Tuskegee syphilis study. *The Hastings Center Report*, 8(6), 21-29.
- Capron, A. (1974). Informed consent in catastrophic disease research and treatment. *University of Pennsylvania Law Review*, 123(20), 364-376.
- Committee on Research Participation. (1995). *Working with human subjects: CRP Flyer #2*. Knoxville: Research Administration, The University of Tennessee.
- Correll, R. R., & McKenzie, J. F. (2005). *Health promotion and education research methods*. Boston: Jones and Bartlett.
- Department of Health and Human Services, Office of Human Research Protections
<http://www.hhs.gov/ohrp/>. January 3, 2008.
- DuBois, J. M. (2002). When is informed consent appropriate in educational research?: Regulatory and ethical issues. *Irb*, 24(1), 1-8.
- Fernandez, C. V., Kodish, E., & Weijer, C. (2003). Informing study participants of research results: An ethical imperative. *Irb*, 25(3), 12-19.
- Gribben, S., Norvell, S., & Van Vorst, B. (1994, January 17). The widening fallout. *Time*.
- Hornig, S., & Grady, C. (2003). Misunderstanding in clinical research: Distinguishing therapeutic misconception, therapeutic misestimation, and therapeutic optimism. *Irb*, 25(1), 11-16.
- Ingelfinger, F. J. (1978). Ethics of experiments on children. *New England Journal of Medicine*, 288, 791.
- Jacelon, C. S. (2007) Older adults' participation in research. *Nurse Research*, 14(4), 64-73.
- Kolodny, R. (1977). Ethical requirements: Informed consent. In W. H. Masters, V. E. Johnson, & R. Kolodny (Eds.), *Ethical issues in sex therapy and research*. Boston: Little, Brown.
- Levine, R. J. (1973). Ethical considerations in the publication of the results of research involving human subjects. *Clinical Research*, 21, 763.
- Mann, H. (2007). Deception in the single-blind run-in phase of clinical trials. *Irb*, 29(2), 14-17.

- McCartney, J. J. (1978). Research on children: National commission says "yes, if" *The Hastings Center Report*, 8(5), 26–31.
- Pappworth, M. H. (1969). Ethical issues in experimental medicine. In D. R. Culter (Ed.), *Updating life and death*. Boston: Beacon Press.
- Powelson, R. (1995, October 4). Radiation victims to be compensated. *Knoxville News-Sentinel*, p. 1.
- Retarded kids fed radiation in '50's. (1993, December 27), *Chicago Tribune*, p. 14, Section 1.
- Singer, E. (1978). Informed consent: Consequences for response rate and response quality in social surveys. *American Sociological Review*, 43, 144–162.
- Sjoberg, G. (1959). Operationalism and social research. In L. Gross (Ed.), *Symposium on sociological theory*. New York: Harper & Row.
- Slaughter, S., Cole, D., Jennings, E., & Reimer, M. A. (2007). Consent and assent to participate in research from people with dementia. *Nursing Ethics*, 14(1), 27–40.
- Sobel, D. (1981). Time out of joint: 25 days in a sleep lab. *The Hastings Center Report*, 11(1), 5–7.
- Thorne, B. (1980). You still takin' notes? Fieldwork and problems of informed consent. *Social Problems*, 27, 284–297.
- Veatch, R. M. (1977). Medicine, biology, and ethics. *The Hastings Center Report*, 7(6): suppl. 2–3.

Conducting Experimental and Quasi-Experimental Research

KEY TERMS

analysis of covariance	experimental group	multiple analysis of variance
analysis of variance	experimental method	parallel group trial
blind study	experimental mortality	placebo
confounding variable	external validity	population validity
construct validity	extraneous variable	random assignment
contamination	Hawthorne effect	randomized controlled trial
control group	history effect	randomized matching
crossover trial	independent variable	selection-maturation
dependent variable	instrumentation effect	interaction
differential selection	interaction effect	sequence effects
diffusion of treatment	internal validity	statistical regression
double-blind study	maturation effect	statistical validity
ecological validity	main effect	testing effect

Case Study A

Takisha was hired as a consultant to the state department of education. She was asked to determine if the new computer-based modules in smoking prevention were better than the current teaching techniques used in ninth grade. The department gave her a one-year contract to ascertain which of the two approaches was best for the students and the teachers.

Case Study B

Carol worked as a clinical trials nurse in a large, multispecialty physician group practice. A pharmaceutical company asked that the practice conduct a Phase III trial on a new product being developed for adult attention deficit disorder (ADD). Carol realized that

many factors play a role in this disorder but also knew that this physician practice had a large number of patients who could enroll in the study.

Characteristics of Experimentation

The **experimental method** of conducting research is an attempt to account for a factor in a given situation. It is generally considered to be the most highly regarded research method for hypothesis testing. The experiment carried out by the investigator is really a plan to garner evidence concerning the stated hypotheses. The natural environment is controlled and manipulated so that the researcher can observe and measure the results. When a true experimental situation is determined, the investigator is measuring the relationship between two or more variables in an attempt to discover the effect one variable might have on others.

Scientists began experiments with observation of the natural setting but realized that extraneous events were not being controlled. The next step was to perform experiments in the laboratory, where these extraneous factors could be controlled or at least taken into account. Physical and biological scientists used the laboratory method, and when, in the latter part of the 1800s, psychologists began using experiments in the laboratory, experimental psychology was born. However, experiments are not limited to the laboratory; they are achieved in the classroom and elsewhere but with much caution. It is understandable that children in classrooms cannot be randomly assigned to groups and randomly exposed to different teaching styles because doing so may lead to nonequivalent groups. Although there are some problems inherent in “real-world” research, our behavior takes place in the real world, and thus experimentation should occur in lifelike situations. Behavioral scientists must exercise extreme caution and regard for variables and controls because of these limitations.

An experiment has three characteristics: (1) a manipulated **independent variable**; (2) control of all other variables (**dependent variables**); and (3) the *observed effect* of the manipulation of the independent variable on the dependent variables. In our case study, if Carol wanted to examine the effects of a health science curriculum on the increase in knowledge of students, she would manipulate the curriculum (the independent variable) to determine the effect on achievement (the dependent variable). As can be deduced from the preceding discussion, the major issue in an experiment is the control of the independent and confounding variables.

Graziano and Raulin (2000) make two distinctions in experimental research designs. First is the difference between independent-group (or between-subjects) designs and correlated-group (or within-subjects or matched-subjects) designs. Independent-group designs have different participants in each group. In correlated-group designs, identical or closely matched subjects are in each group. Their second distinction is between univariate (or single-variable) designs and multivariable (or factorial) designs. The former have a single independent variable, while the latter have two or more independent variables.

Control in Experiments

Without controlling variables, the experimental method cannot exist. Therefore use of **control groups** is of utmost importance in an investigation. Control allows the scientist to arrange the experiment so that the effect of the variables can be studied. Because health scientists deal with humans outside of the laboratory, not *every* variable can be controlled. However, it is acceptable to attempt to control those variables that might have a significant impact on the experiment. For example, if Takisha in Case Study A wanted to determine the effects of the computer-based modules on ninth graders, she would need to have one group that would be exposed to the modules (Group A, the **experimental group**) and another group (Group B, the control group) that would receive the traditional instruction. All the students should be alike in variables that are likely to have an impact on the nature of the training: computer ability, reading ability, motivation, time, etc. However, other variables, such as artistic ability, height, and vocal ability, would be variables that could be ignored. Takisha would seek to choose two groups of students who would be most similar in the *significant variables*.

In Takisha's case, the experimenter is attempting to study the relationship between the independent and dependent variables. To do this, she would have to control for confounding variables. An **extraneous variable** is one that may affect the dependent variable and is not related to the major purpose of the experiment. Here Takisha would have to control for the extraneous variable of intelligence because it would have an influence on the dependent variable. If the students in Group A were more intelligent than those in Group B and Group A performed better, those gains could not be directly attributed to the new computer modules but to the higher intelligence level of the students in Group A. In this experiment, we have a confounding variable. A **confounding variable** is one in which independent and extraneous variables may each have an effect on the outcome of the experiment, and these effects cannot be separated.

When an experiment is carried out, the investigators must take precautions to be sure that there is as much equivalence as possible among the groups in the study. Several procedures are used to ensure equal groups: random assignment to group, randomized matching and analysis of covariance.

Random assignment is the assignment of experimental subjects to groups such that every member of the population has an equal chance of being assigned to any of the groups. The investigator numbers all the subjects in the population and uses a table of random numbers to draw the necessary number of subjects for each group. These groups are then considered to be *equivalent* in a statistical sense. In other words, the groups are so equal that any difference between the groups must be the result of chance alone and not of bias on the part of the investigator.

Randomized matching occurs when subjects are matched on as many extraneous variables as could possibly affect the dependent variable. Then the matched pairs are assigned to an experimental condition. Variables used for matching usually include sex, age, socioeconomic status, and reading or pretest score. In a school situation where groups preexist (classrooms), investigators match groups on the extraneous variables. The researcher will determine that scores of groups on standardized tests (IQ, reading,

pretest scores) are not significantly different in terms of means and standard deviations. Although group matching is not as ideal as individual randomization in certain real-world situations, this is the best method left to the investigator.

Analysis of covariance (ANCOVA) can be used to control for differences among the groups in the experiment. ANCOVA is a statistical method that analyzes differences of the experimental groups on the dependent variable only after initial differences on the pretest measures are taken into account. ANCOVA would probably be used in our case study of Takisha because the method is very useful for intact groups (classrooms). However, ANCOVA only partially controls the extraneous variables that may confound the independent and dependent variables, and attempts at random assignment should be made.

The Hawthorne Effect in Controlling Situations

The experimental situation itself must be controlled to ensure that the differences observed are caused by the dependent variable and not the extraneous situational variables. In the famous Hawthorne experiment, it was determined that any attention paid to subjects in experiments may cause them to behave in a way that they believe is expected of them. At the Hawthorne plant of the Western Electric Company, the workers *increased* their productivity no matter what the experimenters attempted. The study team eventually deduced that the increased productivity was caused by the attention the workers received as subjects in a study, an observation known as the **Hawthorne effect**.

Investigations involving the use of drugs like our Case Study B, routinely use a **placebo** (nonchemical look-alike) so that all subjects believe they are taking the drug. Otherwise, subjects might only react to the fact that they are taking a drug and act as might be expected, which would confound the results of the study. Again, control becomes a large part of the experiment. There are several methods to attempt to control situational variables: (1) hold the variables constant; (2) manipulate the variables systematically; and (3) randomize the situations.

Holding the variables constant is achieved by treating all subjects alike, regardless of group assignment, except for their exposure to the treatment. Ideally, the same teacher should teach the computer modules in Takisha's case study. Additionally, all tests, instructions, and general procedures should be as identical as possible for each group.

Manipulating the variables systematically includes controlling the order in which the experiment is given to the subjects. If the subjects were to take a series of achievement tests in relation to the health curriculum, it might be beneficial to split the group in half. One group would receive the decision-making tests first. This is an attempt to separate the groups from the main independent variable.

Randomizing the situational variables can provide a method to deal with the extraneous condition of having the same teacher for each group. The investigator could randomly assign half the experimental group and half the control group to each teacher. In this manner, extraneous conditions are not able to affect the dependent variable.

Advantages and Disadvantages of the Experimental Method

It is important to recognize the benefits and pitfalls you might encounter when conducting experimental research. In Case Study A, Takisha might discuss these issues with her superiors so that they, as well as she, would be aware of the possible effects an experiment might have on the school system. *Advantages* include the following:

1. *Convenience*: The investigator may decide to carry out the experiment whenever or wherever feasible. Of course, real-world research puts some limitations on convenience, but nonetheless, the experimenter may choose the time and location.
2. *Replication*: By repeating the experiment, or parts of it, the validity of the results increases. This is because the results are based on several observations rather than just one.
3. *Adjustment of variables*: Being able to vary an aspect of the experiment allows the investigator to attempt several steps at a rather rapid pace. For example, in the previously mentioned Hawthorne experiment, the investigators were able to vary the aspects of the workers' environment in succession.
4. *Establishment of cause-and-effect relationships*: This can be accomplished because the experimenter manipulates the independent variables and then observes the effects on the dependent variable. A caution here: Make sure your independent variable is valid because, if it is not, the effects may not be attributable to that variable.

Disadvantages of the experimental method include:

1. *Cost* can at times be a hindrance to experimental research. Many times, the training of experimenters and obtaining of equipment are expensive. When this happens, investigators must either carry out a bare-bones experiment or not conduct the study at all.
2. *Inability to generalize* the results of a study usually occurs because the samples used were not representative of the population. In our case study, if Takisha used a group of private school students in her experiment, it would be an inappropriate sample because the results of the experiment could not be generalized to public schools.
3. *Securing cooperation* from those in the experiment and from significant others (parents, administrators, supervisors) can be a major stumbling block for conducting a study.

Internal and External Validity

When designing an experiment the investigator must be certain that the study is technically sound. This is called the *validity*—internal and external—and should be dealt with to prevent problems that may cast doubt on the implications derived from the results of the study. Two other types of validity are statistical and construct (Graziano & Raulin, 2000). **Statistical validity** refers to the accuracy of the conclusion drawn

from a statistical test. **Construct validity** addresses the degree to which the underlying theory of the research effort explains the observed results.

Internal Validity

Internal validity can be defined as control for all influences between the groups being compared in an experiment, except for the experimental group. In our case study, Takisha would be comparing two methods of teaching smoking prevention to two groups. The only differences between these two groups would be the teaching method. Figure 5.1 illustrates this concept.

Internal validity is extremely difficult to achieve outside of the laboratory because there are too many extraneous variables to control. As we attempt to control for internal validity and tighten those controls, external validity (to be discussed later) suffers. As is so often true of the research in the real world, the investigator must compromise. Not all extraneous variables can be eliminated, but the experimenter, in designing a study, should take into consideration the many confounding variables. Confounding occurs when the independent variable varies with at least one other variable. These confounding variables, or threats to the research design, were originally discussed by Campbell and Stanley (1963) and include the following:

The **maturation effect** refers to factors that may influence subjects' performance because of the time that has elapsed. The change that has occurred within the participants is the problem because people normally change over time, regardless of interference (e.g., an experiment). This threat is especially apparent in longitudinal studies of young or adolescent children. Usually this problem can be attenuated by including a control group in the research design. The members of this additional group should be as comparable to the subjects in the experimental group as possible—they should have similar characteristics in their maturational development.

Group	Teaching Method for Smoking Prevention
A	Computer Modules
B	Traditional

Figure 5.1 Internal Validity Design

The **history effect** is defined as those events that occur at the same time as the study. These external events can interfere with the subjects' performance in the experiment. Sometimes these events are unpredictable, such as a catastrophe in the community (e.g., a tornado). Events such as this make history very difficult to control. In our case study, Takisha has decided to pretest the students on a day when they are having an examination in mathematics. Undoubtedly, the subjects will be under stress, and this may interfere with their performance on the pretest. One way she could attempt to limit the effect on the internal validity of the experiment would be to use a control group that would be exposed to the same historical experiences during the time of the study. While this may not account for an unexpected event (e.g., a lunchroom fight), it can control for some historical events.

It also should be noted here that the concept of history is within the experiment, including the procedures and materials used to conduct the study. All methods, test instruments, and situations should be the same for every subject in the study, regardless of group assignment.

Another complication involves what is known as the **testing effect**. Testing at the onset of an experiment, or actually before it begins, can have an effect on the subjects' performance at the time of the posttest. What is thought to occur is that the subjects practice taking the test and thus are "test-wise". Pretests may even give the subjects information (e.g., about cardiovascular health knowledge instruments) and thus threaten the internal validity of the experiment. This occurs because the investigator does not know if the results of the posttest resulted from the dependent variable, from the practice or knowledge gained from the pretest, or even from a combination of both situations.

To eliminate this test practice threat, two suggestions are offered: (1) use a comparison group that is exposed to the dependent variable but does not receive the pretest; and (2) increase the length of time between the administration of the pretest and posttest. While neither of these plans is absolutely foolproof in reducing the threat to the internal validity of the experiment, they do lessen the chance that the results of the study are suspect to testing practice.

Instrumentation effects can be a cause for concern in the internal validity of an experiment. The term refers to any changes in the instrument or measuring device used to test the effect of the dependent variable. In addition, changes in those who might be observers or raters in an experiment could unknowingly change their rating system. When considering a written instrument, the posttest should remain the same as the pretest, and procedures for recording the data (e.g., use of scanner sheets) should remain the same each time data are collected.

When observers or interviewers are used to collect data, they must be cognizant that they are fallible to fatigue, boredom, and awareness of the "right" answers. It is advisable that investigators check for intrarater and interrater reliability (Graziano & Raulin, 2000; Rubinson, Stone, & Mortimer, 1977) of these types of data collectors to alleviate the possible threat to internal validity.

Statistical regression presents a threat to internal validity when subjects are assigned to a group because of their extreme scores on tests. As an example, students who scored in the highest and lowest quartile on a nutrition knowledge test were chosen as subjects for the experimental group, and those in the middle 50% were eliminated from the

study. At the time of the posttest, the scores of the highest quartile would decrease toward the mean, while the scores for the students in the lowest quartile would increase toward the mean. This condition would produce differences between the groups on the posttest measures regardless of the intervention. So this will not occur, the investigator should make sure that the groups are composed of subjects whose scores represent the full range of possible scores.

Differential selection is actually a bias in selecting individuals for group selection. This usually occurs when participants volunteer for experimental group membership. In this case they usually are more highly motivated, which may cause them to be a biased group. Selection bias can also occur when intact classroom groups are assigned to either an experimental or control group. For example, the students in the seventh-hour science class also may be in the advanced algebra class, hence introducing a bias to the study. One of the best ways to avoid this particular threat to internal validity is to randomly assign subjects to each group. However, this usually cannot be facilitated in a school, for obvious reasons. Then the investigators must limit their generalization of findings to the particular sample in the study.

Experimental mortality is the loss of subjects in an experiment (also called *attrition*), especially if there is a differential loss between the experimental and control group. Many times subjects involved in studies in schools will be “lost” due to absence on the day of a test, or they will have moved out of the school district. If this happens equally between the groups, experimenters can select replacement subjects. Other methods to ensure against biased samples resulting from attrition are to choose large groups of subjects and make sure they are representative. In addition, it is wise to follow a sample of those who left the study to obtain comparison data.

Selection-maturation interaction occurs when the maturation of subjects becomes the confounding variable. As an example to explain this threat to internal validity, let us use Case Study A. The ninth-grade students selected for the experimental group to test the health computer modules were from Takisha’s school district. The control group of ninth graders was selected from another school district where, because of the different ages of the beginning students, the ninth graders were nine months younger than those in Takisha’s school district (the experimental group). If the experiment were to proceed and results indicated that the experimental group’s gains in knowledge were not as great as those in the control group, how could we interpret this data? The extraneous variable of maturation has confounded the results of Takisha’s experiment. To avoid this situation, investigators must be sure to select groups of comparable maturity levels.

Diffusion of treatment takes place when participants in a particular research condition communicate with participants in a different research condition. Those in an experimental group may inadvertently discuss events with participants in the control group, when the latter did not even know they were in such a group. This could affect how they might respond.

Sequence effects can be confounding because participants’ performance in later conditions may be the result of their role in a previous condition of the study. Within-subject designs are prone to this confounding. A study in which participants are exposed to three different conditions is conducive to this type of validity problem (Graziano & Raulin, 2000).

Another threat to internal validity, not discussed by Campbell and Stanley (1963), is that of **contamination**. This bias occurs when the researcher has previous knowledge concerning the subjects in the experiment. The investigator may inadvertently treat the groups differently or even give the subjects hints as to correct responses on surveys. In medical research, great care must be taken to assure the subjects do not know who receives the experimental medication and who receives a placebo; this is known as a **blind study**. A **double-blind study** is an even better safeguard against contamination. In this case the person who administers the treatments and records which subjects are in either the placebo or experimental group is not the experimenter, but rather an aide in the study. In our study, Takisha would have to make sure not to involve herself in the teaching and/or testing of the computer modules.

External Validity

External validity is the researcher's ability to generalize the findings of an experiment. In an attempt to control threats to internal validity in research in the behavioral sciences—real-world research—the investigator runs the risk of creating an unreal situation from which generalization to other settings is impossible. Most researchers outside of the laboratory tend to compromise and set up a rigorous experiment in a realistic situation.

Campbell and Stanley (1963) suggested four threats to external validity: (1) the reactive effect of testing; (2) the interaction effects of selection biases and the experimental variable; (3) the reactive effects of experimental arrangements; and (4) multiple-treatment interference. Bracht and Glass (1968) delineated Campbell and Stanley's four factors into a more specific set of threats to external validity. The following discussion is based on their work.

Population validity refers to the extent to which the results of an experiment can be generalized from the sample used in the study to a larger group of similar people. There are two types of population validity, according to Bracht and Glass. *The extent to which results can be generalized from the experimental sample to a defined population* is the first type. In our case study, if Takisha randomly selected a group of ninth graders to expose to the experimental module and found positive gains in terms of skills in decision making, she would like to generalize these findings to all ninth-grade students. However, she can only generalize to those students from whom the sample was drawn: ninth grade students in her school district. This sample is defined as the *experimentally accessible population*. When we read reports of experiments, we sometimes without thinking, generalize the sample to include all subjects (e.g., ninth-grade students in all junior high schools in New York). The latter group (ninth-grade students in all junior high schools in New York) is called the *target population*. When making these generalizations, the researcher must be sure that the two groups are representative of each other. This can be accomplished, although with intact classrooms in schools it is virtually impossible. However, large-scale studies, such as the one conducted by Ireson, Hallam, Mortimore, Hack, and Clark (1999), can be generalized to the target population. The schools in this experiment were truly representative of a cross-section of the schools in the United States.

Bracht and Glass wrote about a second type of population validity; *the extent to which personological variables interact with treatment effects*. Personological variables such as ability, sex, anxiety level, extroversion-introversion, and independence can have

an effect on students' performance. Hence if Takisha wanted to generalize the findings to another grade level (e.g., eighth grade), it would be unwarranted. This phenomenon has gained support as its own branch of research called *aptitude-treatment interaction (ATI) research*.

Ecological validity is the second major type of threat to external validity that Bracht and Glass described. They define ecological validity as the *extent to which the results of an experiment can be generalized from the set of environmental conditions in the experiment to other environmental conditions*. If the results can be obtained only under a very limited set of conditions, those results have low ecological validity. Bracht and Glass described several factors that may contribute to the ecological validity of an experiment:

1. *Explicit description of the experimental treatment*: Researchers must describe the experimental treatment in exact detail so that it can be replicated by other experimenters.
2. *Multiple-treatment interference*: When it appears that subjects will be exposed to more than one experimental treatment that may affect the generalizability of the findings of the study, the investigator should choose an experimental design in which one treatment is assigned to each subject.
3. *Hawthorne effect*: This phenomenon, which was discussed earlier, can prevent the findings in a study from being generalized because subjects simply reacted to being in a study.
4. *Novelty and disruption effects*: An experimental treatment may be effective just because it is different from what the subjects normally receive. This can cause low generalizability because, as the novelty wears off, so does the effectiveness of the treatment.
5. *Experimenter effect*: This refers to the inability of the person who administers the treatment (physician, teacher, nurse practitioner) to be involved in subsequent investigations. This is also an area of concern for low generalizability.
6. *Pretest sensitization*: Just as the pretest can affect internal validity, so can it act in an adverse way on the posttest performance of subjects. The usual occurrence is that the pretest positively affects the scores on the posttest.
7. *Posttest sensitization*: As in the case with pretest sensitization, at times the posttest can affect the subjects' scores and thus cause concern for generalizability of the experiment.
8. *Measurement of the dependent variable*: When instruments are used to measure the effects of a treatment, they may limit the generalizability of the results to the other tests if they are particularly well adapted to the dependent variable.
9. *Interaction of time of measurement and treatment effects*: Experiments may use two or more posttests to measure the effects of an intervention. Usually, the first posttest is administered immediately after the intervention is concluded, and then again several weeks or months later to measure the subjects' retention. These second measurements have changed the effects of the intervention and therefore pose a threat to ecological validity.

Investigators must take into account the many threats to internal and external validity when they design an experiment. Because most behavioral scientists work in the real world, some threats must remain, while others can be minimized and even excluded. If investigators find that a discrepancy exists between the experimental condition and real-world setting, then they should note this in the report as a limitation to the generalizability of the study.

Constructing Experimental Designs to Control Variables

The major purpose in constructing an experimental design is to control as many extraneous variables as possible. Because experimental research in the health sciences is usually conducted outside a laboratory, the investigator must take great pains to ensure that the proper controls are employed. However, in attempting to control for so many variables, real-world research can produce artificial results. This occurs because the environment and/or the subjects are sometimes put into unnatural situations. Snow (1974) has developed an alternative called *representative design*, whereby experiments accurately reflect real-life environments and the natural characteristics of the learners.

We shall discuss several experimental designs, which can be divided into the following categories: preexperimental, true experimental, factorial, and quasi-experimental designs. To simplify this process, we will include symbols and terms to describe designs. These symbols are widely used in the literature and include:

X: Independent variable that is manipulated

O: Process of observation or test

R: Random selection of subjects to groups

Xs and Os across a given row: Apply to the same people

--: Dashed lines between groups indicate no random assignment to groups.

The left-to-right dimension indicates the temporal order, and the Xs and Os, when vertical to one another, are given simultaneously.

Preexperimental Designs

The least adequate experimental designs fall into this group; there is no control group, and extraneous variables can cause threats to internal validity. As we progress from these types of experiments to those that do not have such weaknesses, you will be able to build an experiment that will avoid such problems.

The One-Shot Case Study Design

X O

In this type of experiment, a treatment is given to one group. Then observations (O) are made on the subjects in that group to detect the effects of the treatment. Those observations are made in the form of a posttest. There is no control group, but rather

the investigator makes inferences or comparisons of the results based on what the results would have been had the intervention not been given. In Case Study A, Takisha would simply have a group of ninth-grade students complete the computer modules and then determine the results (treatment effects) through a posttest. The sources of invalidity in this design are history, maturation, selection, and mortality. This is the very weakest of experimental designs and should be avoided. Carol, in Case Study B, could not even consider this design for her clinical trials study.

The One-Group Pretest-Posttest Design

$$O_1 \quad X \quad O_2$$

$$O_1 = \text{pretest}$$

$$O_2 = \text{posttest}$$

Although this design improves the one-shot case study experiment, it is still considered to be weak. The group is administered a pretest to measure the dependent variable and then the treatment is introduced. The same test is readministered at the conclusion of the intervention (posttest). Takisha, in attempting to test the computer module, did give the students a pretest and posttest. She found that the students increased their knowledge significantly after being exposed to the intervention. Is she correct in believing that the gains in scores resulted from the new module? Probably not, because this design does not control for student maturation, history, the testing situation, or statistical regression. Case Study B cannot use this design.

The use of the one-group pretest-posttest design is sometimes necessary in school systems that will not allow their students to be treated differently. In other words, some school districts insist that a new or experimental program must be made available to all students. In this case, it is necessary for the investigators to make estimates of the gains the subjects would make and then set their experimental significance levels against that standard. Therefore, this type of design can be safely used because most of the extraneous factors that cause threats to validity have been estimated to reach certain levels by the investigators. To test the differences of scores between the pretest and posttest, the usual method of analysis would be to employ the *t*-test for correlated means because the same subjects take both tests.

The Static-Group Comparison Design

$$\begin{array}{c} X \quad O \\ \hline O \end{array}$$

The static-group comparison experiment includes two treatment groups (experimental and control) in which each is given a posttest only. However, the subjects are not randomly assigned to the treatment groups, as indicated by the dashed lines (---) above. Here, selection and mortality interfere with the internal validity of the experiment because the results of the experiment may not be attributable to the intervention but to the differences in the subjects in each group. It is necessary that the groups be equivalent; and, because random assignment was not employed, the investigator cannot ensure that the groups are equal.

In our case study, Takisha selected Ms. Evans's class as the control group and Mr. Jackson's class as the experimental group, and each group would be given a posttest after the computer modules were taught. She analyzed the data using a t -test of the posttest mean scores and found that Mr. Jackson's class did score significantly higher on the tests than did Ms. Evans's class. Does this mean that the new computer modules are better than the usual series that Ms. Evans used? Not necessarily so, because Mr. Jackson's students may have had more knowledge than Ms. Evans's students. They may not have been equivalent at the start of the study, a serious limitation of this type of experimental design.

True Experimental Designs

In these types of real experimental designs, control groups and experimental groups are involved in the study, and subjects have been randomly assigned to each group. These are the strongest types of designs, but Takisha will have difficulty using them because she is dealing with intact groups (i.e., subjects in classrooms). However, she can assign *classrooms* randomly to groups, and then check students' records for standardized test scores to determine the homogeneity of the classes. In Case Study B, Carol requires a true experimental design for a Phase III clinical trial.

The Posttest-Only Control Group Design

R	X	O ₁
R		O ₂

In this true experimental design, the experimental group experiences the intervention, but the control group does not. In addition, subjects are randomly assigned to each group, and they both receive the same posttest. This is a very powerful design because it controls for all threats to internal validity with the exception of mortality. This is especially useful when the researcher has a large group of subjects because random assignment of the subjects allows for equivalence of the groups.

The pretest is not used in this design, which can be beneficial in several ways. There is no interaction between the pretest and the independent variable, and therefore the design should be used when there is likelihood that pretest reactivity would occur. In addition, there are cases in which pretests are not appropriate (studies of very young children in whom learning has not yet happened) or not available. A disadvantage of this design is that you cannot determine if change has occurred. Takisha would have to use a table of random numbers to select subjects to be assigned to experimental and control groups. Sixty students were chosen from an initial pool of 400 ninth graders. The experimental group would receive the computer modules, while the control group would be taught traditionally. All factors would be equated; and, at the end of the intervention, both groups would be posttested. The data would be analyzed by using a t -test comparison of the mean posttest scores of the groups. If Takisha had used more than two groups, she would use analysis of variance. The results of the statistical scores show that the experimental group's scores were significantly higher than those of the control group. What can she conclude in regard to the new computer modules?

In setting up her clinical trials design, Carol could use the posttest-only control group but would be lacking significant clinical information prior to the intervention. Therefore, it is unlikely she would choose this design.

The Pretest-Posttest Control Group Design

R	O ₁	X	O ₂
R	O ₃		O ₄

This design includes a pretest addition to the previously discussed true experimental design: posttest-only control group design. The subjects in the experimental and control groups are randomly assigned, and each group is given both a pretest and posttest. This design eliminates all threats to validity and thus provides an excellent setting for conducting a study. Randomization of the subjects will ensure that there was no systematic bias in the groups; however, in some cases, there *may be* initial differences between groups, as shown on pretest scores. Once again, Takisha randomly selected her subjects, assigned them to groups, administered the pretests, had the appropriate teacher introduce the computer modules, and posttested each group. She analyzed the data by using analysis of covariance. The posttest mean scores were compared with the pretest scores as a covariate. In this experiment, no significant differences were found on the scores between the groups, but there were some that increased, showing positive changes in the experimental group's scores. See if you can interpret these results.

This design allows Carol in Case Study B to collect relevant clinical information about both groups; the intervention group receiving the new pharmaceutical agent and the control group receiving a placebo.

The Solomon Four-Group Design

R	O ₁	X	O ₂
R	O ₃		O ₄
R		X	O ₅
R			O ₆

This is a very sophisticated experimental design that takes into account factors associated with external as well as internal validity. The design is set up to determine several factors at once: assess the effect of the treatment in relation to the control group, determine the effect of the pretest, and explain the interaction between the treatment conditions and the pretest. Random assignment of subjects to groups and inclusion of groups that are not pretested add to the significance of the results of the experiment. This results from combination of the previously discussed designs: posttest only and pretest-posttest. The design actually has two experiments going on at once and thus provides replication of the study.

However, there is a major disadvantage to this design: finding enough subjects to complete the four equivalent groups and randomly assigning them into groups. Even if enough subjects can be located, the time to conduct two experiments may not be available to the investigators. Because several school health science studies (Haight, Michel, & Hendrix, 1998; Huss & Ritchie, 1999; Peterson & Rubinson, 1982) have been conducted

using the Solomon Four-Group design, Takisha has some frame of reference from which to conduct her investigation.

The analysis for this design is not simple but can easily be handled with today's computers. Campbell and Stanley (1963) suggest disregarding the pretests, except as a treatment, and analyzing the posttest scores with a simple analysis of variance design of 2 (pretest) \times 2 (treatment), as shown:

	No X	X
Pretested	O_4	O_2
Unpretested	O_6	O_5

From the mean scores in the columns, one estimates the main effect of X and, from the mean scores in the rows, the main effect of pretesting. The cell means provide an estimate of the interaction of testing with X . If the main and interactive effects of pretesting are negligible, an analysis of covariance of O_4 versus O_2 with the pretest scores as the covariate, should be performed.

The Randomized Controlled Trial: Clinical Trials

The **randomized controlled trial** (RCT) is generally thought of as an experimental design for medical and pharmaceutical investigation but has been broadened to incorporate prospective clinical designs to test the efficacy of an intervention against a controlled condition (Duffy, 2006). As clinical entities embrace evidence-based practice, the RCT is viewed as the best type of evidence on which to base decisions and establish guidelines. Generally, a clinical trial is a study of human volunteers to address specific health questions and is usually the most expedient and safe way to find treatments that improve health. The reason many people volunteer is so they can play a more active role in their own health and perhaps gain access to new interventions well before they are available to the general public. However, every clinical trial has a set of inclusion and exclusion criteria as to who can participate. The former dictates who can be allowed in the study, while the latter disallows someone from participating. The criteria, such as age, gender, type and stage of disease, prior treatment, and other conditions, are simply used to identify participants and determine if they qualify. The principal purpose is to keep the participants safe. In Case Study B, Carol would establish inclusion / exclusion criteria so that she can state her results with greater assurance.

While the emphasis of this section is on Phase III clinical trials, it should be understood that there are four phases or steps when testing a new pharmaceutical agent or surgical device. Figure 5.2 illustrates the four phases across several criteria. Phase I is the introduction of the new drug or device into a human being. Of course, *in vitro* (laboratory) and *in vivo* (animal) testing is done beforehand to ensure safety. Phase I is most concerned about the safety of the drug or device in humans (Rich, 2004). This trial is usually not controlled or randomized. The participants are very closely monitored as they are given incremental doses of the drug.

	Phase I	Phase II	Phase III	Phase IV
Objectives	Safety of drug or device	Safety & efficacy of dosage and frequencies of administration Device efficacy	Safety & efficacy over a long period Risk / benefit ratio Get FDA approval	After FDA approval For new indications, expanded safety, new information
Design factors	Not randomized Not controlled Monitored closely Single site	May or may not be randomized Placebo may be used May be blinded Well-defined entry criteria Few sites	Full scale RCT Randomization Placebo control Double-blinding Broader entry criteria Multiple sites	Generally observational Uncontrolled surveys No comparisons among interventions Multiple sites
Duration	Brief (usually less than one month)	Average 2 years	Several years	Ongoing with FDA approval
Participants	Healthy volunteers	Individuals with targeted disease	Individuals with targeted disease	Individuals with targeted disease, as well as new age groups, etc.
Sample size	20–80	100–300	Several thousand	Several thousand

Figure 5.2 Clinical Trial Phase Comparisons

Phase II still has safety and efficacy as high priorities and only occurs if safety has been demonstrated in Phase I trials. Unlike Phase I, these participants usually have the condition under study. Phase II trials provide much needed evidence of clinical significance. From a design perspective, these trials are often placebo-controlled and double-blinded (neither the investigator nor the participant knows whether he or she is receiving the placebo or the drug being studied). Of course the researcher has access to a randomization code that can be broken in case of emergencies to see if the participant was on the investigative drug (Ginsberg, 2002).

A Phase III trial is usually a full-blown RCT (Duffy, 2006). They prove the safety and efficacy of the intervention (drug / device) over a longer period of time. The researcher evaluates the risk / benefit ratio with a large number of participants at multiple sites. The new drug or device is compared with current standard treatment (control group) if one exists. Randomization, placebo-control, and double-blinding are customary (Rich, 2004). Carol would no doubt be one of several researchers contributing data to the overall study. In conjunction with the pharmaceutical company, she would establish inclusion / exclusion criteria for her adult patients, such as prior diagnosis of ADD (persistent, inappropriate inattention, impulsivity, non-goal directed behavior) affecting self-esteem, family and social relationships, delinquent behavior, drug abuse, and increased risk of failure. These included patients may require further

screening such as a physical exam, electrocardiogram, and routine lab tests. They would then be randomly assigned to the intervention or control groups. In a **parallel group trial**, such as the pretest-posttest control group design, only one set of patients receives the new drug. The comparison is between the two groups. However, Carol could design a **crossover trial** in which all participants receive the new ADD drug. In this design, patients are randomized into an intervention group or a control group, with the intervention group taking the medication for a selected period of time (such as one month). After that period of time, the original intervention group becomes the control group taking the placebo while the original control group serves as the intervention group receiving the new drug (Duffy, 2006; Silva et al., 2006). The crossover design is shown in Figure 5.3.

The crossover design allows for evaluation after intervention with one group and then again following intervention with the second group. Note that participants in this design serve as their own controls, and within-group comparisons are made. Moreover, when participants serve as their own controls there is greater power, thereby requiring fewer people to produce statistically and clinically significant results (Jadad, 1998). Consequently, it would be to Carol's advantage to employ this design rather than the parallel design.

Phase IV trials are conducted after approval by the U. S. Food and Drug Administration (FDA) and are used to monitor adverse effects, check for new usage indications, gain more information about the product, or educate clinical researchers. The length of the study will vary with the objectives.

The RCT is an effective design, and the need for clinical trials is growing—particularly since many sites fail to meet the pharmaceutical industry's needs for adequate numbers and clean data. Investigators need to be aware of how to plan the clinical trial (Boissel, 2004), as well as the methodological idiosyncrasies (Birch, 2006; Hahn, Puffer, Torgerson, & Watson, 2005).

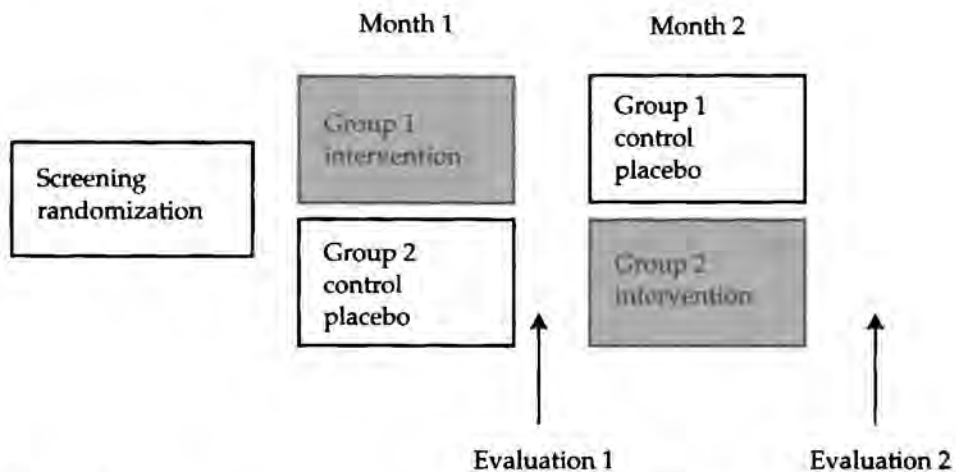


Figure 5.3 Crossover Design for Randomized Clinical Trials (RCT)

Factorial or Multivariable Designs

Factorial designs are more complicated than the previously discussed designs in which there was a single variable and one independent variable that was manipulated to have an effect on the dependent variable. A factorial design is one in which two or more variables are manipulated simultaneously to allow study of the independent effect of each variable on the dependent variable. In addition, the effects caused by the interaction among the several variables are assessed. The effect of each independent variable on the dependent variable is called a **main effect**, while the effect of the interaction of two or more independent variables on the dependent variable is termed an **interaction effect**.

Imagine that a researcher investigating distraction in students with ADHD (attention deficit hyperactivity disorder) hypothesizes that the complexity of the task at hand plays a role. The investigator feels that the number of students in the room is a significant factor in distraction, too. While each independent variable can be studied separately, the researcher would like to know how each interacts with the other in so far as impacting overall distraction levels (interaction effect). In other words, how does the number of students in the room combined with task complexity interact to affect distraction in these students? In most real life situations, there is more than a single factor playing a role. Factorial designs allow the researcher to investigate several factors or variables in a single experiment.

Using the example of distraction, the researcher can design a 3 × 2 factorial table. Figure 5.4 illustrates the combination of independent variables. The total number of conditions is six, derived from the number in the room (three different settings) and the types of tasks (simple and complex). Every level of every variable is paired with every level of every other variable. Only factorial designs allow the investigator to test for interactions among variables.

	Factor B (Task Complexity)	
	Simple	Complex
Factor A (Number in the Room)		
1		
10		
20		

Figure 5.4 Factorial Design with Two Independent Variables

Quasi-Experimental Designs

There are times when researchers cannot control all sources of internal and/or external validity. These occasions call for the investigator to use a quasi-experimental design, which although not as strong as the true experimental designs, is certainly preferable to the preexperimental designs. Experiments in this group use designs in

which random assignment of subjects to groups has not been accomplished. These experiments are usually carried out where intact groups are available (i.e., schools). In our case study, Takisha would have to use a quasi-experimental design, as she had determined that school officials will not allow students to be reassigned from their original classes.

Campbell and Stanley (1963) originally described 13 quasi-experimental designs; however, we have decided to explain only those that would have a direct bearing on the health sciences: time series, equivalent time samples, nonequivalent control groups, and counterbalances. Each design has strengths and weaknesses that enable investigators to choose the design that would be most appropriate for a particular setting.

The Interrupted Time Series Design

$$O_1O_2O_3O_4 \quad X \quad O_5O_6O_7O_8$$

If Takisha wished to have the entire school system involved in her study, she would want a comparable school district to serve as a control group. Unfortunately, she cannot find one that is both comparable and willing to join in the experiment. Therefore she decides to use the interrupted time series design, as diagrammed above.

As evidenced, there is one experimental group in which observations occur before and after the intervention. Several sources of validity are not threatened when using this design, but the general weakness includes a threat to the effect of history; other influences might have an effect on the dependent variable after it is introduced. This can be controlled by adding a control group, but in our case study this is impossible. Another minor weakness of interrupted time series designs is that of the effect of instrumentation. During this experiment, there are many points at which data are collected, and those who administer the tests may do so differently and even score them improperly. This is especially true of data and record keeping in hospitals, where many different personnel may be involved in this activity.

A caution that all researchers should consider when using interrupted time series designs is the cyclical nature of performance and attitudes, which may be attributed to seasonal variations. For example, more hospital admissions occur during the Christmas holidays, and schoolchildren will not respond well to serious interventions the day before a vacation. Several other interrupted time series designs may be used. (For a thorough discussion of these, refer to Cook & Campbell, 1979; and Graziano & Raulin, 2000.)

There are a few limitations associated with interrupted time series designs. They include:

1. Treatments are rapidly implemented but rather slowly diffused.
2. Effects are not instantaneous.
3. Data observations are too few—50 are recommended.

The statistical analysis for interrupted time series designs depends on the number of observations within the experiment. If 50 to 100 observations are available, it is recommended that the *autoregressive integrated moving average* (ARIMA) models be used. Refer to Box and Jenkins (1976) for a complete description of this technique. However,

if fewer than 50 observations occur and the errors are independent, *analysis of variance* (ANOVA), with repeated measures, should be used to analyze the data. If the errors are correlated and the number of observations is small, *multiple analysis of variance* (MANOVA) is a satisfactory statistical technique.

The Equivalent Time Sample Design

$$X_1O \quad X_0O \quad X_1O \quad X_0O$$

The equivalent time sample design is similar to the time series design; however, there are repeated introductions of the treatment or independent variable. As illustrated, the treatment (X_1) is introduced and then responses observed. After a period of time, an observation is done again without the effect of the treatment (X_0). This design is useful when the researcher believes the independent variable or treatment brings about results that are reversible or transient. For example, suppose an educator addresses a group of students in a driver education course about the importance of wearing a seatbelt and includes a short film showing the accidents in which seatbelts were not worn. She may find that seatbelt usage is high following the treatment. However, over a period of time, there may be a decrease in usage. Her hypothesis that the education intervention is transient could be tested with this design. In reality, the time interval between the treatment and the control (X_0) is generally random rather than set intervals. This design can be used with a large number of participants or even just one participant.

The equivalent time series design controls for all threats to internal validity, therefore covering the threat of history, which the time series design does not. In other words, extraneous events should not have an effect on the outcome of the experiment.

Two types of generalizations occur in this type of experiment: (1) across occasions; and (2) across subjects. Analysis of the data may be accomplished by comparing the two experiences against a between-occasions-within-experience error term. This provides information about the changes that occur over time.

The Nonequivalent Control Group Design

$$\frac{O_1}{O_3} \text{-----} X \text{-----} \frac{O_2}{O_4}$$

In this design, both groups are pretested and posttested, no control group is exposed to the intervention, and none of the subjects have been randomly assigned to groups. This design and its many variations (adding nonpretested experimental and control groups) are frequently used in situations such as our case study. As mentioned previously, classrooms are intact, as are other groups: prisoners, some patients in hospitals (dependent on illness), and so on. However, what does become random is the assignment of the intervention. While this design can appear weak, it does become sound in that subjects are usually chosen from a pool of those with somewhat similar characteristics (e.g., age, experience, grade in school). This homogeneity can be tested by comparing scores of the groups on the pretest. The interaction of selection and maturation can be a threat to the validity of the nonequivalent control group design.

To statistically analyze this design, the investigator should use ANCOVA, which would reduce the effects of initial differences between the groups because it makes adjustments to the posttest scores of each group. If you use ANCOVA, remember that only the *measured* variables are used as covariates. There is the possibility that subjects in the groups may differ on other, nonmeasured variables that have not been used in ANCOVA. The investigator also can use multiple regression to analyze the data from the results of the experiment. This method is advantageous compared to ANCOVA only because it has less stringent assumptions than ANCOVA. In multiple regression, the posttest scores would be the criterion variable, and the investigator could then decide if the treatment was significant to the prediction.

The Counterbalanced Design

	Time ₁	Time ₂	Time ₃	Time ₄
Group A	X ₁ O	X ₂ O	X ₃ O	X ₄ O
Group B	X ₂ O	X ₄ O	X ₁ O	X ₃ O
Group C	X ₃ O	X ₁ O	X ₄ O	X ₂ O
Group D	X ₄ O	X ₃ O	X ₂ O	X ₁ O

Counterbalanced designs, also called Latin squares, are usually used when two or more interventions are to be tested. All the subjects will receive all the treatments for a specific time period. The order of administration of treatments is varied between subjects so that order effects are not confounded with treatment effects. These have been called *rotation experiments*. In our Case Study A, Takisha would use counterbalancing to test the effectiveness of two methods of teaching smoking prevention to ninth-grade students. The two units, however, must be equal in all aspects—difficulty, age-relatedness, factual information, and the like. After each class completed the unit, a cognitive test would be administered to the subjects. This design is especially useful for intact groups and actually is slightly better than the nonequivalent control group design in that it might “rotate out” differences that could exist between groups. In addition, students are matched to themselves, making statistical analysis more discriminatory.

There are some weaknesses in carrying out an experiment like the one just described. A carryover effect from one teaching method to another may affect the scores on the tests. In addition, it is sometimes very difficult to determine that units of study are, in effect, equal in all aspects as prescribed. Also, it could be possible that students would tire of taking as many tests as required in the counterbalanced design.

Case Discussion for Case Study A

Takisha’s objective was to determine if the new computer-based modules in smoking prevention were better than the current teaching techniques used in ninth grade. To select her sample, Takisha reviewed her sampling frame of all ninth grade-students in her own school district. Her design choice was the pretest-posttest control group design. She decided to randomize matching of schools to control for as many extraneous variables

as possible. From the matched pairs, she randomly assigned one from the pair of schools to the experimental or treatment group to receive the computer-based modules. It is important to keep in mind that she is dealing with intact groups rather than individual students, so the unit of analysis becomes the schools. The teachers of the new modules will receive extensive in-service training prior to commencement of the program. Another in-service will be held for teachers of the traditional method. The teachers will administer knowledge and attitude, instruments selected for their reliability and validity. The initial testing will be immediately prior to instruction on smoking prevention, and the second follows the completion of the program.

Case Discussion for Case Study B

Carol worked as a clinical trials nurse and was asked to conduct a Phase III trial on a new product being developed for adult attention deficit disorder. She was well aware that a randomized controlled trial (RCT) design would be used, requiring randomization, placebo control, and a double-blind design. As in all clinical trials, Carol created inclusion / exclusion criteria to acquire her study sample from all the patients available through the large multispecialty practice. Those meeting the criteria through screening were assigned to an intervention group or a control group. She decided to use a crossover design rather than a parallel design so participants serve as their own controls and within-group comparisons are made. Carol realized that when participants serve as their own controls there is greater power, thereby requiring fewer people to produce statistically and clinically significant results. During the course of the study, she would ensure that participants followed the research protocol presented to them at the time of informed consent. The protocol simply describes the schedule of tests, procedures, medications and dosages, and the length of time for the study. Of course, all participants are monitored closely for safety reasons, as well as to determine the effectiveness of the new ADD medication.

SUMMARY

This chapter attempted to introduce and explain the very important concept of *control* in conducting experimental research. Good research studies ensure and take into account the advantages and disadvantages of conducting an experiment. While experimentation originally began in the laboratory, it has been successfully conducted in naturalistic settings such as schools, factories, ships, and hospitals. When conducting an experiment, the investigator must be concerned with several threats to internal validity, including variables, history, selection, instrumentation, testing, differential selection, experimental mortality, selection-maturation interaction, and contamination. Another type of validity, external validity, was discussed with specific reference to population validity and ecological validity.

Experimental designs can be classified into four categories: preexperimental, true experimental, factorial or multivariable, and quasi-experimental. Ten designs were discussed with respect to controlling extraneous variables: the one-shot case study,

one-group pretest-posttest, static-group comparison, posttest-only control group, pretest-posttest control group, Solomon Four-Group, interrupted time series, equivalent time samples, nonequivalent control group, and counterbalanced. Investigators may use each of these designs based on need, availability of subjects, proposed hypotheses, and suitable data analysis capabilities.

The randomized controlled trial (RCT) was presented as means to provide evidenced-based practice or medicine with the information necessary to make sound clinical decisions. While there are four phases in clinical trials, the emphasis was on Phase III trials requiring randomization, placebo control, and double-blind design.

CRITICAL THINKING QUESTIONS

1. How do true experimental designs differ from (a) preexperimental and (b) quasi-experimental designs?
2. Why would patients want to enroll in a clinical trial?
3. Why do Phase III clinical trials generally require randomization, placebo control, and a double-blind design?
4. Why are true experimental designs difficult to implement when students in grades K–12 are the participants?
5. What role would informed consent play in both Takisha's study and Carol's study?

SUGGESTED ACTIVITIES

1. Explain the differences and similarities in the following designs:

Design 1				Design 2		
R	O ₁	X	O ₂	R	X	O ₁
R	O ₃		O ₄			O ₂

2. You are conducting a smoking-cessation program, which has three groups of smokers. Each meets on a different evening once a week for a total of six weeks. You pretest each of the intact groups prior to implementing the smoking-cessation program. Upon completion of the program, you immediately administer a posttest and then repeat it eight weeks later. The design would look as follows:

Monday group	O ₁	X ₁	O ₂	O ₃
Tuesday group	O ₄	X ₂	O ₅	O ₆
Thursday group	O ₇	X ₃	O ₈	O ₉

- a. What is the meaning of each of the following sets of data?
Set 1: O₁ = O₄ = O₇ = O₂ = O₃ = O₅ = O₈, but O₆ is greater than O₅, and O₉ is greater than O₆.

Set 2: O_1 , O_4 , and O_7 are not equal because O_1 is greater than O_4 , which is greater than O_7 ; and O_2 is greater than O_5 , which is greater than O_8 . However, O_3 , O_6 , and O_9 are equal.

- b. Are you able to check for the effect of pretesting using this design? Explain your answer.
 - c. Does the lack of randomization make any difference in your study since you are pretesting the groups?
3. You have developed the following design to compare the effects of two different drugs for hypertension on three groups, all of which are randomized and pretested. All are then given two posttests, three months apart.

Group 1	R	O_1	X_1	O_2	O_3
Group 2	R	O_4	X_2	O_5	O_6
Group 3	R	O_7		O_8	O_9

- a. What is the advantage of using randomized subjects versus intact groups?
 - b. Why did you decide to pretest rather than posttest only?
 - c. Explain these results: $O_3 = O_6 = O_9$ but $O_2 > O_8$ but $< O_5$
 - d. Interpret this: $O_2 = O_5$ but $> O_8$; $O_3 > O_6$ and O_9
 - e. How would you explain the results if O_7 , O_8 , and O_9 were not equal as observations in the control group?
4. Go the website <http://clinicaltrials.gov/> sponsored by the National Institutes of Health. Click on each of the following links: *Understanding Clinical Trials*, *What's New*, and *Glossary*. Review this information to gain a better understanding of clinical trials.
 5. Go the website <http://clinicaltrials.gov/> sponsored by the National Institutes of Health. Click on the following link: *List by Condition*. Next, click on *Behavior and Mental Disorders*. Now click on *Attention Deficit Disorder with Hyperactivity*. This brings you to a long list of studies showing which ones are active. Select an active study, and review inclusion / exclusion criteria and related clinical trial information.

REFERENCES

- Birch, S. (2006). A review and analysis of placebo treatments, placebo effects, and placebo controls in trials of medical procedures when sham is not inert. *Journal of Alternative and Complementary Medicine*, 12(3), 303–310.
- Boissel, J. P. (2004). Planning of clinical trials. *Journal of Internal Medicine*, 255, 427–438.
- Box, G., & Jenkins, G. (1976). *Time-series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Bracht, G., & Glass, G. (1968). The external validity of experiments. *American Educational Research Journal*, 5, 437–474.
- Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cook, T., & Campbell, D. (1979). *Quasi-experimentation*. Boston: Houghton Mifflin.

- Duffy, M. (2006). The randomized controlled trial: Basic considerations. *Clinical Nurse Specialist*, 20(2), 62–64.
- Ginsberg, D. (2002). *An Investigator's Guide to Clinical Research* (3rd ed.). Boston: Centerwatch.
- Graziano, A. M., & Raulin, M. L. (2000). *Research methods: A process of inquiry*. Boston: Allyn & Bacon.
- Hahn, S., Puffer, S., Torgerson, D. J., & Watson, J. (2005). Methodological bias in cluster randomized trials. *BMC Medical Research Methodology*, 5(10), 1–8.
- Haight, B. K., Michel, Y., & Hendrix, S. (1998). Life review Preventing despair in newly relocated nursing home residents. *International Journal of Aging and Human Development*, 47(2), 119–142.
- Huss, S. N., & Ritchie, M. (1999). Effectiveness for a group for parentally bereaved children. *Journal for Specialists in Group Work*, 24(2), 186–196.
- Ireson, J., Hallam, S., Mortimore, P., Hack, S., & Clark, H. (1999, April 19–23). *Ability grouping in schools: An analysis of effects*. Paper presented at the annual meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED 430989).
- Jadad, A. R. (1998). *Randomized controlled trials: A user's guide*. London: British Medical Journal Books.
- Peterson, F., & Rubinson, L. (1982). An evaluation of the effects of the American Dental Association's dental health education program on the knowledge, attitudes, and health locus of control of high school students. *Journal of School Health*, 52(1), 363–369.
- Rich, K. (2004). An overview of clinical trials. *Journal of Vascular Nursing*, 22, 32–34.
- Rubinson, L., & Stone, D. (1975). An evaluation of the positive aspects of the American Dental Association's oral health program, level II. *Proceedings of the National Symposium on Dental Health Education in Schools*, 1–16.
- Rubinson, L., Stone, D., & Mortimer, R. (1977). Statistical reliability of NPI administered to a sample of sixth grade children. *Dental Hygiene*, 51(3), 109–113.
- Silva, R. R., Muniz, R., Pestreich, L., Childress, A., Brams, M., Lopez, F. A., & Wang, J. (2006). Efficacy and duration of effect of extended-release dexamethylphenidate versus placebo in school children with attention-deficit/hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology*, 16(3), 239–251.
- Snow, R. (1974). Representative and quasi-representative design for research on teaching. *Review of Educational Research*, 44, 265–291.

Data Collection Through Surveys and Self-Reports

KEY TERMS

analytical design	noncontact rate	semistructured interview
branching questions	open-ended	sentence completion
descriptive design	ranking	structured interview
dichotomous questions	rating	unrestricted or open
flow plan	refusal rate	unstructured interview
Internet surveys	response rate	
multiple-choice	restricted or closed form	

Case Study

Elizabeth was recently hired by the community hospital as the Director for Outreach Education to plan, implement, and evaluate hospital-sponsored health education programs. Her immediate administrator explained that resources would be available to her for 1 year and subsequently that her outreach programs would have to stand alone—at least economically. In other words, Elizabeth had 1 year in which to make her programs self-sufficient. Although some programs had been conducted in the past for the two-county catchment area, no one knew if they were successful or, if they were, what the people really wanted or needed. Elizabeth knew that she had to find out a lot about (1) her catchment area, including the nature of residents—demographics and health needs, (2) the resources available to her in terms of both personnel and money, and (3) the best way to deliver her programs so that she could reach the greatest number of people and remain within her budgetary constraints. She realized that research was necessary if she was to be successful.

Characteristics of Survey Research

One kind of research that often appears in the health science literature is survey research. For many, this type of descriptive investigation is viewed as being unworthy and misusing funds. Not unlike other research methods, this perception is correct if the

survey is conducted in poor fashion. Trow (1967) observed the following about this form of research:

The errors and inadequacies of survey research in education appear at many points from the way problems are initially chosen and defined to the choice of the subject population, the selection of the sample, the design of the individual questions and the questionnaire as a whole, and the analysis of the resulting body of data. (p. 89)

Johanson, Green, and Williams (1998) have found similar blunders in survey research. They point out errors and offer correction for many of them, including nonresponse and interpretation.

To prevent such problems, the survey approach should include (1) a clearly delineated research problem; (2) appropriate questions to respondents to gain information; (3) a well-systematized data collection technique; (4) a generation of group-level statistics; and (5) results that are generalizable to the larger population (Aday, 1989). These characteristics are very similar to those of other types of research. The major differences are in type of data collection—existing data, participant or nonparticipant observation, and case studies. Like other research, it is appropriate to tie survey research to theory (Russell & Van Gelder, 2008), which adds even greater credibility to a highly-used and valuable research approach even in medicine (McMahon & Fagerline, 2004).

Researchers who plan to use survey methodology should review existing health surveys to learn about good designs for small and large budgets and to become familiar with secondary analysis of health survey data sets. At the national level, the National Center for Health Statistics–National Health Interview Survey (NCHS–NHIS) or the Youth Risk Behavioral Survey conducted by the Centers for Disease Control and Prevention (CDC) could be reviewed. Similarly, a review of the Victoria Healthy Youth Survey, a population-based longitudinal survey of youth aged 14 to 21 in Victoria, British Columbia, could be beneficial (Nixon, Cloutier, & Jansson, 2008).

Survey Flow Plan

A **flow plan** is used to outline the design and subsequent implementation of a survey. It begins with the objectives of the survey, lists each step to be taken, and concludes with the final report. In short, the flow plan is an organizational device. Keep in mind that several decisions may have to be made at the same time. The major components of this process are:

1. *Planning the survey:* This section includes (a) survey objectives, (b) monetary resources, (c) time resources, and (d) personnel resources. This section should be completed in detail.
2. *Overall design:* A survey should be designed to match the objectives and be in concert with data needs, sample size requirements, data collection, resources, interviewer selection, data analysis, budget, and method of reporting the results.
3. *Method of data collection:* The method chosen should match the survey objectives and fit within resource constraints.
4. *Planning data analysis:* This step describes how the data are to be analyzed.

5. *Drawing the sample*: From the survey objectives and design come (a) the questionnaire population, (b) the sample size and selection, and (c) interviewers, when appropriate.
6. *Questionnaire construction*: The questions formulated for a survey are of the utmost importance and require detailed attention. All questions should match the objectives.
7. *Pretest questionnaire*: The survey should be pretested with a sample comparable to the intended population to be surveyed.
8. *Questionnaire revision*: Revisions should be based upon the findings from the pretest. If they are extensive, a second pretest should be conducted.
9. *Administering the survey*: The method chosen (e.g., regular mail, e-mail, telephone, personal interview, or even the Internet) should fit the nature of the data to be gathered and, of course, the objectives.
10. *Code preparation*: This is the initial step in data reduction. It is the translation of question responses and respondent information to specific categories for analysis. The coding should be consistent and conventional. A precoded questionnaire should be used if at all possible.
11. *Verification*: This is an important step to check for bias, particularly when interviewers are employed. According to Aday (1989), the two principal means of cleaning the data are range checking and contingency checking. As the name implies, range checking looks for the valid range of numbers/codes used for a particular answer. For example, if the codes are to be “1” or “2” and a “3” shows up, then a mistake has been made. Contingency checking cleans data for related questions. For example, if a question were to be skipped by certain respondents, this approach would check to be sure they were omitted from the survey. For example, the respondent’s age may be recorded as 12 years old but the educational level may be recorded as college graduate. For interviewers, the answers collected by one interviewer can be compared with those of others to see if bias is present. Some respondents could be reinterviewed. Also, the researcher could check with some outside criterion, such as whether the respondents really did what they said they would do (e.g., get an annual Pap exam). The errors should be located, and when appropriate, the responses should be sent back to field operations for corrections.
12. *Data entry*: This will vary according to resources, but in most instances data will be entered into a software database program on a computer. The key here is to use a program that is user-friendly, advantageous for analysis, and can be “watched” for errors.
13. *Tabulation*: Initially, a frequency count should be conducted to ascertain how many answers are in each of the categories for every question.
14. *Analysis*: This varies according to the purpose of the study, but it generally includes percentages, averages, relational indices, and tests of significance.
15. *Recording and reporting*: All the prior steps should be outlined in the report, with special emphasis on hypotheses, hypothesis testing, reliability of results, and implications of results for the subjects and further research.

Figure 6.1 summarizes the flow plan steps.



Figure 6.1 Survey Flow Plan

Survey Design

The nature of the research questions should determine the research design. The research questions should address the what, who, where, and when of the survey. On occasion, the researcher may wish to incorporate what findings are to be expected. Research design falls into two broad categories: descriptive and analytical.

In Elizabeth's case, one of her objectives is to determine the health needs of the bicounty residents. Let's suppose that an overview of hospital data shows that a high accident rate is one of the major reasons for emergency room visits and hospital admissions for children. However, she would like, as part of her survey, to find out more about injury and injury prevention. Table 6.1 illustrates the relationship between her research questions and descriptive designs. As shown, she may use any of three types of **descriptive designs**: cross-sectional, longitudinal, or group comparison. The primary difference among the three choices are time dimension and group focus. Cross-sectional occurs at one point in time, whereas longitudinal takes place over a period of time. Group comparison simply compares groups on the issue, in this case childhood injuries. Descriptive designs emphasize what characteristics the group or groups possess.

On the other hand, Elizabeth may want to make her design analytical rather than descriptive. Table 6.2 shows how three **analytical designs** relate to research questions. The nature of the design—cross-sectional, case control, or prospective—must fit the research question asked. Analytical designs go beyond description to address the relationship of the variable in question (here, injuries) to other factors or variables. Analytical designs, like experimental ones, explore why certain groups have particular characteristics.

In addition to these designs, there is the experimental study design. If using this design, Elizabeth would want to find out the impact her program has on injury reduction. Her research question might read as follows: "Is the incidence of injury less among children in the bicounty area during the summer break for those children who attended

Table 6.1 Research Questions for Descriptive Survey Designs

Elements of the Research Question	Descriptive Designs		
	Cross-Sectional	Longitudinal	Group Comparison
What	What is the prevalence of injuries	What is the incidence of injuries	Is there a difference in characteristics between those who suffer injuries and those who do not
Who	among children	among children	among children
Where	in the bicounty area	in the bicounty area	in the bicounty area
When	in the last month of summer (no school)?	during June, July, and August?	in the last month of summer (no school)?

Table 6.2 Research Questions for Analytical Survey Designs

Elements of the Research Question	Analytical Designs		
	Cross-Sectional	Case Control	Prospective
What	Are injured children more likely than non-injured children	Are injured children more likely than non-injured children	Is the incidence of injury greater
Who	among children	among children	among children
Where	in the bicounty area	in the bicounty area	in the bicounty area
When	in the last month of summer	in the last month of summer	during June, July, and August
Why	to be at home?	to have a history of injuries?	for those who attend summer camps?

the injury control program?” This implies pre- and postmeasurement or data collection. Obviously several variables would have to be controlled.

Data Collection Methods

The method of data collection should be linked with the objectives and research questions (including sample coverage). In addition, the availability of the method and the budget for the study impact methodology selection. If Elizabeth finds that migrant workers who speak little English come into the counties during the summer months to help on the farms, neither mail questionnaires nor telephone interviews would be good methods. Instead, personal interviews would be necessary. In contrast, in seeking very sensitive data about health issues, it would be best for her to use a mail questionnaire to ensure anonymity. Telephone interviewing is to her advantage if the questions are not too sensitive and her target audience is readily accessible. Of course personal and telephone interviews require a large number of well-trained personnel, which can be prohibitive to many programs. The price tag is of concern to most graduate students, also.

Issues to Consider in Mail Surveys

Whether to use a mail survey or an interview technique should be determined before sample selection and questionnaire design and construction are done. Consequently, Elizabeth should have made this decision some time ago in her survey of the two counties. This section will deal with some of the advantages, disadvantages, and factors involved with mail surveys.

Advantages of Mail Surveys Briefly, some of the advantages of a mailed questionnaire are (1) a savings of money and time, especially as compared with the interview

technique, (2) no interviewer bias, (3) greater assurance of anonymity, (4) completion by the respondent at his or her convenience, (5) accessibility to a wide geographic region, (6) accurate information because respondents can consult records before answering, and (7) identical wording for all respondents. In short, there are some definite pluses for using the mail survey technique. On the other hand, drawbacks do exist.

Disadvantages of Mail Surveys While a mailed questionnaire may appeal to Elizabeth, she should consider its disadvantages: (1) lack of flexibility, (2) likelihood of unanswered questions, (3) low response rate as compared to interviewing, (4) inability to record spontaneous reactions and/or nonverbal responses, (5) lack of control over the order in which questions are answered and over the immediate environment, (6) no guarantee of return by the deadline date, and (7) inability to use a complex questionnaire format.

Factors Influencing Mail Surveys Further consideration of whether or not to do a mail survey should include the seven variables enumerated by Sellitz, Jahoda, Deutsch, and Cook (1959), which affect the adequacy of data and the number of questionnaires returned.

One factor is sponsorship of the questionnaire. The organization or individuals involved may enhance or detract from the legitimacy of the study, thereby influencing questionnaire completion. A second variable is questionnaire color. It appears that color makes little or no difference in the rate of return. A third factor is questionnaire length; it seems that a less cluttered questionnaire, although longer, will bring a higher return rate than a shorter version. Ease of completion and return serves as a fourth factor. It is suggested that directions be explicit and that a stamped, self-addressed envelope be supplied. Incentives, which can range from money to a copy of the survey results, usually increase the response rate. However, monetary incentives should be seen as a goodwill gesture and not payment for time (O'Connor, Sharp, & Olson, 1999)! Moreover, incentives should be sent on the first mailing. The nature of the respondents also affects the number of questionnaires returned and the adequacy of data collected. If a highly select group is used, such as directors of mental health centers, responses tend to be more favorable than if the general public is used. Sellitz et al. (1959) also noted that the cover letter (discussed later) is of great importance.

The time and type of mailing, as well as the nature of the follow-up, also play a role in the success of mail surveys. As may be expected, first-class mail provides a greater return than any other class of mail. Concomitantly, using a hand-stamped envelope rather than a business-reply envelope may increase the return slightly. This difference seems to be shrinking, however, and with rising postal rates, the researcher should seriously consider the use of business-reply envelopes to save money.

Mailings must be well timed. Obviously major holidays should be avoided. Surveys received during the latter part of the week are more likely to receive quick response than those received early in the week. The months of February and April offer the lowest rate of return and March the highest, although for K-12 school surveys September may be the best month.

As a final note on factors influencing mailed questionnaires, follow-up letters or telephone calls should be employed. This should be standard procedure. The literature reveals that an increase of 20% can be expected with one follow-up or more. Frequently

the researcher may send out a reminder letter, followed by an additional questionnaire and letter, and then either another letter or a phone call.

A reasonable rate of return is a highly questionable topic because so many factors affect it. Some researchers hold that a 90% return rate is needed, others claim that 50 to 60% is permissible, and still others believe that a lower rate is acceptable depending on the target population. Whatever the response rate, it is important that there be a demonstrated lack of response bias. Of course, the greater the return, the less opportunity for response bias. The amount of time and money available to the researcher will dictate response rate too.

To help demonstrate a lack of bias in the data collected, the survey researcher should make every attempt to gain information about nonrespondents (Barriball & White, 1999). In some studies there may be little difference between respondents and nonrespondents; in others the difference may be large. A survey examining the attitudes of employees who refused to complete an employee survey found that noncompliers had a greater likelihood of quitting, less company loyalty, less satisfaction with supervisors, and more negative beliefs about how their employer handled the survey data (Rogelberg, Luong, Sederburg, & Cristol, 2000). While there are different strategies to address nonresponse (Barriball & White, 1999), Sully and Grant (1997) suggest using a "reasons for not responding" form. They found that this yielded responses that increased to the level obtained with traditional follow-up procedures and also collected demographic information on the nonrespondents.

In summary, the mail survey is a technique with much potential and several advantages. Nonetheless, drawbacks do exist, and many factors influence the rate of return and adequacy of response. The decision to use or not use a mail survey must be couched within the framework of the entire study and in particular the objectives of that study.

The Personal Interview as a Research Technique

Traditionally, interviews have dealt with an individual on a face-to-face basis. It may be the principal method of investigation in some studies, but in others it is more of an exploratory tool to acquire more information (e.g., a pilot study to develop a more extensive questionnaire). Less traditional is the group interview. While not appropriate for all occasions, it may be an excellent technique if the researcher is concerned with a behavior that takes place in a group interaction setting. The advantages of the group technique over the individual approach are (1) greater efficiency in time and money, (2) observation of group interaction patterns, (3) reflection of group behavior in results, and (4) productivity of others can be stimulated. On the other hand, the group approach may (1) intimidate and suppress responses, (2) promote conformity, (3) polarize opinions, and (4) be susceptible to manipulation by an influential group member (Isaac & Michael, 1981).

Advantages of Personal Interview Studies Some of the major advantages of the individual interview study are (1) personalization of the study to the participant; (2) flexibility so that further probing may occur or questions can be repeated; (3) a response rate that is usually higher than with a comparable mail survey; (4) observation of both verbal and nonverbal behavior; (5) control over question order that cannot be accomplished by a mail questionnaire; (6) spontaneity and "no help from others," as contrasted with either the mail survey or the group interview technique; (8) recording of the time of the

interview (this may be important if events affecting the object under study have occurred); and (9) ability to use more involved and complex questionnaires.

Disadvantages of Personal Interview Studies As with all research methodologies, the interview study has some inherent disadvantages. Some of these are (1) cost in terms of money and time (including training period and travel allowance); (2) openness to manipulation or interviewer bias; (3) vulnerability to personality clashes; (4) lack of anonymity; (5) inconvenience to the respondent as well as lack of opportunity to consult records (e.g., medical records for immunization or booster shots); (6) lack of standardization in questions because probing or question repetition may bring about a rewording that produces different responses from different respondents; (7) lack of access to respondents because of distance or other factors that may make the mail survey appear more desirable; and (8) difficulty in summarizing the findings.

Factors Influencing Personal Interview Studies As in the mail survey, several factors may influence the quality of the data received through the interview technique. The principal one, of course, is the effect of interviewer characteristics.

Berg (2001) noted that interviewee's perception of the interviewer plays a principal role not only in the decision to consent to interview but also in the presentation of biased information. Other researchers (Burns & Grove, 1993) have shown that, in addition to race and gender, style of dress, age, hairstyle, and speech mannerism all play a role.

Leenan et al. (2008) combined the personal interview technique with a mail technique by hand-delivering the cover letter, conducting the interview when possible, and collecting biophysical data for hypertension. This was a unique approach to gathering data using a cross-sectional design for three visible minority groups.

Interview Structure

Several interview structures may be employed by the researcher, all of which may be placed into one (or more) of three categories. All three will be presented briefly. The researcher should note that reliability increases with objectivity.

Unstructured interviews offer broad freedom to the respondent in terms of both response and time. This type of interview is usually reserved for obtaining information that is very personal and/or potentially threatening. As to be expected, this format is the most susceptible to subjective bias or error.

The middle road is the **semistructured interview**, which contains a core of structured questions from which the interviewer may move in related directions for in-depth probing. This can produce accurate information on certain questions with a built-in opportunity for exploration. Training is important so that the interviewer knows when and how to probe as well as how to avoid the introduction of interviewer bias.

In the **structured interview** a well-defined pattern is followed, similar to a questionnaire. The interviewer only strays from the pattern to clarify questions or allow for elaboration. The type of information sought through this technique must be factual and specific. The interview itself is usually brief.

Whatever structure is used, Berg (2001) has constructed commandments for interviewing; a synopsis is presented here. You should spend a few moments with small talk

so that you never enter an interview cold. All interviewers should practice a lot to become proficient. You should always remember your purpose and be respectful to the interviewees. While you need to think about your appearance, you should act naturally allowing questions to “pop up” rather than appear overrehearsed. Being a good listener is extremely important, as is being cordial and appreciative at the end of the interview. Of course, you should not accept single-syllable answers. Studies by Leenan et al. (2008) and Nixon et al. (2008) demonstrate how sensitive data can be handled.

Telephone Interviewing

The use of the telephone in interviewing has increased greatly over the past few years. Its chief advantage over face-to-face interviewing is cost savings. One study by Graves and Kahn (1979) estimates a savings of 50% by telephone, while others purport a reduction of 75 to 80% (Klecka & Tuchfarber, 1978; Taylor, Wilson, & Wakefield, 1998).

A second advantage is that telephone interviewing is much faster than either a mail survey or personal interview study. Third, the researcher can select subjects from a much broader area because travel is not involved. Fourth, the respondent remains more anonymous in a telephone interview. Fifth, monitoring and quality control are much easier in telephone interviews because all calls can be made from a central location. Sixth, if no one is home, frequent callbacks can be made with little expense, as contrasted to an interviewer having to return to the household or business. Seventh, the researcher has access to secure buildings and dangerous neighborhoods with the telephone interview. Finally, the telephone interview may be better than the face-to-face interview for collecting sensitive data.

However, the telephone interview does have some drawbacks. Generally, respondents are less motivated and often may see the interview as a hoax or cover for some ulterior motive. Of course, the use of checklists and visual aids are eliminated, as compared with the interview study. Further, the telephone interviewer has very little control over the situation; all the respondent has to do is hang up the telephone. However, Graves and Kahn (1979) in their study found that, while a 74% response was obtained in personal interviews, a 70% response was received through the telephone interview. Because the same questionnaire was employed for each method, the results could be compared easily. It was discovered that the results were very similar over a wide range of topics.

Conklin (1997, 1999) has done extensive work on telephone survey use among colleges and universities, particularly as it relates to mail surveys and nonrespondents. She found little difference between the two approaches. Similarly, Fowler, Gallagher, and Nederend (1999), when comparing telephone and mail responses, found data collection method had little bearing on the key results.

The telephone is productive for qualitative interviews, especially if used as a follow-up to face-to-face interviews (Rubin & Rubin, 1997). To accomplish a qualitative telephone interview, Berg (2001) recommends following three steps. First, the interviewer must establish legitimacy; second, the interviewee must be convinced that it is important to participate in this research activity; and third, detailed information must be collected if it is to be of value to the investigation.

A potential problem with telephone interviews is that of unlisted numbers. Random digit dialing (RDD), however, has allowed the researcher to circumvent this dilemma. The researcher simply selects four-digit numbers from a table of random numbers and

adjoins them to the prefix (first three numbers) and then dials that number. This makes unlisted numbers available. Another problem, today, is that increasing numbers of people are abandoning their land-based lines for cell phones, making it more difficult to reach them.

Computer-assisted telephone interviewing (CATI) has been around for quite some time. This method allows the questionnaire to be displayed on a screen, and the interviewers simply ask the question and enter the answer via the keyboard or mouse. Using CATI can virtually eliminate the recording of data in the wrong place and the incorrect asking of questions. For example, if the question, "Do you smoke?" arises in a health interview and the answer is affirmative, the interviewer may be required to turn two pages to a set of questions about smoking (e.g., about inhaling or frequency). If the interviewer fails to do this, that data may be lost and/or the respondent may be asked other questions that could be embarrassing and shorten the interview. However, a personal computer can be programmed so that an answer automatically moves the interviewer to the next appropriate question. Response accuracy is increased, especially because the interviewer does not even see the questions that do not have to be asked. Other advantages are the detection of "wild" or meaningless codes, the production of error-free data ready for analysis, and the use of help menus. In order to achieve these advantages, the researcher needs to have a computer that will perform RDD, store and retrieve telephone lists, dial automatically, present questions, check codes, input to a data set, and perform interviewing management.

While CATI is a widely recognized approach, computer-assisted personal interviewing (CAPI) is new by comparison. Lightweight laptop computers with large memories have made this technique much more user-friendly. The interviewer brings the questionnaire to the screen and enters data for each answer, similar to the CATI approach. The U. S. Census Bureau has been experimenting with CAPI. The U. S. Department of Agriculture used laptop computers over a decade ago in their Nationwide Food Consumption Survey. The National Household Education Survey (National Center for Education Statistics, 1997) is a national study on educational issues that uses both RDD and CATI.

The use of a computerized self-administered questionnaire (CSAQ) requires the respondents to have access to a computer. Respondents are given instructions on how to access the questionnaire and steps involved in completing it. The World Wide Web now has several computerized self-administered questionnaires on many topics. As computers become more accessible and software more usable, it is likely that all forms of data collection will be computer-assisted or computer-driven, as exemplified by Greene, Speizer, and Wiitala (2008) in their mixed mode approach.

The Computer Revolution: Web-Based Surveys, Including Electronic Mail

The Internet has dramatically changed how we communicate with one another. Electronic mail (e-mail), websites, listserves, chatrooms, and newsgroups allow researchers almost instant communication. Several studies have shown the promise of the web-based survey (Fricker, Galesic, Tourangeau, & Yan, 2005; Fricker & Schonlan, 2002).

Nesbary (2000) outlined three types of **Internet surveys**. The first is the e-mail survey, which is the oldest form of Internet survey and is very similar to regular mail surveys, since questionnaires are employed. The principal difference, of course, is the speed and nature of transmission. Another difference may be that e-mail is a better way to obtain reflective data. Heflich and Rice (1999) found that using a semistructured interview protocol via e-mail produced reflective dialogues with deep, qualitative data. Furlong (1997) believes e-mail, when compared to postal mail, is faster, increases the likelihood of the recipient getting the survey, and may encourage participants to respond more rapidly. Research by Meha and Sivadas (1995) supports this contention. However, Mavis (1998) found postal surveys to be superior to e-mail surveys with regard to response rate. The decision to use e-mail, at least in the next few years, must be situation-specific (Hughes & Pakieser, 1999). In other words, the researcher must consider survey cost, convenience, timelines, response rate, and most importantly, representative sampling.

The second form of Internet survey discussed by Nesbary is the disk-based survey. The questionnaire is placed on a compact disc (CD) in either a word-processing format or an executable (.exe) format. A respondent simply opens the questionnaire in the appropriate format and fills in the blanks. A database format (Access, Paradox) can be used in the executable form, allowing the researcher great flexibility in developing and answering questions. Of course, both formats could be sent by e-mail as attachments instead of being sent as CDs.

The third type of Internet survey is the forms-based survey. This survey instrument is located on the researcher's website. The researcher can contact respondents by e-mail to alert them to the website and ask them to respond. Of course, depending upon the nature of the research, only people who frequent the website can be asked to respond. For example, a cancer site may ask all people who visit to complete the questionnaire. This technique allows for the flexibility of a database format and the accuracy of a written survey. Needless to say, it can be a cost saving in terms of both time and money.

Pealer and Weiler (2003) suggest an eight-step approach to web surveys which are modified slightly from steps necessary for most survey techniques. Kittleson (2003) encourages all researchers using electronic techniques to become very familiar with the information technology (IT) portion of web-based surveys in order to obtain what is wanted. While the response rates may vary in web-based surveys, it has been found that it is dependent upon the target group (Daley, McDermott, McCormack-Brown, & Kittleson, 2003). Greene et al. (2008) used a mixed mode survey combining the telephone with the Web, finding that this approach can significantly increase response rates over single-mode surveys.

Response and Other Rates

When reading about **response rates** in survey research, it is important to identify how they were determined (Frey, 1989). The more common way is comparing the number of completions to the number of potential respondents who were eligible. This formula is as follows:

$$\text{response rate } 1 = \frac{\text{number of completions}}{\text{number in sample}} \times 100$$

A second response rate that is sometimes used compares the number of completions to the completions plus partial completions and the number of refusals, less all uncompleted interviews. This formula is expressed as:

$$\text{response rate 2} = \frac{\text{number of completions}}{\text{number in sample} - (\text{noneligible and nonreachable})} \times 100$$

Response Rate 1 is the preferred method and should be used by Elizabeth in her survey research. It establishes a much more honest evaluation of returns. As a general rule, response rates for personal interviews are highest, followed by telephone and self-administered questionnaires, especially those done by mail. This should not be too surprising since interviewers can be persuasive and respondents are much less likely to end a face-to-face interview.

Refusal rates are of importance in survey research because refusals are the most common reason for nonresponses. This rate is calculated as follows:

$$\text{refusal rate} = \frac{\text{number of respondents refused}}{\text{number of eligible respondents contacted}} \times 100$$

The **noncontact rate** is simply the ratio of nonresponse not attributed to direct refusals from the potential respondents. This rate is important since it lets the researcher know the degree to which respondents are accessible or can be located. The formula is:

$$\text{noncontact rate} = \frac{\text{total not contacted}}{\text{total known eligible}} \times 100$$

In reporting your research, you should show that the loss of data from nonrespondents is not detrimental to the findings. This might be done by showing that no obvious differences exist among respondents and nonrespondents in such factors as age, race, gender, socioeconomic status, education, and so forth.

Survey Sampling

The results of survey research, in fact perhaps the results of all research, rest heavily on the sampling foundation. That is, if our sampling is flawed, then our results will not be as useful. Ideally, the researcher would like to observe the entire population to add more weight to the findings. For example, in Elizabeth's situation she may wish to obtain answers from all residents in the two-county area. However, limitations of resources, time, and money frequently preclude a study of the entire population. This would be even more evident if the study were to survey her entire state. Subsequently, a subset of some predetermined size must be selected from the population of interest. The sample or subset should represent the total population so that the data collected from the sample will be as accurate as that from the entire population.

The logic involved is simple. Nonetheless, the importance of this step cannot be overemphasized. As stated by Leedy (1980), "The results of a survey are no more trustworthy than the quality of the population or the representatives of the sample" (p. 35). A knowledgeable researcher commences with a population and works down to the

sample. In other words, the population of interest is designated and then a sample is derived. The neophyte, on the other hand, often works from the bottom up by attempting to ascertain the minimum number of respondents needed for a successful study. The inherent problem with this approach is that it is next to impossible to assess the representatives of the sample because the entire population has not been identified. In the case study, Elizabeth must determine her target population. Is it to be all residents? Or does she just wish to include taxpayers? Her population could be limited to all those who may potentially use the services of the hospital. Once this decision is made, she may then proceed to the selection of the sample.

The savings in both time and money are obvious reasons to deal with a sample of the population. There are additional advantages, too, as described by Bailey (1994). The sample may achieve a greater response rate owing to greater cooperation than might occur in the full population survey. This in itself would tend to make the results more accurate. In health surveys with sensitive items, this point is particularly important. Concomitantly, the researcher can keep a low profile by using a sample. That is, less people may be offended, thereby negating an opportunity for several people to organize a common resistance. In the case of interviewing, a sample reduces the number of interviews and interviewers. This is beneficial in that supervision of an enormous number of interviewers is difficult at best, and necessary attention to details such as follow-ups becomes very cumbersome as numbers increase.

There is little doubt that in many instances the use of a sample is much more advantageous than an entire population survey. Yet, the benefits only hold true if the sample is drawn with precision.

The National Health Interview Survey (1999) was redesigned to improve the reliability of the statistics for racial, ethnic, economic, and geographic domains. Patrick, Pruchno, and Rose (1998), in comparing five recruitment strategies, found that even nonprobability sampling can be very successful for recruiting large, diverse groups if very careful planning, implementation, and good funding exist. Wang and Fan (1997) discuss six criteria for survey sampling design evaluation. Four of the six involve sampling procedures. The four criteria are (1) clearly specifying the population, (2) explicitly stating the unit of analysis, (3) specifying a method to determine sample size, and (4) giving a detailed description of selection procedures. The next chapter addresses sampling techniques and sample size in detail.

Questionnaire Design and Construction

Questionnaire design and construction involve much more than drafting the questionnaire itself. The researcher needs to complete prequestionnaire planning, draft the questionnaire, prepare the final copy, and then pretest it. The following points illustrate the necessary components:

1. Prequestionnaire planning:
 - a. Define the problem, and hypothesize solutions.
 - b. Determine the information needed to test the hypothesis.
 - c. Review previous research, and speak with resource personnel.
 - d. Develop preliminary questions.

2. Drafting the questionnaire:
 - a. Considerations for researcher, respondent, and interviewer
 - b. Types of questionnaires
 - c. Types of questions
 - d. First draft of the questionnaire
 3. Preparing the final questionnaire:
 - a. Physical layout
 - b. Reproduction and materials
 - c. Identification of respondents in the questionnaire
 4. Pretesting:
 - a. Determining the value to questionnaire design
 - b. Evaluating the pretest
 - c. Revising the questionnaire, if necessary (which may warrant another pretest)
- The emphasis in this section will be on drafting the questionnaire.

Researcher, Respondent, and Interviewer Considerations

Once the hypotheses have been carefully specified and the sample drawn, the next step in the research chain is development of the data collection instrument. Herein, the major consideration is questionnaire relevance to the researcher, the respondent, and the interviewer (when appropriate). The researcher must be certain that the questionnaire is relevant in the goals and objectives of the study as set. No matter how well worded or designed the questionnaire, if it fails to produce the data relevant to the objectives, it is worthless.

Overall, the study and subsequent questionnaire must be relevant to the respondent. This is not always self-evident because research objectives are frequently housed in scientific jargon; therefore, they must be clarified and justified in lay terms. This can be accomplished by means of a cover letter. There must be a connection between the respondent and all those questions that apply to the respondent. That is, can the respondents understand the questions? Are they likely to know the answers? Are they willing to respond to the questions? The answer to the last question could be “no” if the questions are not relevant and thereby offer no motivating force. To make the questions applicable, skips or contingency questions (e.g., “If you answered ‘yes’ to this question, skip to Question 22”) can be used. With this method the respondent has to read and answer only items that are personally relevant.

In drafting the questionnaire that is to be used in an interview schedule, all elements that could lead to interviewer bias must be removed. Questions have to be phrased so as not to be misconstrued. Further, questions should follow a logical order, with a smooth transition from topic to topic. Needless to say, all directions to the interviewer and the respondent should be clear and concise.

Types of Questionnaires

Generally, questionnaire forms are closed, open, or a combination of the two. The **restricted, or closed, form** provides fixed-alternative questions that can be answered by a

simple “yes” or “no” or by checking an appropriate box. Some of the advantages of this form are (1) ease of completion for the respondent; (2) simplification of coding and analysis, particularly because the questionnaire can be precoded; (3) greater chance that respondents will answer sensitive questions (e.g., about age or income) because they are usually categorized rather than demanding an exact number; and (4) a minimum of irrelevant responses.

On the other hand, some disadvantages of the closed form are (1) given a list of potential answers, the unknowledgeable respondent may guess or randomly select an answer; (2) variations in answers among respondents may be reduced since only certain categories are available; (3) there may be too many answer categories to be printed simplistically; (4) the respondent may become frustrated since there is no room for a separate, nonprovided opinion; and (5) the respondent may circle the wrong answer (e.g., circle a 3 instead of a 4).

It is suggested that the categories of “don’t know” and “other” be included in a closed-form questionnaire. In this way, the respondent is not forced to work with just the alternatives provided. Further, it gives the researcher the opportunity to receive more relevant information.

In the **unrestricted** or **open** questionnaire form, the response categories are not specified, and the respondent is allowed to answer in his or her own words. Some of the advantages of this type of questionnaire are that it is (1) usable when all the response categories are unknown, (2) preferable for controversial, sensitive, and complex issues, and (3) allows for respondent creativity, clarification, and detail. The disadvantages include (1) difficulty in coding and analysis; (2) greater demands on the respondent in terms of time, writing ability, and thought; (3) questions may be too general for the respondent to comprehend or answer; and (4) data collected may not be relevant to the objectives of the study.

Many questionnaires have *combined* forms, including both closed and open items. A questionnaire that is primarily closed should have at least one open-ended item to allow the respondent to express a personal opinion or thought. Each health science researcher must decide which type is more likely to supply the information desired.

Types of Questions

A variety of questions and response category formats are available to the questionnaire builder. However, no matter which format is used, one of the most important factors is that the respondents understand the questions (Johnson & Fendrich, 2005). In closed-form questionnaires the usual types are dichotomous, multiple choice, rating, and ranking. Open-form questionnaires generally consist of a blank space where the answer is to be written. On occasion, sentence-completion questions are incorporated into both forms.

To illustrate the different types, suppose that Elizabeth in our case study were to develop a questionnaire with primarily a closed format and a few open-ended items. The basic rule she would follow for writing questions would be to provide all possible answers in as clear a fashion as possible.

In **dichotomous questions**, the answer comprises two parts, one of which is to be selected by the respondents. Examples of this type are:

Circle the appropriate number:

- 1. Gender: Male 1 Female 2
- 2. Setting: Urban 1 Rural 2

With **multiple-choice** questions, each potential answer is listed for the respondent such as:

How far would you be willing to travel for a health program of interest to you?
(Check One)

- 1 mile or less []
- 2 to 4 miles []
- 5 to 7 miles []
- 8 to 10 miles []
- 11 or more miles []

It can be readily noted that this format takes much more space than if answers are placed side by side; however, the answers are recognized quickly.

Many questionnaires include **rating** questions in which the respondent indicates a particular view about the psychological object. For example:

Several educational services are listed below. Please indicate the importance of each to you:

	Very Important	Somewhat Important	Not Important
1. Diabetic training	_____	_____	_____
2. Smoking cessation	_____	_____	_____
3. Stress reduction	_____	_____	_____

This format allows several items to be categorized as a series with directions stated only once.

Another fixed-alternative approach is that of **ranking**. Here the respondent simply orders the given answers in rank. For example:

- 1. The following are some of the health problems faced by residents of our two counties. Please place them in order of greatest problem (rank it 1) to the smallest problem (rank it 5) within the counties, as you see them:
- _____ Unintentional injury
- _____ Alcoholism
- _____ Drug addiction
- _____ Teen pregnancy
- _____ High blood pressure

For the **sentence-completion** questions, an example item would be:

1. In regard to education, I feel the Community Hospital should
-

Frequently, the **open-ended** questions are placed at or near the end of the questionnaire. For example, Elizabeth might ask the following question:

1. In the space below, please write out any particular interests or concerns that you have about attending programs at the Community Hospital:
-

Branching questions need to be handled with care so that the correct target group responds to the questions. An example for Elizabeth could be as follows:

In the past month, have you or a member of your family been injured?

(1) Yes (2) No

If yes: Which of the following was done?	Yes	No
1. Took care of the problem at home	1	2
2. Visited our family physician	1	2
3. Went to the emergency room	1	2
4. Was admitted to the hospital	1	2
5. Other; please name _____	1	2

Or, this could be written in the following way.

In the past month, have you or a member of your family been injured?

1. Yes (complete section A)
2. No (go to section B)

It is important that the health science researcher consider the target population when deciding on the type of questionnaire and the types of questions. Further, when writing the questions, several pitfalls should be avoided. The following list can serve as a guide:

1. Phrase questions to be comprehended by all those in the target population.
2. Avoid double-barreled questions (two questions in one).
3. Be careful of double negatives.
4. Define terms that could be easily misinterpreted.
5. Underline or boldface a word if special emphasis is demanded.
6. Watch for inadequate alternatives to a question.
7. Do not use adjectives that fail to have an agreed-upon meaning.
8. Be sure questions are not leading questions.
9. There should be no ambiguity in the questions.

On completion of the questions, it is necessary to combine them into the final questionnaire in an order that will bring about the greatest response. To assist in this task, some general rules should be followed:

1. Put sensitive questions as well as open-ended ones near the end of the questionnaire.
2. Place questions in a logical order when possible.
3. Simpler questions should be ahead of more difficult ones.
4. Avoid establishing a response set.
5. Request information needed for subsequent questions first.
6. Vary questions by length and type.
7. Separate reliability-check question pairs (i.e., pairs of questions in which one is stated positively and the other negatively). For example, if one of your questions is, "The Mental Health Center should offer educational classes (agree/disagree)," you should wait until later in the questionnaire to include the item, "the Mental Health Center should *not* offer educational classes (agree/disagree)."

Attitude Scale Construction

Likert Scaling

Scales may be included in a questionnaire to gather data in a different fashion. One type of scale was devised by Rensis Likert (1932), who employed ordinal scaling and summated rating techniques to develop an attitude scale. In *summated rating*, several items are used in an attitude scale, and, to ascertain an individual's score, the researcher adds (sums up) each item score circled. For example, if the attitude scale comprises 10 items and the person circles 1 for every item, the summated rating score would be 10, whereas the summated rating score of 50 would be obtained by the person who circles 5 for each of the 10 items.

One of the major problems of summated rating is that all the items may fail to measure the same concept. Perhaps some of the items in the 10-item scale just mentioned really do not measure what the others measure. Likert created a technique to eliminate such items and thereby improve internal consistency. This will be discussed in more detail later.

In constructing a Likert-type scale, the first step is to assemble a large number of items considered relevant to the attitude under investigation. These items or statements should fall in approximately equal numbers with respect to their relative favorableness or unfavorableness toward the object of interest.

Next, each statement must be weighted from 1 to 5, with 3 as the neutral position. It makes no difference whether a rank of 1 is the favorable or unfavorable end of the continuum, as long as the weighting is consistent. For purposes of illustration, however,

let us allow the higher score to indicate a stronger agreement with the attitude being scaled. Therefore, a positive statement would be scored by the following key:

strongly agree	agree	undecided	disagree	strongly disagree
5	4	3	2	1

A negative statement would be scored as follows:

strongly agree	agree	undecided	disagree	strongly disagree
1	2	3	4	5

The reason for reversing the negative items or statements is to provide a total score that reflects positiveness toward the object in question. In a scale of attitudes toward sexuality, for example, the program participants with a positive sexual attitude would agree with positive statements and disagree with negative ones, while patients with an unfavorable (negative) sexual attitude would disagree with positive items and agree with negative ones. To obtain a total score, the relative weightings of each response are summed. Subsequently, in this example the higher total score represents a more positive sexual attitude than a lower total score.

Now, all of these items should be administered to a sample of the population for whom the scale is intended. As a general rule, the sample size should be at least twice the number of statements desired in the final scale, and most Likert-type scales have 20 to 22 items.

Selection of statements for the final scale is based on an objective check for internal consistency and statement differentiation between the highest and lowest scores. If selection is being done by hand, the scores can be separated into quartiles so the upper 25% can be compared with the lower 25%. The median score for each statement is calculated. If any statement has the same median score for both the high and low groups, it should be eliminated from the scale. For best differentiation, only those statements that have widely different median scores for the highest and lowest groups are retained. An easier way is to use a computer and check for internal consistency by an item-total correlation. Herein, the responses for *each* statement are correlated with the *total* scores obtained by the subjects on the whole test. This technique reveals the amount of agreement between each individual item and the total test—that is, the degree to which each item measures what the total test measures. Those statements receiving a low correlation should be eliminated.

As a final note, retain items with a high correlation and be certain to have approximately an equal number of positive and negative statements covering a range of topics within the attitude being measured.

Some of the advantages of Likert scaling are (1) it is simple to construct; (2) each item is of equal value so that respondents rather than the item are scored (unlike Thurstone scaling, discussed under *Interval-Ratio Scaling*); (3) it permits the use of latent attitudes in that items that are not manifestly related to the attitude being measured can be employed; and (4) it is likely to produce a highly reliable scale.

Van Alphen, Halfens, Hasman, and Imbos (1994) suggest that scoring can be improved by using the Rasch model. In that model both respondents and items are scaled on the same continuum, whereas in the Likert scaling all items have the same weight. The Rasch model is based on item response theory, and the Likert model is from classical test theory. The Thurstone and Chave (1929) method (see below) is another way to weight the items.

Interval-Ratio Scaling

Another method of quantification is the interval scale. Based on equal units of measurement, the interval scale demonstrates how much of a given characteristic is present. Further, the difference in amount of the characteristic present in the persons with scores of 25 and 26, for example, is assumed to be equivalent to the difference in persons with scores of 75 and 76. In other words, interval scales indicate the relative amount of a trait, which is something that neither nominal nor ordinal scales permit. However, because a true zero fails to exist in ordinal scales, it is inappropriate to claim that a score of 75 is three times a score of 25.

In contrast, a ratio scale possesses not only the equal interval properties of an interval scale, but also a true zero. For example, a zero on a pharmacy weight scale shows the complete absence of ounces or grams. Moreover, a ratio scale has the properties of real numbers, which can be added, subtracted, multiplied, divided, and expressed in ratio relationships. Therefore, a pharmacy scale that measures 10 grams has twice as much weight on it as one that measures 5 grams.

The most precise ratio scales are common in the physical health sciences (e.g., those for measuring blood pressure). The behavioral health sciences, which measure characteristics such as attitudes, are limited to interval scales or even less precise types, such as ordinal or nominal scales.

Thurstone and Chave (1929) developed the Thurstone technique of scaled values, frequently referred to as the method of equal-appearing intervals. Herein, a large number of attitude statements are collected and entered on separate cards. The statements, via the cards, are then submitted to a panel of judges, who are required to rank the statements into 11 equidistant piles. The 11 piles represent a continuum from extremely favorable to extremely unfavorable, with the middle pile being neutral.

The scale value for any particular statement is the median of the frequency distribution by the judges. If more than one statement has the same scale value, then the statement with the smaller Q statistic or interquartile range is retained for the final scale. It is believed that the smaller the interquartile range, the greater the degree of agreement among the judges with respect to that particular statement's position along the continuum.

The general goal is to obtain 20 to 22 statements to form the final scale. The 20 or 22 statements are then randomly arranged and administered to a group of respondents, who can either agree or disagree with each item. To score, the mean or median value of those statements with which the respondent agreed is calculated. In other words, the respondent is asked to only check those items to which he or she agrees, and the mean or median value of the checked statements is the scale score for that respondent.

A partial Thurstone-type attitude scale may appear thus:

Sexual Attitude Statements for Program Participants

1. Sexual relations are simply sources of frustration. (10.0)*
2. Hoping for a sexual relationship is senseless. (8.2)
3. Sex may or may not be important in a relationship. (5.0)
4. Partners should try new and different sexual approaches. (2.5)
5. Sex is the most essential part of a relationship. (1.0)

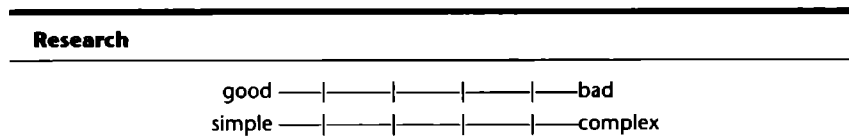
Some of the statements are unfavorable toward sexuality (Statements 1 and 2), while others are favorable (Statements 4 and 5) and still others neutral (Statement 3).

One of the advantages of the Thurstone scale is that the statements are weighted or valued rather than the respondents. Nevertheless, the Thurstone scale is disadvantageous in that it is much more cumbersome to construct than the Likert scale. Further, there is the question about the dependence of the scale values upon the opinions of the judges (Murphy & Likert, 1937). Moreover, it has been shown that this method is no more reliable than the Likert technique (Ferguson, 1941).

Some Thurstone-type scales have been produced for current use in the health science field. One such instrument is a scale to appraise the attitudes of college students toward euthanasia (Tordella & Neutens, 1979). Another scale measures information priorities among health consumers seeking cancer care (Bilodeau & Degner, 1996; Degner, Davison, Sloan, & Mueller, 1998). Degner et al. (1998) explored the use of both the Likert and Thurstone approaches.

Semantic Differential Scaling

The semantic differential scale was developed by Osgood, Suci, and Tannenbaum (1957) to measure attitudes. The semantic differential (SD) has three elements: (1) the attitudinal concept to be measured; (2) a pair of opposite adjectives; and (3) a series of undefined scale positions, usually seven in number, between each of the polar adjective pairs. For example, for the concept of research, the polar adjective pairs forming the opposite ends of the seven categories could be “good-bad” or “simple-complex.” If this were the case, then the SD might look as follows:



The polar adjective pairs should be selected according to the objectives of the study. In addition to the polar adjective pairs developed by the originators of the SD, Jenkins, Russell, and Suci (1958) created an atlas of semantic profiles for 360 words. In both

*This is the scaled value for the statement. The participant checks the statements with which he or she agrees, and the scale values of the statements checked are used to ascertain a scale score by finding the average or median score.

instances, Osgood et al. and Jenkins et al., the pairs of polar adjectives can be used to measure three dimensions—evaluative (e.g., good-bad), potency (e.g., hard-soft), and activity (e.g., fast-slow).

Heise (1970) has researched question format to find that it makes no difference whether one concept is followed by all adjective-pair scales, one concept is followed by one pair only, or all concepts are rated on one polar adjective pair. Further, if one concept is followed by all adjective pair scales, the ordering of concepts makes no difference. However, it is recommended that evaluative, potency, and activity scales be mixed or combined to prevent response sets. Concomitantly, the polar adjective pairs should be randomly arranged so that left and right positions on the total scale do not encourage a response set.

Sorokin (1976) has developed a SD to measure attitudes toward aspects of sexuality. Two of the concepts in his scale are shown for illustration:

Sexuality Scales

Concept 9: Getting Married

pleasing —|—|—|—|—|— annoying
 constructive —|—|—|—|—|— destructive
 desirable —|—|—|—|—|— undesirable

Concept 10: Having Sexual Relations With the One You Love

pleasing —|—|—|—|—|— annoying
 constructive —|—|—|—|—|— destructive
 desirable —|—|—|—|—|— undesirable

Different polar adjective pairs could also have been selected, such as:

Concept 10: Having Sexual Relations With the One You Love

important —|—|—|—|—|— unimportant
 constrained —|—|—|—|—|— free
 active —|—|—|—|—|— passive
 approach —|—|—|—|—|— avoid

Here three principal factors or dimensions as well as a fourth situational dimension are represented. “Important-unimportant” represents the evaluative factor, and “constrained-free” and “active-passive” reflect the potency and activity dimensions, respectively. The situational component is rendered by the adjective pair “approach-avoid.”

In administering and scoring the SD, the patients or subjects should be instructed to put down their initial impression. The scale positions are converted to numerical values so that various statistical assessments can be completed:

important —7 | 6 | 5 | 4 | 3 | 2 | 1— unimportant

The scores can be analyzed for differences between concepts, between scales, between subjects, or any such combination. Subsequently, a semantic differential generates a large amount of data. For information on analysis techniques, consult Kerlinger (1964), Nunnally (1962), Snider and Osgood (1969), and Osgood et al. (1957).

Two advantages of the SD are that it is simple to construct and easy for the respondent to complete. Further, it allows for several types of analyses to take place. However, that also serves as a disadvantage because some analyses can be very complex. If the researcher employs this method of attitude scale construction, it is important to outline the scoring technique in detail for the practitioner.

Cover Letter

When the questionnaire is complete, a cover letter should be developed (Fanning, 2005). Though the instrument constructed is of prime importance, the beginning researcher should realize that an inappropriate cover letter or introductory statement may cause the respondent to discard the questionnaire without even looking at it or to ask the interviewer to leave. As a matter of note, the letter should be on an organizational letterhead to indicate the legitimacy of the survey in the case of a mailed questionnaire. When an interview is conducted, the introductory statement serves as a public relations technique.

In regard to content, the cover letter or introductory statement should contain (1) identification of the person or organization conducting the study, (2) the reason the study is being conducted, (3) why it is important for the respondent to complete the survey, (4) assurance that there are no right or wrong answers, (5) assurance of confidentiality of information given, (6) assurance of anonymity of the respondent, when applicable, (7) length of time it will take to complete, (8) the date of return (for a mail survey), and (9) a notice of how to obtain results. Figure 6.2 shows the cover letter used by Elizabeth.

Scott (1961) reviewed several studies about cover letters and found that a “permissive” letter obtains a greater response than does a “firm” letter. Also, a “short, punchy letter” is better than a longer, logical appeal. Handwriting the address does not seem to increase the response rate. No difference in response rate was found whether a true signature or facsimile was used or whether the letter was addressed as “Dear Mr. Smith,” “Dear Friend,” or “Dear Bulletin User.” In short, he concluded that “the content of the letter is very much more important than its trappings” (p. 152). However, Wiersma (2000) has recommended that the letter be on official letterhead and be signed by a person in a professional position who is in some way associated with the respondents.

Pretesting and Questionnaire Revision

The constructed questionnaire remains at rough draft stage until a pretest is done to identify flaws and to allow for corrections. Though the sample for the pretest is frequently fellow students, faculty, or coworkers, it is recommended that a subsample for

[Date]

Dear [Respondent Name]:

As you are aware, the Community Hospital is responsible for many health services in our bicounty area. We are conducting this survey to determine the need for selected health services and health education programs.

Enclosed is a questionnaire that we are asking you to complete as part of this survey. The questions are very easy to answer and should not take more than 20 minutes of your time. There are no right or wrong answers. You have our assurance that the information that you provide in this survey will be kept anonymous. Your answers will help us to plan health services and education throughout the bicounty region and to promote better health for all of our residents.

Since a limited number of these questionnaires are being sent out to select residents of the bicounty area, your individual opinion is highly important to the success of this undertaking. We, therefore, request that you please complete and return this questionnaire in the enclosed, self-addressed, stamped envelope *no later than [insert date]*. If you have any questions about this survey or want to have a copy of the results, please contact Elizabeth _____ at 555-9306.

Thank you for your cooperation.

Sincerely,

Jane Doe
Executive Director
Community Hospital

Figure 6.2 Cover Letter

the target population be employed for better results. Further, it should be administered in the same fashion as intended for the actual study (i.e., mail, telephone, or interview).

Those in the pretest sample should complete the questionnaire as directed and then do a critical analysis of all aspects of the instrument: sensitivity of issues, question wording and order, response categories, reliability checks, physical layout, length of time for answering, and instructions. Any comments given in the margins or elsewhere should receive special attention, particularly if several respondents hold the same view. In addition, the researcher should seek indicators of other problems by calculating the “no response” or “don’t know” answers. Pattern of response should be observed for set responses.

Only when the corrections have been made should the questionnaire be used in the research project. If several alterations were required, another pretest should be conducted. All of this takes time and should be built into the time frame for the entire study. Omission of this step could prove to be a grievous error if the final data fail to correspond to the objectives.

Focus Groups

Focus groups, as an interview methodology, are becoming popular. This technique capitalizes on communication among the participants and the researcher. Rather than ask each person in the group to respond to the same question, although that could be done, the researcher encourages the participants to talk to each other—ask questions, give examples, provide comments. Focus groups arrive at opinions and can get to the reasons underlying those opinions. One of the highlights of the focus group technique is that it is culturally sensitive. It has been used extensively in cross-cultural research and in studying differential use of health services within a population (Kitzinger, 1995; Wilson, Pittman, & Wold, 2000). Critical comments can be generated with greater ease with this approach than with interviews (Taylor et al., 1999; Watts & Ebbutt, 1987). This is particularly important in evaluation studies and in marketing research.

The number of groups used in a singular study may range from 6 to 50, depending upon the nature of the research project and that old nemesis, available resources. Usually just a few groups are used. Sampling for focus groups can follow patterns similar to other types of research. That is, random sampling can be done. Depending upon the study objectives, imaginative sampling may be required. For example, Kitzinger (1990) found that it was important to include lesbians and women who were sexually abused in a study of women's experience with cervical smears. Focus groups can also be advantageous when sampling people who cannot read or write.

The number of participants per group should be between four and eight, and the sessions generally should last 2 hours, although they may be shorter or longer in duration. The sessions should be relaxed and the participants should be told that you are seeking interaction, not a question-answer format. Initially the researcher may sit back and observe but at some point will likely intervene to direct and guide the process. To stimulate discussion, you may want to have a series of large cards, each with a statement. The participants could place the cards in order of most important to least important, or they could separate the cards into "agree" and "disagree" piles. For example, Elizabeth could have cards dealing with quality of care at the hospital or the role of community health education carried out by the hospital.

Sim (1998) raises three issues in analyzing focus group data. First, it is difficult to infer attitudinal consensus. Second, measuring strength of opinion is very problematic. Third, attempts to generalize from focus group data can be met with methodological and epistemological objections. Therefore, analysis is generally qualitative in nature. Elizabeth, for example, could pull together and compare themes and how they relate to variables within the sample and population. Differentiation between group consensus and individual statements is important. Percentages and similar frequency data are not used. Deviant data (comments, themes) should be noted because they can be valuable to the findings. Focus groups can be a stand-alone approach or can be combined with other survey techniques.

The Delphi Technique

The Delphi technique designed by Helmer (1967) is a method of reaching group consensus on any psychological object. It was originated to circumvent the traditional roundtable approach of group consensus with its inherent problems of the power of

individuals to sway the group, the bandwagon effect of majority opinion, the manipulation of group dynamics, and the unwillingness of individuals to alter publicly stated positions.

While the Delphi technique is not new (Dalkey & Helmer, 1963), it is relatively recent to health science research. Two very different examples are those of Sowell (2000) and Annemans, Glaccone, and Vergnenegre (1999). The technique was used by Sowell to identify HIV/AIDS research priorities from the perspective of nurses in AIDS care. Annemans et al. employed a modified technique with clinicians to investigate the cost effectiveness of Taxol, an anticancer drug. Anderson, Goddard, Garcia, Guzman, and Vazquez (1998) used the Delphi technique to identify diabetes care and education issues for Latinos with diabetes. At the international level, the World Health Organization (WHO) employed it to determine essential public health functions (Bettcher, Sapiric, & Goon, 1998). At the state level, Hahn, Toumey, Tayens, and McCoy (1999) used this approach to seek agreement among Kentucky legislators regarding tobacco control and tobacco farming policy. Elizabeth could use this technique as part of her overall data collection. For example, she could have a member of each hospital department involved in ambulatory care address the issue of necessary community health promotion programs in order to get consensus.

Generally, group members are identified who will generate the consensus position, and each member interacts individually to provide collective feedback. Individuals then reconsider their initial positions in light of group trends and can make adjustments accordingly. Eventually, this leads to an informed consensus isolated from the forces of the traditional approach.

Specifically, the sequence of events is as follows:

1. Identify the group members whose consensus opinions are sought. If they are representatives of a group (e.g., heads of departments of health sciences in universities), the sampling technique must be appropriate.
2. In the first questionnaire, each member of the group generates a list of concerns, goals, or issues toward which consensus opinions are desired (e.g., knowledge competencies for health educators or research issues confronting health education). The combined lists are edited, randomized, and placed in a format acceptable for a second questionnaire.
3. In the second questionnaire, each member rates or ranks the items derived from the initial questionnaire.
4. In the third questionnaire, the results of the second questionnaire are presented, revealing the preliminary level of group consensus to each item as well as repeating each member's previous response. The individual group member then rates or ranks each item a second time. If the member differs greatly from the group trend, a brief explanation should be given.
5. In the fourth questionnaire, the group trend becomes quite evident, as the results of the third questionnaire are presented for each item as well as the member's latest ranking or rating. Along with this is a listing by item of the major reasons for dissent from the group trend. In this questionnaire each member ranks or rates each item for a third and last time, keeping the group's emerging pattern in mind.

6. The results of the fourth questionnaire are calculated and presented as the final statement of group consensus.

If this technique is employed, it is necessary to have all the knowledge and skill required for survey design as well as questionnaire construction and design.

Case Discussion

As the newly hired Director for Outreach Education, Elizabeth definitely has her work cut out. The following steps would need to be taken or addressed if she were to carry out successful survey research:

1. She would need to identify her objectives and resources. One of her objectives is to demarcate the health needs of the people residing in the bicounty area. Another objective is to determine which health education/promotion programs are needed, while a third is to ascertain the best type of program delivery (when, where, size, and so forth).
2. Her overall design is likely to be descriptive—the “what” characteristics—with both cross-sectional and group-comparison items. Elizabeth may want to conduct a more detailed analytical design after she has analyzed information from her initial survey. The analytical approach might be best for a more focused survey effort.
3. It appears that her resources are limited, especially in regard to personnel. Nonetheless, it would be fruitful to recruit volunteers from the various departments to carry out a telephone survey. Assuming that the hospital has computer resources, she could seek assistance in designing her questionnaire so that it could be easily placed into an existing software package. Her volunteers would need to be trained to conduct CATI. If the bicounty area had remote places with a poverty-struck population, then personal interviews would have to be arranged to collect those data.
4. Elizabeth’s data analysis will be primarily descriptive in nature, with some cross-tabulations, such as gender and health needs, age, accessibility to programs, and the like.
5. In drawing a sample, her sampling frame would be all households residing in the bicounty area. For the CATI, she would use random sampling of households. She would need to ascertain the number of households and take a random sample. The remote households may require purposive sampling or convenience sampling (discussed in Chapter 7).
6. Her next step is to draft the questionnaire following the guidelines established in this chapter.
7. Using a subset of her intended population, Elizabeth should pretest the questionnaire and make revisions accordingly.
8. The next step is the actual administering of her precoded questionnaire.

9. The data collected should be cleaned or verified by both range and contingency checking.
10. The data will have been entered into the computer via CATI or CAPI, in the case of remote locations with no telephones.
11. Her data analysis will reflect percentages, means, and cross-tabulations because the design is descriptive in nature.
12. Her final step will be to write the report and present it to the appropriate people at the hospital.

SUMMARY

This chapter discussed survey research, commencing with the characteristics of such research. It was seen to be a more complicated matter than frequently perceived. The survey flow plan consists of 15 steps that outline the overall approach to doing a survey.

Overall design showed how both descriptive and analytical designs should relate to research questions. The descriptive designs included cross-sectional, longitudinal, and group comparison. Analytical designs addressed cross-sectional, case control, and prospective approaches.

The mail survey discussion included advantages, disadvantages, and many factors that influence the rate of return and the adequacy of data received. The latter included sponsorship, time and type of mailing, questionnaire length, color and format, ease of completion and return, incentives, nature of respondents, and follow-up procedures. There is little consensus as to what is an appropriate minimum number of returns. The chapter did discuss the issues of return rates and how they may be calculated.

The interview study was discussed, including the advantages and disadvantages of both the one-on-one interview and the group interview. The effects of interviewer characteristics—race, ethnicity, gender, social status and distance, age, clothing, and grooming—were presented and seen to have an effect on the interview process and results. Unstructured, semistructured, and structured interviews were presented according to reliability and ease of use.

The telephone interview was seen to have several advantages over the face-to-face interview methodology, despite some drawbacks. The increase in accessibility and the use of RRD combined with CATI has made the telephone interview more popular as a research technique. Computer-assisted surveys were addressed in regard to the telephone interview, personal interview, and self-administered questionnaires.

Four necessary components were presented for questionnaire design and construction: (1) planning the prequestionnaire, (2) drafting the questionnaire, (3) preparing the final questionnaire, and (4) pretesting. Emphasis was given to drafting the questionnaire with attention to researcher, respondent, and interviewer considerations; types of questionnaires (open or closed); and types of questions (dichotomous, multiple choice, rating, ranking, sentence completion, and open-ended). A checklist of nine pitfalls in question writing was presented, as were seven general rules for arranging the final draft.

The importance of the cover letter accompanying the instrument, as well as its necessary content and trappings, was discussed. Pretesting was viewed as a critical analysis so that further modification may be made of the questionnaire.

Focus group interviews were addressed briefly. As a final note, the Delphi technique was reviewed in some detail, with reference given to various health science studies. This unique group consensus approach appears to be one that is of great benefit.

CRITICAL THINKING QUESTIONS

1. Why is survey research often seen as a “poor or ineffective” research technique, and how would you contradict that perception when asked?
2. How would combining a telephone survey with a web-based approach possibly enhance the rate of participation?
3. Self-report bias is a concern in collecting health-related information, such as drug usage. What steps can be taken by the researcher to reduce such bias?
4. In what ways, if any, would the guidelines for conducting a “snail-mail” survey differ from an e-mail survey or web-based survey?
5. What are the differences between descriptive and analytical survey designs?

SUGGESTED ACTIVITIES

1. Conduct a literature search to identify investigations using (a) web-based approach, (b) mail survey, (c) telephone survey, and (d) a mixed approach. Compare and contrast the various approaches.
2. Go to the CDC website (www.cdc.gov), and click on *Data and Statistics*. Click on *Surveys* and then on *Behavioral Risk Factor Surveillance System (BRFSS)*. Click on *Questionnaires* to review the most recent one. Look closely at the design and question format, and be prepared to explain what you have learned about formatting, wording of questions, logging of answers, and data analysis.
3. Go to the CDC website (www.cdc.gov), and click on *Data and Statistics*. Click on *Surveys* and then on *National Health Care Surveys*. Select one of the surveys conducted by the National Center for Health Statistics (NCHS), and review the section titled *Survey Methodology*. Be prepared to discuss issues regarding instrument, design, collection, and reliability.
4. The State Department of Health has asked you to conduct a survey of registered dietitians regarding their perception of child and adolescent obesity in the state. Write out:
 - a. The objective of the survey
 - b. At least one research question

- c. The overall design of the survey
 - d. A proposed method of data collection
 - e. How you would determine sample size
5. Go to your Office of Instructional Technology or similar office to get their ideas of how an online survey should be conducted to get the best results.

REFERENCES

- Aday, L. A. (1989). *Designing and conducting health surveys: A comprehensive guide*. San Francisco: Jossey-Bass.
- Anderson, R. M., Goddard, C. F., Garcia, R., Guzman, J. R., & Vazquez, F. (1998). Using focus groups to identify diabetes care and education issues for Latinos with diabetes. *Diabetes Educator, 24*(5), 618–625.
- Annemans, L., Giaccone, G., & Vergnenegre, A. (1999). The cost effectiveness of paclitaxel (Taxol) + cisplatin is similar to that of teniposide + cisplatin in advanced non-small cell lung cancer. *Anti-Cancer Drugs, 10*(6), 605–615.
- Bailey, K. (1994). *Methods of social research*. New York: Free Press.
- Barriball, K. I., & White, A. E. (1999). Non-response in survey research: A methodological discussion and development of an explanatory model. *Journal of Advanced Nursing, 30*(3), 677–686.
- Berg, B. L. (2001). *Qualitative research methods for the social sciences* (2nd ed.). Boston: Allyn & Bacon.
- Bettcher, D. W., Sapirie, S., & Goon, E. H. (1998). Essential public health functions: Results of the international Delphi study. *World Health Statistics Quarterly, 51*(1), 44–54.
- Bilodeau, B. A., & Degner, L. F. (1996). Information needs, sources of information, and decisional roles in women with breast cancer. *Oncology Nursing Forum, 23*(4), 691–696.
- Burns, N., & Grove, S. K. (1993). *The practice of nursing research* (2nd ed.). Philadelphia: W. B. Saunders.
- Conklin, K. A. (1997, May 18–21). *Telephone survey research*. Paper presented at the annual forum of the Association for Institutional Research, Orlando. (ERIC Document Reproduction Service No. ED 410886)
- Conklin, K. A. (1999). Community college telephone survey research: An overview of methodology and utility. *Community College Journal of Research and Practice, 23*(4), 423–433.
- Daley, E. M., McDermott, R. J., McCormack-Brown, K. R., & Kittleson, M. J. (2003). Conducting Web-based survey research: A lesson in internet designs. *American Journal of Health Behavior, 27*(2), 116–124.
- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi Method to the use of experts. *Management Science, 3*, 458.
- Degner, L. F., Davison, B. J., Sloan, J. A., & Mueller, B. (1998). Development of a scale to measure information needs in cancer care. *Journal of Nursing Measurement, 6*(2), 137–153.
- Fanning, E. (2005). Formatting a paper-based survey questionnaire: Best practices. *Practical Assessment, Research and Evaluation, 10*(12), 1–14. Available at <http://pareonline.net/getv.asp?v=10&n=12>.
- Ferguson, L. (1941). A study of the Likert technique of attitude scale construction. *Journal of Social Psychology, 13*, 51–57.

- Fowler, F. J., Gallagher, P. M., & Nederend, S. (1999). Comparing telephone and mail responses to CAHPS survey instrument: Consumer assessment of health plans study. *Medical Care*, 37(Suppl. 3), MS41–49.
- Frey, J. H. (1989). *Survey research by telephone* (2nd ed.). Newberry Park, CA: Sage Publications.
- Fricker, S., Galesic, M., Tourangeau, R., & T. Yan (2005). An experimental comparison of Web and telephone surveys. *Public Opinion Quarterly* 69(3), 370–392
- Fricker, R., & Schonlan, M. (2002). Advantages and disadvantages of Internet research surveys: Evidence from the literature. *Field Methods* 14(4), 347–367.
- Furlong, D. K. (1997, May 18–21). *Between anecdote and science: Using e-mail to learn about student experiences*. Paper presented at the annual forum of the Association for Institutional Research, Orlando. (ERIC Document Reproduction Service No. ED 409320)
- Graves, R. M., & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.
- Greene, J., Speizer, H., & Wiitala, W. (2008). Telephone and Web: Mixed-mode challenge. *Health Services Research* 43(1), 230–249.
- Hahn, E. J., Toumey, C. P., Tayens, M. K., & McCoy, C. A. (1999). Kentucky legislators' views on tobacco policy. *American Journal of Preventive Medicine*, 16(2), 81–88.
- Haugland, S., & Wold, B. (2001). Subjective health complaints in adolescence: Reliability and validity of survey methods, *Journal of Adolescence*, 24, 611–624.
- Heflich, D., & Rice, M. (1999). *Online survey research: A venue for reflective conversation and professional development*. Paper presented at the international conference of the Society for Information Technology and Teacher Education, San Antonio, TX. (ERIC Document Reproduction Service No. ED 432283)
- Heise, D. (1970). The semantic differential and attitude research. In G. F. Summers (Ed.), *Attitude measurement*. Chicago: Rand McNally.
- Helmer, O. (1967). *Analysis of the future: The Delphi technique*. Santa Monica, CA: Rand.
- Hughes, J., & Pakieser, R. (1999). Factors that impact nurses' use of electronic mail. *Computers in Nursing*, 17(6), 251–258.
- Isaac, S., & Michael, W. (1981). *Handbook in research and evaluation* (2nd ed.). San Diego, CA: Edits.
- Jenkins, J., Russell, W., & Suci, G. (1958). An atlas of semantic profiles for 360 words. *American Journal of Psychology*, 71, 618–699.
- Johanson, G., Green, S., & Williams, J. (1998, April 13–17). *Blunders and missed opportunities in survey research*. Paper presented at the annual meeting of the American Educational Research Association, San Diego. (ERIC Document Reproduction Service No. ED 420693)
- Johnson, T., & Fendrich, M. (2005). Modeling sources of self-report bias in a survey of drug use epidemiology. *Annals Epidemiology*, 15, 381–389.
- Kerlinger, F. (1964). *Foundations of behavioral research*. New York: Holt, Rinehart, & Winston.
- Kittleston, M. J. (2003) Suggestions for using the Web to collect data. *American Journal of Health Behavior*, 27(2), 170–172.
- Kitzinger, J. (1990). Recalling the pain: Incest survivors' experiences of obstetrics and gynaecology. *Nursing Times*, 86, 38–40.
- Kitzinger, J. (1995, July 29). Introducing focus groups. *British Medical Journal*, 299.
- Klecka, W. R., & Tuchfarber, A. J. (1978). Random digit dialing: A comparison to personal surveys. *Public Opinion Quarterly*, 42, 105–114.
- Leedy, P. (1980). *Practical research: Planning and design*. New York: Macmillan.
- Leenen, F. H., Dumais, J., McInnis, N. H., Turton, P., Stratyckuk, I., Nemeth, K., Lum-Kwong, M. M., & Fodor, G. (2008). Results of the Ontario survey on the prevalence and control of hypertension. *Canadian Medical Association Journal*. 178(11), 1441–1450.

- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 21, 140.
- Mavis, B. E. (1998). Postal surveys versus electronic mail surveys: The tortoise and the hare revisited. *Evaluation Health Professions*, 21(3), 395–408.
- McMahon, L. F., & Fagerline, A. (2004). Reflections from the population: Survey methods. *Journal Clinical Gastroenterology*, 38(4), 307–308.
- Meha, R., & Sivadas, E. (1995). Comparing response rates and response content in mail versus electronic mail surveys. *Journal of the Market Research Society*, 37, 429–439.
- Murphy, G., & Likert, R. (1937). *Public opinion and the individual*. New York: Harper.
- National Center for Education Statistics. (1997). *National household education survey data files and electronic codebook*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.
- National Health Interview Survey. (1999). National health interview survey: Research for the 1995–2004 redesign. *Vital and Health Statistics Series 2: Data Evaluation and Methods Research*, 126, 1–119.
- Nesbary, D. K. (2000). *Survey research and the World Wide Web*. Boston: Allyn & Bacon.
- Nixon, M. K., Cloutier, P., & Jansson, M. (2008). Nonsuicidal self-harm in youth: A population-based survey. *Canadian Medical Association Journal*, 178(3), 306–313.
- Nunnally, J. (1962). The analysis of profile data. *Psychological Bulletin*, 59, 311–319.
- O'Connor, K. G., Sharp, S., & Olson, L. (1999). The effect of monetary incentives on a survey of pediatricians. *Health Services Research*, 16, 42.
- Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Patrick, J. H., Pruchno, R. A., & Rose, M. S. (1998). Recruiting research participants: A comparison of the costs and effectiveness of five recruitment strategies. *Gerontologist*, 38(3), 295–302.
- Pealer, L., & Weiler, R. M. (2003). Guidelines for designing a web-delivered college health risk behavior survey: Lesson learned from the University of Florida Health Behavior Survey. *Health Promotion and Practice*, 4(2), 171–179.
- Rogelberg, S. G., Luong, A., Sederburg, M. E., & Cristol, D. S. (2000). Employee attitude surveys: Examining the attitudes of noncompliant employees. *Journal of Applied Psychology*, 85(2), 284–293.
- Rubin, H. J., & Rubin, I. S. (1997). *Qualitative interviewing: The art of hearing data* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Russell, C. L., & Van Gelder, F. (2008). An international perspective: Job satisfaction among transplant nurses. *Prog Transplant*, 18(1), 32–40.
- Scott, C. (1961). Research on mail surveys. *Journal of the Royal Statistical Society*, 124(A), 143–195.
- Sellitz, C., Jahoda, M., Deutsch, M., & Cook, S. (1959). *Research methods in social relations*. New York: Holt, Rinehart, & Winston.
- Sim, J. (1998). Collecting and analysing qualitative data: Issues raised by the focus group. *Journal of Advanced Nursing*, 28(2), 345–352.
- Snider, J., & Osgood, C. (Eds.). (1969). *Semantic differential technique*. Chicago: Aldine.
- Sorokin, W. (1976). *Personal health appraisal*. New York: John Wiley & Sons.
- Sowell, R. L. (2000). Identifying HIV/AIDS research priorities for the next millenium: A Delphi study with nurses in AIDS care. *Journal of the Association of Nurses in AIDS Care*, 11(3), 42–52.
- Sully, P., & Grant, P. (1997). Ethical and methodological improvements to mail survey research: An alternative follow-up method. *Canadian Journal of Program Evaluation*, 12(2), 81–104.
- Taylor, A. W., Wilson, D. H., & Wakefield, M. (1998). Differences in health estimates using telephone and door-to-door survey methods. *Australian and New Zealand Journal of Public Health*, 22(2), 223–226.

- Taylor, W. C., Yancey, A. K., Leslie, J., Murray, N. G., Cummings, S. S., Sharkey, S. A., et al. (1999). Physical activity among African American and Latino middle school girls: Consistent beliefs, expectations, and experiences across two sites. *Woman and Health, 30*(2), 67–82.
- Thurstone, L., & Chave, E. (1929). *The measurement of attitudes*. Chicago: University of Chicago Press.
- Tordella, M., & Neutens, J. (1979). An instrument to appraise attitudes of college students toward euthanasia. *Journal of School Health, 49*(6), 351–352.
- Trow, M. (1967). Education and survey research. In C. Clock (Ed.), *Survey research in the social sciences*. New York: Russell Sage Foundation.
- van Alphen, A., Halfens, R., Hasman, A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing, 20*(1), 196–201.
- Wang, L., & Fan, X. (1997, March 24–28). *The effect of cluster sampling design in survey research on the standard error statistic*. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 409320)
- Watts, M., & Ebbutt, D. (1987). More than the sum of the parts: Research methods in group interviewing. *British Medical Journal, 13*, 25–34.
- Wiersma, W. (2000). *Research methods in education* (7th ed.). Boston: Allyn & Bacon.
- Wilson, A. H., Pittman, K., & Wold, J. L. (2000). Listening to the quiet voices of Hispanic migrant children about health. *Journal of Pediatric Nursing, 15*(3), 137–147.

Sampling Designs and Techniques

KEY TERMS

alpha (α) level	mixed sampling	sampling error
beta (β) level	multistage cluster sampling	sampling frame
cluster or area sampling	nonprobability samples	sampling unit
confidence level	population	simple random sampling
convenience sampling	power	snowball sampling
dimensional sampling	probability samples	stratified random sampling
effect size (ES)	purposive sampling	systematic sampling
level of statistical significance	quota sampling	Type I error
	sample	Type II error

Case Study

Amy, a health researcher for the city-county public health department, was alarmed at the increasing rate of obesity across the nation. Just walking the streets in the city or county, she could see that this was a public health concern for her area also. This concern was confirmed by the number of people who came to the public health clinic experiencing problems with being overweight. Amy addressed the problem with both the city and county governments by suggesting that the public health department conduct a survey of residents using some of the same questions from the National Health and Nutrition Examination Survey (NHANES). The survey could include items about eating habits, dietary supplements, weight history, and, of course, physical activity and fitness. Both governments agreed to support the survey after she explained the health consequences of being overweight and the costs to the community. They asked Amy to develop a proposal to include the questionnaire, who she would survey, and how many she would survey, and they requested she separate the city participants from the county participants in presenting the data after analysis.

The Purpose of Sampling

Ideally, the researcher would like to observe the entire population to add more weight to the findings. For example, while it might be good for Amy to have everyone living in the city and county complete the survey, it is unlikely that she could capture the entire **population** due to limited resources, time, and money. Subsequently, a subset of the population must be selected. This subset is actually a **sample** of the population and as such should represent the total population so that the data collected will be as accurate as if taken from the entire population.

The logic involved is simple. Nonetheless, the importance of this step cannot be overemphasized. A knowledgeable researcher commences with a population and works down to the sample. In other words, the population of interest is designated, and then a sample is derived. The neophyte, in contrast, often works from the bottom up by attempting to ascertain the minimum number of respondents needed for a successful study. The inherent problem with this approach is that it is next to impossible to assess the representativeness of the sample because the entire population has not been identified.

While savings in time and money are obvious reasons for sampling, there are additional advantages, too. The sample may achieve a greater response rate owing to greater cooperation than might occur in the full population survey. This, in itself, would tend to make the results more accurate. In health surveys with sensitive items, this point is particularly important. Concomitantly, the researcher can keep a low profile by using a sample. That is, less people may be offended, thereby negating an opportunity for several people to organize a common resistance. In the case of interviewing, a sample reduces the number of interviews and interviewees. This is beneficial in that supervision of an enormous number of interviewers is difficult at best and necessary attention to details such as follow-ups becomes cumbersome as numbers increase. There is little doubt that using sampling, in most instances, is more advantageous than using an entire population. However, the benefits are realized only if the sample is drawn with precision.

The Sampling Frame

The **sampling frame** is a list of all the persons (objects) from whom the sample is to be drawn. Understandably, the sample cannot be more accurate than the sampling frame from which it is selected. In constructing the sampling frame, the researcher lists every person in the population but does so only once so as not to increase someone's likelihood of being chosen. If the study is small, it is recommended that the investigator construct the list personally to avoid omissions and repetitions that may be on existing lists. In a large study however, it is much more difficult and perhaps virtually impossible to procure an accurate and complete list. In the case study, it is probable that Amy would not be able to obtain a listing of all residents in the city and county area. People are born daily, and others die, move, or give incorrect addresses so they cannot be contacted. As the size of the study increases to include a city, county, state, or nation, the construction of the sampling frame becomes more formidable.

Employment of existing lists, such as telephone directories and county directories, is only a partial answer at best. Both poor people who may not own telephones and wealthy people who may have unlisted numbers will be excluded from the study. Another group that may be excluded is the growing number of people who no longer have land-based lines and use their cell phones only. Those who have two telephones and multiple listings would have a greater opportunity for selection than those with a single listing. Unless the directory was recently compiled, address changes and people who have left the county would confound the list. If the researcher employs existing lists, an attempt should be made to ascertain the number of persons excluded and whether they differ in any systematic way from those on the list. If there is no common bond, that is, if people are excluded randomly from the total population, then little harm will be done.

A possible alternative to the time and monetary demands of listing individuals is to compile residence addresses of households. Residences are relatively stable, and a complete listing could serve as the sampling frame. Further, no groups should be excluded, and the risk of bias should be decreased greatly. In attempting to survey the people in her area, Amy could use this approach and select a sample from the sampling frame of residential addresses.

Another consideration when selecting a sample from the sampling frame is the sampling unit. The **sampling unit** could be an individual, an intact group such as a classroom or school, an organization such as the local medical society, or even a geographical region such as a city or county (Bowling, 2002; Cottrell & McKenzie, 2005). The sampling unit is particularly important for data analysis. For example, Amy will have one sampling frame for the county and one for the city. Per the directions of both governments, she will draw a sample from each sampling frame. While each sample contains a large number of individuals, her actual sampling unit would be the city and the county. Amy's data analysis would be at the city-county level rather than at the individual resident level. Similarly, in most K-12 studies, the sampling unit is a classroom or even the school rather than the individual students. As discussed in a previous chapter, the classrooms or schools may be randomized as intact groups since the students cannot be randomized individually in most studies. In clinical trials and similar research the patients are randomized and are truly the sampling unit.

Figure 7.1 illustrates the reduction process by which a researcher such as Amy could refine her sampling unit from the population as a whole.

Sampling Techniques

Once the population has been defined and the sampling frame established, the next step in arriving at a target group for research purposes is to select a method of sampling. Basically, there are two types of sampling techniques, which have several different procedures. **Probability samples** are those wherein the probability of selection of each respondent, address, or even object is known. In contrast, **nonprobability samples** reflect an unknown probability of selection.

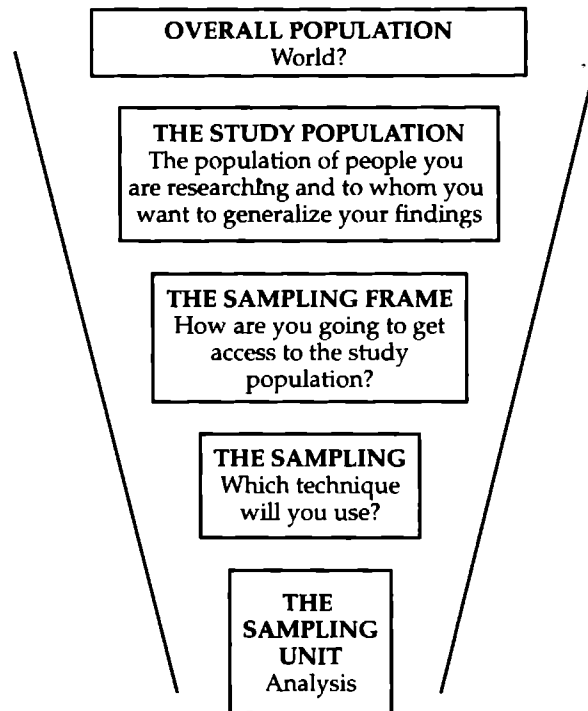


Figure 7.1 The Pathway and Relationship from Population to Sampling Unit

Probability Sampling

Probability sampling techniques include random sampling, systematic sampling, stratified random sampling, and cluster sampling. The common thread in all of these techniques is the random selection of participants or objects such as schools or hospitals. Because selection is random, each participant or object has an equal chance of being chosen and a known probability of being chosen.

Random Sampling **Simple random sampling** is the basic building block of all probability sampling designs. In a random sample, each person (or address or object) in the population has an equal chance of being chosen for the sample. This is accomplished without bias for any personal characteristics. Of course, the underlying necessity is an adequate sampling frame with no one listed more than once and no one excluded. If either of these occurs, then by definition the sampling fails to be random.

An additional point about simple random sampling is that it is sampling without replacement. For example, if the sampling frame comprises 300 people, the first person has a 1 in 300 chance of being selected. After 150 people have been chosen, the remainder will have only a 1 in 150 chance of being selected. This is considered adequate because the opportunity for selection is equal at any given stage of the sampling process.

While there are many methods for random selection, such as the flip of a coin, a lottery, or the spin of a roulette wheel, the usual one used by researchers is the table of random

numbers. Each person in the sampling frame is assigned a number through identifiable characteristics (names, age, gender) to avoid bias. From the example, each of the 300 people in the sampling frame would be given one number. The researcher would then employ the table of random numbers to commence selecting the sample.

In cases for which the sample size is small, the health science researcher could simply select numbers by hand from a random numbers table. However, as the size of the sample increases, the task is much better accomplished by using computer software that generates random numbers.

Systematic Sampling One alternative to the process of simple random sampling is **systematic sampling**. As the name implies, it is the selection of specific items in a series according to some predetermined sequence. The origin of the sequence must be controlled by chance. In other words, systematic sampling can be employed only when units in the sampling frame are random. In our case study, the residential addresses would have to be randomly ordered within the sampling frame. If they were not random, then Amy would be unable to use systematic sampling. However, if the items in the sampling frame were randomly listed, the health science investigator could choose $1/k$ th of them, with k being any constant. If k were 2, then the sample would comprise one-half of the population. Similarly, if k were 5 the sample would be 20% of the entire population.

Generally, the investigator randomly selects the first item from among the k items in the sampling frame. Next, by definition, a $1/k$ th sample is established by choosing every k th item in the sampling frame. Subsequently, Amy would randomly select her first address for inclusion in the sample. Needless to say, like Amy all health science investigators would need to determine sample size before beginning sample selection.

Although simple random sampling is more accurate and does not require the assumption of a randomized sampling frame, systematic sampling involves less work, thereby providing more information per dollar. Further, for the inexperienced survey researcher, it may reduce error because it is simpler to perform. In short, the greater the complexity of the method, the greater the opportunity for error. Nonetheless, it must be emphasized that systematic sampling is more dependent on the adequacy of the sampling frame than is simple random sampling. Because any ordering of the sampling frame is retained in systematic sampling, the results can be totally nonrepresentative. If evidence of biased ordering is found in the sampling frame, then steps must be taken to correct it. The most obvious step is to randomize the sampling frame (which is expensive and time consuming) or, if this is untenable, to return to simple random sampling or perhaps to draw a stratified random sample.

Stratified Random Sampling At times it is advisable to use **stratified random sampling**. This technique subdivides the population into smaller homogeneous groups in order to get a more accurate representation or to include parameters of special interest. Herein, the population is broken down into nonoverlapping groups called *strata*, and then a simple random sample is extracted from each stratum.

The first step in stratified random sampling is to identify the strata (sometimes called *stratification parameters*). For example, Amy is likely to subdivide the population of the city-county area into city and county as separate areas. If she used residential addresses, they would be split exclusively into the appropriate category of city or county, and a simple random sample could be taken from each list. Although a simple random

sample could have been employed for a listing of combined city-county residents and still have good representation, the stratified design could save time and money by requiring a smaller sample size.

If so desired, more than one stratification variable or parameter could be used. That is, the health science researcher can stratify on two or more variables simultaneously. In addition to the city-county parameter, Amy could stratify on gender also. Now she would have four groups rather than just two. Table 7.1 illustrates the four groups or strata. Once the groups have been formed, a simple random sample can be taken within each group or stratum.

Table 7.1 Stratified Random Sampling

Gender	Geographic Area	
	City	County
Male	Cell 1	Cell 2
Female	Cell 3	Cell 4

In proportional stratified sampling, each sample drawn should represent the population in the proportion in which it exists within the total population. For example, assume that Amy's target population is between the ages of 20 and 74 years of age and imagine that in the city there were 110,000 females in that age strata and 90,000 males for a total of 200,000. Similarly, using the same age parameters, suppose there are 48,000 females and 52,000 males for a total of 100,000 in the county. Combined there would be 300,000 participants who would be in the original sampling frame. For Amy to establish a sample size of 2000, the sample should include approximately the same proportions as the entire study population. Table 7.2 illustrates the proportional stratified sampling design.

Table 7.2 Proportional Stratified Sampling

	City	County
Male	90,000	52,000
Proportion	30%	17%
Sample Size	600	340
Female	110,000	48,000
Proportion	37%	16%
Sample Size	740	320

Each cell is drawn randomly in proportion to the total study population. Cell 1, comprising 90,000 males, makes up 30% of the study population and subsequently has a sample size of 600 (30% of 2000). The remaining cells follow a similar pattern.

In other situations, the researcher may decide to stratify by gender, socioeconomic status, racial origin, education, or religious preference. Obviously, those in the health sciences may wish to stratify by health parameters such as smoking-nonsmoking,

hypertensive patients-nonhypertensive patients, pregnancy-nonpregnancy, and many others. The characteristics of the entire population must be considered together with the objectives of the research before the stratified sample design is used. On occasion, stratified random sampling is disproportionate in that a larger proportion of the population is sampled in one stratum than in another. Two reasons for this design are (1) differences in population size; and (2) homogeneity among strata. When the population of a particular stratum is very small, proportionate sampling may leave the researcher with a sample size that is statistically unworkable. Similarly, a larger proportion of the population would have to be sampled in a very heterogeneous stratum. For example, if the health science investigator were to select a stratified sample of nonsmokers, cigarette smokers, and pipe smokers in a company that has 1,000 nonsmokers, 800 cigarette smokers, and 200 pipe smokers, a greater proportion of pipe smokers would have to be sampled than either nonsmokers or cigarette smokers to obtain a sufficient size for sampling adequacy. This dilemma of disproportionate sampling will be discussed further in the *Sample Size* section of this chapter.

Overall, stratified random sampling is a technique to maintain the same proportionality on stratification parameters in the sample as occurs in the population. The researcher may stratify by demographic characteristics or by health variables. In any case, the characteristics of the entire population must be considered together with the objectives of the research before the stratified sample design is used.

Cluster or Area Sampling Cluster or area sampling is a variation of the simple random sample and is especially useful when (1) the population to be studied is infinite; (2) a list of members of the population is nonexistent; or (3) the geographic distribution of the population is widely scattered. For example, if an investigator proposed to survey all public school health educators in the United States, a simple random sample would be impractical.

In **multistage cluster sampling**, the investigator can first randomly sample 20 of the 50 states. In the second stage, from a sampling frame that lists all counties within the 20 states, a random sample of 100 counties could be selected. Then, in the third stage, a random sample of 50 school districts could be drawn from all the school districts within the 100 counties. The fourth stage could consist of random selection of 100 school health educators in the 50 school districts. The successive random sampling of states, counties, school districts, and finally health educators is relatively inexpensive and efficient.

Cluster sampling samples among clusters. While it has some advantages over simple random sampling, it does hold the possibility of more error. This is because it is not a single sample but rather two or more, each open to error. Further, there may be sample bias because of the unequal size of some of the subsets or clusters selected. The first stage of sampling may be representative, but the second stage may not be. The researcher must be concerned about the sample size and accuracy at every stage of the cluster sample.

In summation, all of the techniques we have discussed—simple random sampling, systematic sampling, stratified random sampling, and cluster sampling—may be combined into a single procedure to suit the needs of the researcher. In so doing, the investigator must be familiar with the idiosyncrasies of each method.

Nonprobability Sampling

In some instances, the researcher may decide to employ nonprobability sampling. In this method, the probability that a person will be chosen is not known, with the result that a claim for representativeness of the population cannot be made. Concomitantly, sampling error (the degree of departure from representation) is unknown. Subsequently, the researcher's ability to generalize findings beyond the actual sample is greatly limited. This may be a major disadvantage, depending upon the purpose of the study. Nevertheless, nonprobability sampling has an advantage over probability sampling in that it is less expensive, less complicated, and lends itself to spontaneity (spur-of-the-moment investigations). It is particularly useful in small studies or pilot investigations to perfect questionnaires. Nonprobability sampling includes convenience sampling, quota sampling, dimensional sampling, purposive sampling, and snowball sampling.

Convenience Sampling A common example of **convenience sampling** is the captive-audience approach, such as using a classroom full of health science students. While the researcher forgoes representativeness in this case, time and money are saved. It is a simple matter of selecting the closest and most convenient persons.

Quota Sampling **Quota sampling** is the nonprobability sampling equivalent of stratified sampling. Initially, the researcher determines which strata are relevant to the investigation and then proceeds to establish a quota for each stratum that is proportionate to its representation in the population. For example, in Amy's study, 30% of people in the city were male and 37% female. In the county, 17% were male, and 16% were female. Preferably the sample would be drawn proportionate to the gender realities of the study population. Once the quota is set, Amy simply finds the addresses (or phone numbers or e-mails) that fit into the stratum. Imagine Amy decided to have a quota sample of 200 participants. That would mean 60 males and 73 females living in the city and 35 males and 32 females residing in the county.

The researcher must make every effort to prevent bias. Bias is most likely to occur when the route of least resistance is chosen (e.g., avoiding houses in questionable neighborhoods or ones that contain unfriendly people). Confining the research to friends and acquaintances is not acceptable.

Dimensional Sampling **Dimensional sampling** is principally a multidimensional form of quota sampling wherein the variables (dimensions) of interest in the population are delineated. Each variable and combination thereof must be represented by at least one case. It is a method in which only a small sample is required so that each case selected can be examined in more detail.

Purposive Sampling This technique falls somewhere between quota sampling, in which various strata are to be filled, and convenience sampling, wherein the nearest and most available people are used. In **purposive sampling**, the researcher employs his or her own discretion to select the respondents who best meet the purposes of the study. This is a great advantage to the experienced researcher who can apply prior knowledge and skill.

Snowball Sampling There is a multistage technique that literally “snowballs.” In the first stage of **snowball sampling**, a person possessing the requisite characteristics is identified and interviewed. This person then identifies others who may be included in the sample. The next stage is to interview these persons, who in turn identify still more respondents who can be contacted and interviewed in following stages.

Mixed Sampling Designs When a population or sample is very large, a mixed model of judgment and probability sampling is often used. Discretion procedures are frequently employed in the early stages and probability procedures in the later stages. This **mixed sampling** approach offers a savings in time, money, and effort as well as a sample that can be representative of the entire population.

Sample Size

The determination of sample size usually perplexes many researchers because they often have no conception of a minimally adequate sample size. They need to understand that correct sample size is dependent on both the nature of the population and the purpose of the study. Usually, a trade-off is discovered between the desire for a large sample and the feasibility of a small one. An ideal study would have a sample large enough to represent the population so generalization may occur yet be small enough to save time and money, as well as to reduce the complexity of data analysis.

Considerations in Sample Size It is a popular misconception that a sample is a small carbon copy of the original population, identical in every way. If this were the case, then the researcher would not have to worry about having a sample size that is representative of the population under study. Needless to say, one can never be certain of representativeness unless the entire population is used. An obvious deduction at this juncture is that the larger the sample, the greater the likelihood of representativeness. This is especially true if the population is quite heterogeneous on the given variable; the greater the heterogeneity, the greater the necessity for a larger sample. For populations in which there is no heterogeneity on a variable (complete homogeneity), a sample size of even one would suffice. Some sample designs are more amenable than others in making the population more homogeneous. For example, the stratification in random sampling narrows the heterogeneous population under study into groups or strata that are more homogeneous thereby allowing for a smaller sample size.

Sampling Error **Sampling error**, sometimes called *level of precision*, is the degree to which the sample means of repeatedly drawn random samples differ from one another and from the population mean. For example in studying hypertension, imagine that you took 10 random samples from the study population and calculated the mean of the 10 samples. As you would expect, it is highly likely that the 10 means would be different from one another as well as different from the population parameter mean. While most would tend to cluster around the population parameter mean, some would be relatively high by comparison while others would be relatively low. This variation is a result of sampling error. It is not a mistake in the sampling process but rather an inevitable

variation when a number of randomly selected sample means are compared. This variation or sampling error reflects lack of representativeness. Norwood (2000) noted that sample sizes of 30 or more per variable are workable. While this may be mathematically correct, keep in mind that the larger the sample size the smaller the sampling error simply because the larger the sample the closer it is to the study population in all measured characteristics.

Sample Size for Studies with Hypotheses: Analytical Studies

When the research effort has a null and alternative hypothesis such factors as alpha (α), beta (β), effect size, power ($1 - \beta$), and directionality should be taken into account (Cottrell & McKenzie, 2005; Hulley, Cummings, Browner, Grady, & Newman, 2007). When testing the null hypothesis to determine if it is true, the researcher sets a cut-off point called the alpha level to make the decision. This **alpha level** is the guide for rejecting or not rejecting the null hypothesis. In making the decision, the researcher can make an error, however. If the health science researcher rejects the null hypothesis when in fact it is true, a **Type I error** occurs. The probability is called alpha (α) or the **level of statistical significance**. That is, the probability of this error occurring is the same as the alpha level set by the researcher. In other words, if the null hypothesis is rejected at the .01 level there is a 1% risk of rejecting it when it is actually true. Similarly, using the .05 alpha level of significance, the researcher is taking a 5% risk of rejecting the null hypothesis even when it is true.

On the other hand, if the researcher does not reject the null hypothesis when in fact it is false, a **Type II error** occurs. The term **beta (β) level** refers to the chance of making a Type II error. If β is set at .10, the researcher is taking a 10% chance of making a Type II error (failing to reject the null hypothesis because the statistical test is not significant when in fact the null hypothesis is not true). These two types of errors require a balancing act by the researcher. As the researcher decreases the possibility of making a Type I error, the possibility of making a Type II error automatically increases. In health science education research, α is generally set at .05 while β is set at four times α which would be .20 (Cottrell & McKenzie, 2005; Windsor, Clark, Boyd, & Goodman, 1994). Table 7.3 illustrates both types of error.

A related factor when considering sample size for analytical studies is called power. **Power**, represented by $1 - \beta$, is simply the probability of correctly rejecting the null hypothesis. Since this is done through statistical tests (rejecting or not rejecting the null hypothesis), $1 - \beta$ is sometimes referred to as the power of a statistical test (Polit & Beck, 2004). If the researcher follows tradition setting $\beta = .20$ then the power of the statistical test is .80. Accordingly, the researcher has an 80% chance of correctly rejecting the null

Table 7.3 Type I and Type II Errors

Researcher's Decision	True Conditions	
	Null Hypothesis True	Null Hypothesis False
Reject the null hypothesis	Type I error	Correct decision
Fail to reject null hypothesis	Correct decision	Type II error

hypothesis. This may be restated to say that the researcher has an 80% chance of finding a relationship when in fact it exists. It should be noted that power also depends on the precision of the research design as well as the statistical procedure used (Graziano & Raulin, 2000).

Effect size is yet another factor in determining sample size. **Effect size (ES)** goes beyond significance testing in an attempt to quantify the size of the association between the sample and the population or any two groups under investigation. ES or degree of association between the two groups (such as sample and population) or between any two variables (such as weight and blood pressure) may be determined by prior research. If that fails, the researcher may want to conduct a pilot study. If that is not feasible, a standardized effect size of small, medium, and large has been established (Cohen, 1988; Cottrell & McKenzie, 2005). Several tables and websites are available for researchers to use for power analysis and should be consulted since it is beyond the scope of this chapter.

In addition to these factors, the researcher should consider directionality (Cottrell & McKenzie, 2005; Hurlburt, 2003). If a “one-sided” test of significance is employed (because the researcher believes the effects of the study will point in one direction, such as a hypertensive medicine that will reduce blood pressure versus not knowing whether it will reduce, increase, or have no effect), a different formula or table would be used, thus generally indicating a smaller sample size than needed for a two-sided test of significance. Finally, the researcher must take study design into account. Study design (such as crossover versus parallel clinical trials or stratified versus nonstratified sampling) impacts sample size with crossover, and stratified approaches generally allow for a smaller sample size.

Since sample size is such an important and complex topic, it suggested that researchers employ computer programs or consultants. Hulley et al. (2007) suggest that the following steps be taken when working with tables, formulas, or consultants for analytical research:

1. State the null and alternative hypotheses (one or two sided)
2. Choose the best statistical test
3. Select an appropriate ES
4. Set both alpha and beta
5. Use a table, formula, or consultant to determine sample size

Hulley et al. (2007) provide several different tables to arrive at sample size, as well as methods using simple statistical tests (chi-square, *t*-test, correlation coefficient) for estimating sample size.

Sample Size for Studies Without Hypotheses: Surveys and Descriptive Studies

Descriptive studies and surveys may have research questions but generally do not compare groups or have specific outcome variables. Consequently, the concepts of power, hypotheses, and the like do not apply (Hulley et al., 2007). Instead, these studies look at

information through descriptive eyes such as proportions and means. When descriptive studies or surveys do address analytical issues (example for Amy: what are the predictors of obesity in the city-county population?), sample size should be calculated so there is sufficient power to answer the question.

In surveys, the researchers frequently look for precision by establishing **confidence levels**. This confidence level, also called probability level, is usually set at the 95% or 99% level. At the 95% level, the researcher has confidence that there is a 95% chance that the sample is distributed in the same way as the population. This, of course, is based on the normal curve in which 95% of the curve includes two standard deviations on each side of the mean. The interval is even wider at 99% since it includes three standard deviations on each side the mean. Obviously this confidence level is more likely to include the true population value since it is much wider.

A related concern is how much sampling error the researcher is willing to tolerate. This can be expressed as the confidence interval, which means that the researcher knowingly accepts a specific percentage of error. For example, the media frequently report poll results with a plus-or-minus figure. In the case study, Amy may set a confidence interval of 4% (she is willing to accept a 4% sampling error). Imagine that she finds that 60% of her sample believe that a carbohydrate diet is the best way to control weight. Using her established confidence interval of 4%, that would mean that if she had asked the question of the entire population, between 56% ($60 - 4$) and 64% ($60 + 4$) would have selected that answer. The survey researcher would employ both the confidence level and confidence interval for precision and reliability of the data. Amy would note that she is 95% confident that the true percentage of the population is between 56 and 64%.

Several “calculators” are available on websites, both free and commercial, that calculate sample size based on the researcher establishing the confidence interval, the confidence level, and providing the number in the total population. For example, in Amy’s study, suppose that she is doing a simple random sample of the target population in the city (which number 200,000) and set a confidence interval of 4% and a confidence level of 95%. She would simply apply this information to calculate her sample size. The calculators would find that she requires a sample size of 598 people. Keep in mind these calculators require a true random sampling. Hulley et al. (2007) provide several tables that can be used to ascertain sample size in descriptive studies.

Aday and Cornelius (2006) offers a series of steps using this information and more to determine sample size for health surveys. Those steps are illustrated in the following. Note there are additional considerations, such as design effect, response rate, expected proportion of eligibles, and cost.

Step 1: Identification of Major Study Variables. As noted in Chapter 6, one of the principal steps of survey research is to delineate the variables under study. In Amy’s study, her major area of interest is obesity in the city and county populations with related interest in eating habits, dietary supplements, weight history, and physical activity and fitness.

Step 2: Types of Estimates of Study Variables. Amy has a choice of several different ways to summarize the study variable—percentages, ratios, means, and so on. She should pay heed to the level of measurement that will be used in her analysis. In this example, she chooses to use percentages—the percent who are overweight.

Step 3: Population or Subgroup Selection. The design of the study dictates the population or subgroup in question. For example, is the survey descriptive or analytical in nature? Is it a group comparison, cross-sectional, or longitudinal approach?

Step 4: Relevant Standard Error Formula. There are several ways to calculate sampling (standard) error when determining sample size. The method chosen should correspond with the analysis and study design for the variables—percentages, means, ratios, or differences. Procedures for calculating standard errors of selected types of estimates can be found in several texts (Kalton, 1983; Lee & Forthofer, 2006; Levy & Lemeshow, 2008). Listed below are formulas for the standard error of the mean and the standard error of percentage (Aday & Cornelius, 2006).

The standard error of the mean (SE_M) for a simple random sample is calculated by dividing the standard deviation for the sample by the square root of the sample size:

$$SE_M = \frac{s}{\sqrt{n}}$$

The standard error of percentage (SE_p), or proportion for a simple random sample, is:

$$SE_p = \frac{\sqrt{p \times 100 - p^2}}{n}$$

From Step 2, suppose Amy chose to use percentage or proportion to summarize her variables. Consequently, the standard error of a percentage estimate of 50% for a sample of 100 cases would be:

$$SE_p = \frac{\sqrt{.50 \times (1.00 - .50)}}{100}$$

$$SE_p = \frac{\sqrt{.50 \times .50}}{100}$$

$$SE_p = \sqrt{.0025}$$

$$SE_p = .05$$

This means that for a sample of 100 cases, the standard error of the estimate of 50% ($p = .50$) is 5% (.05). She can be confident that 95% of the time the true value of the population was between 40 and 60% (almost two standard errors above and below). Remember that 68% of the time, the true value would lie between 45 and 55% (one standard error, and in this study that is 5% above and 5% below).

Step 5: Expected Estimate. Health survey researchers should use other studies or theories to arrive at an expected value of the estimate.

Step 6: Tolerable Range of Error in the Estimate. Needless to say, the figure derived in Step 5 is an estimate and subject to error. All investigators must decide what would be a reasonable range of error. The range varies with the precision needed in reporting survey results (for example, $\pm 10\%$ versus $\pm 5\%$). Keep in mind that the more precise the estimates, the greater the need for a larger sample size.

Step 7: Level of Confidence. Herein the researcher establishes the level of statistical confidence—99%, 95%, or 90%. The confidence level chosen is used with the formula for the relevant standard of error estimate to obtain the minimal sample size to have that level of statistical confidence. In our case example, Amy selected the typical 95% level of confidence with a ± 1.96 standard error around the estimated population value. Her calculations would be:

$$\begin{aligned} 95\% \text{ confidence interval} &= 1.96 \times \text{standard error} \\ &= 1.96 \times \frac{\sqrt{p \times (1.00 - p)}}{n} = .05 \end{aligned}$$

where, therefore,

$$\begin{aligned} n &= (1.96)^2 [p \times (1.00 - p)] / (.05)^2 \\ n &= 3.84 (.50 \times .50) / .0025 = 384 \end{aligned}$$

The size of her sample for this illustration is 384. However, as the following steps show, this can be modified for several reasons.

Step 8: Estimated Sample Design Effect. The sample size, n , derived from the formula in Step 7 pertains to a simple or systematic random sample design. More complex designs generally require larger sample designs. Kalton (1983) discusses the effects of various sample designs on the variances (standard error squared). For example, in cluster sampling the design effect is 1.3, which means that the standard error for any estimate based on the sample is 30% higher than that derived from a simple random sample. Consequently, if Amy were to use a cluster sampling, she would need to increase her sample size to 499 (384×1.3). In contrast to cluster sampling, stratified designs usually require fewer subjects than a simple random sampling. This is because the strata are less diverse (i.e., more homogeneous). “The net result of taking the weighted average of the standard errors of these relatively homogeneous strata is that the standard errors for a stratified design will be less than those that result from a simple random sample of the same population” (Aday, 1989, p. 116).

In our example, Amy plans to conduct a simple random sample so her estimate of 384 respondents remains.

Step 9: Response Rate Adjustment. Unfortunately, almost any health survey will have less than a 100% response rate. This step allows the researcher to adjust the size of the sample to accommodate the response rate. For example, Amy feels that she will have a 65% response rate. The adjustment is determined as follows:

$$n = 384 / .65 = 591$$

Step 10: Adjustment for Expected Proportion of Eligibles. At this step, the number of respondents determined from Step 9 are divided by the expected portion of respondents who will actually be found eligible, once they are contacted. Amy estimated that 95% will be eligible, with the revised sample size being:

$$n = 591 / .95 = 622$$

Step 11: Cost Computation. The researcher should repeat these steps for each of the major estimates to be analyzed. The resulting range of sample sizes can be reviewed, as can the cost for each sample size. The final size selected can be based on the number of ideal respondents, the budget, and what compromises can be made. Costs are determined by multiplying the dollar amount per respondent by the total number of respondents. Herein, Amy estimated her cost at \$55 per respondent (if she uses the telephone), with a total cost of \$34,210.

In summary, Amy found that she needs approximately 622 respondents to have 95% confidence that the hypothesis that about 50% of the people in the city-county area have been overweight in the past year is true, should the value for the sample drawn from the area fall between 45 and 55%.

Additional Sample Size Considerations. Two key elements that play a role in sample size are design, specifically stratified random sampling, and weighting where necessary.

Stratified Random Sampling. Stratified random sampling presents several issues in regard to sample size. As noted in the discussion of sampling types, the sample within each stratum is drawn randomly and as such can be considered an independent sample of the population stratum. Stratified random sampling is such that a very heterogeneous population can be subdivided into several relatively homogeneous strata, each demanding a fairly small sample.

Investigative studies involving a single dichotomous stratification parameter (urban-rural, smokers-nonsmokers, private hospitals-public hospitals) with random sampling in each stratum may employ a formula to determine sampling size. The formula considers confidence level and sampling error in calculating a representative sample size:

$$N = (ze)^2 (p)(1 - p)$$

where N = sample size

z = the standard score corresponding to a given confidence level

e = the proportion of sampling error in a given situation

p = the estimated proportion or incidence of cases in the population

Confidence level indicates the probability that the sample proportion will reflect the population proportion with a specific degree of accuracy (sampling error is designated as e in the formula). With a 95% confidence level, $z = 1.96$; whereas with a 99% confidence level, $z = 2.58$, and with a 90% confidence level, $z = 1.65$.

Suppose that a health researcher decided to investigate patient education programs in public and private hospitals in West Virginia. In ascertaining the sampling frame, it was found that private hospitals accounted for 25% of all hospitals in the state. The proportion of private hospitals in the population of all hospitals would be .25 ($p = .25$).

Employing the usual standard of a 95% ($z = 1.96$) confidence level and a sampling error of .10, the following calculations apply:

$$N = (1.96/.10)^2(.25)(.75)$$

$$N = (19.6)^2(.25)(.75)$$

$$N = (384.16)(.1875)$$

$$N = 72$$

As a point of interpretation, a sample size of 72 private hospitals would give representativeness with no more than a plus or minus .10 sampling error with a confidence limit of 95%.

However, a stratified random sampling is often disproportionate in that a greater proportion is sampled in one stratum than another. The two major reasons for this are (1) differences in population size and (2) differences in homogeneity among strata. As an illustration, an investigation was conducted in West Central Illinois to ascertain the relationship of religion and health habits. The population comprised 1,000 Protestants, 800 Roman Catholics, and 200 Mormons. It is evident that a proportionate sample would leave the Mormons misrepresented or at least without a statistically functioning sample. Further, the decision to work with a sample of 100 from each of the religious strata revealed that the odds of being randomly selected varied tremendously. The Protestants had a 1 in 10 chance of being chosen, Catholics had a 1 in 8 chance, and Mormons had a 1 in 2 chance.

One way around this dilemma is to use weighted sampling. With this procedure the additional problem of combining subsamples (strata) into one overall sample for the purpose of data analysis can be overcome. Weights are assigned to each of the strata. Mormons are given a weight of 2 because they have a 1 in 2 chance of selection. Catholics have a weight of 8 because they have a 1 in 8 chance of selection, and similarly Protestants have a weight of 10. To make calculations more workable, each weight is divided by 2 to arrive at smaller numbers. This provides weights of 1, 4, and 5 for Mormons, Catholics, and Protestants, respectively.

The researcher can weight the data during analysis using the appropriate weight for each stratum. For example, suppose the following unweighted data shown in Table 7.4 were obtained. To obtain proportional data, each stratum would be weighted by the appropriate amount. The result would be as seen in Table 7.5. The figures are changed, but

Table 7.4 Distribution of Smokers by Religion

Smoking	Religion		
	Mormon	Catholic	Protestant
Cigarette	2	25	26
Pipe	10	21	34
Nonsmoker	88	54	40
Total	100	100	100

Table 7.5 Weighted Distribution of Smokers by Religion

Smoking	Religion		
	Mormon	Catholic	Protestant
Cigarette	2	100	130
Pipe	10	84	170
Nonsmoker	88	216	200
Total	100	400	500

the relative values of the data are not altered. Thus, weighting provides adequate and equal representation of all strata. Aday (1989) provides a more detailed discussion of weighting.

In conclusion, there are several considerations in determining sample size, which are listed below:

1. Select as large a sample size as possible because the larger the sample, the smaller the sampling error.
2. Cost, in terms of money and time, and subject availability are legitimate concerns in ascertaining sample size.
3. Surveys require a greater sample size than experimental studies because of response failure, item omission, poor interviewing, and so on.
4. When a sample is to be subdivided into smaller groups for data analysis, a large enough sample is required to allow for statistical treatment within each subgroup.
5. Stratification allows for greater homogeneity, thereby requiring a smaller sample size than a simple overall random sampling.
6. Assigning weights is a technique applicable to disproportionate sampling.
7. Begin estimating sample size early in the design of the study.
8. Keep in mind that sample size estimates are for outcome data—not necessarily the number who should be enrolled. In other words, plan for dropouts and missing data.
9. The nature of the research (analytical or descriptive) is important in determining the method used to estimate sample size.

Case Discussion

Amy's task was to conduct a survey of residents using some of the questions from the NHANES. The survey could include items about eating habits, dietary supplements, weight history, and physical activity and fitness. She had to determine who she would survey, how many people she would survey, and be able to present data for the city and county separately as well as an aggregate. She examined several sampling techniques and understood that if she wanted to specify certain variables, such as gender, a stratified random sample would be preferred. In so far as sample size, she needed to

determine if her survey was descriptive, analytical, or a combination of both. If it contained an analytical component, then the concept of power analysis would need to be considered so that there would be sufficient power for her statistical tests. The illustration of 11 steps took this into consideration. Smaller numbers were used to facilitate understanding of these steps than were used in the stratified random sampling illustration. In considering cost, she was thinking of using a telephone survey but perhaps could decrease cost with a mail or Web survey. However, the return rate may be less. On the other hand, she could interview respondents with a greater return rate, but the cost would be much greater.

SUMMARY

Limited time, money, and resources frequently preclude a study of the entire population. Therefore, a sample of the population is required. The sample, however, must be of a size that represents the population and provides opportunity to detect significant differences when appropriate. The sampling frame is a list of all the persons from whom the sample is to be drawn. This can be difficult to obtain in some instances.

Sampling techniques can be subdivided into two broad categories: probability and nonprobability sampling. Probability sampling means that the likelihood of selecting each respondent is known; examples are random sampling, systematic sampling, stratified random sampling, and cluster sampling. In contrast, nonprobability sampling reflects an unknown probability of selection. Types of nonprobability sampling techniques include convenience, quota, dimensional, purposive, and snowball sampling. Depending on the nature of the research objectives, mixed sampling designs may be used.

Sample size was discussed from two broad perspectives: the first was based on power analysis while the second was seen from a health survey perspective. Considerations in determining sample size include sampling error; alpha probability level; and, for power analysis, probability beta level, and directionality. The standard alpha level of significance is .05, while power is generally established at $1 - \beta = .80$.

In regard to health surveys, 11 steps were outlined as considerations for sample size. These steps included:

- Identification of major study variables
- Estimates of study variables
- Subgroup selection
- Relevant standard error formula
- Expected estimate
- Tolerable range of error in the estimate
- Level of confidence
- Sample design effect

- Response rate adjustment
- Adjustment for expected proportion of eligibles
- Cost computation

The steps are like building blocks in that each step depends on the preceding steps for appropriate calculation. Additional considerations were stratified random sampling and weighting. A brief summary list of considerations was presented.

CRITICAL THINKING QUESTIONS

1. What are the differences among study population, sample, and sampling unit?
2. How does sampling unit affect statistical analysis?
3. What are the principal differences between probability and nonprobability sampling?
4. How does determination of sample size differ for analytical studies and survey or descriptive research?
5. What factors should be considered in determining sample size?

SUGGESTED ACTIVITIES

1. Imagine that you had a population of 931 people and that a simple random sample size of 75 is to be selected. How would you use a table of random numbers to select the first 10 people?
2. A population is divided into four strata (urban-rural, male-female) with the population sizes being 15,000, 8,000, 17,000, and 9,000, respectively. A sample of 500 is to be selected using proportional stratified random sampling. What is the sampling fraction, and what would be the number selected from each stratum?
3. You wish to conduct a survey on health education in your school district. You have a list of 6,125 households, and this is to be your sampling frame. The sample size is limited to 600. What is the sampling fraction? How would you go about selecting the sample? If you were encouraged to do a systematic sampling, how would it be conducted? Why might it be better than a simple random sampling?
4. Assume that you are conducting a satisfaction survey of patients who were hospitalized in your facility sometime over the last 6 months. You set the confidence interval at 5% and the confidence level at 95%, and the number of total patients is 4,250. Go to three different websites (use a search engine of your choice) to estimate the sample size. Compare all three website answers.
5. Repeat the 11-step process to estimate sample size by creating your own study, including a null hypothesis, establishing one or more variables, the population size, level of confidence, response rate adjustment, and so on.

REFERENCES

- Aday, L. A. (1989). *Designing and conducting health surveys*. San Francisco: Jossey-Bass.
- Aday, L. A., & Cornelius, L. J. (2006). *Designing and conducting health surveys: A comprehensive guide* (3rd ed.). San Francisco: Jossey-Bass.
- Bowling, A. (2002). *Research methods in health: Investigating health and health services* (2nd ed.). Buckingham, UK: Open University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cottrell, R. R., & McKenzie, J. F. (2005). *Health promotion and education research methods*. Boston: Jones and Bartlett.
- Graziano, A. M., & Raulin, M. L. (2000). *Research methods: A process of inquiry*. Needham Heights, MA: Allyn & Bacon.
- Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2007). *Designing clinical research* (3rd ed.). Philadelphia: Lippincott, Williams & Wilkins.
- Hurlburt, R. T. (2003). *Comprehending behavioral statistics* (3rd ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Kalton, G. (1983). *Introduction to survey sampling*. New York: Sage Publications.
- Lee, E. S., & Forthofer, R. N. (2006). *Analyzing complex survey data*. Thousand Oaks, CA: Sage Publications.
- Levy, P. S., & Lemeshow, S. (2008). *Sampling of populations: Methods and applications* (4th ed.). Hoboken, NJ: John Wiley & Sons.
- Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537–563.
- Norwood, S. I. (2000). *Research strategies for advanced practice nurses*. Upper Saddle River, NJ: Prentice Hall.
- Polit, D. F., & Beck, C. T. (2004). *Nursing research and principles* (7th ed.). Philadelphia: Lippincott Williams & Wilkins.
- Windsor, R., Clark, N., Boyd, N. R., & Goodman, R. M. (2004). *Evaluation of health promotion, health education, and disease prevention programs* (3rd ed.). Boston: McGraw-Hill.

Qualitative Research

KEY TERMS

coding patterns	official documents	photo libraries
inductive reasoning	outcomes	process
natural setting	participant observation	qualitative
nonpersonal documents	patterns	themes
observer's comments	phenomenological approach	

Case Study

Health Data Analysts, a research consortium, has been awarded a contract to examine the efficacy of a hospital's Wellness Center. Steven has been named the project director of this study, and he will head a team comprised of several staff people who are all trained and educated to gather and interpret data for the layperson. The consortium devised a methodology to use qualitative research in order to best reach the goals and objectives of the project.

Some of the previous chapters of this textbook have dealt with collecting and reporting data in a quantifiable manner. This chapter discusses another method of gathering data that is different from experimental and survey research in that it is **qualitative**. As an introduction to this chapter, Table 8.1 compares the two approaches of qualitative and quantitative research. After you review the table, you will be able to see that the need, setting, and type of problem to be studied will eventually determine the research approach you will use.

Characteristics of Qualitative Research

What is different about qualitative research? The following characteristics, adapted from Bodgan and Biklen (2007), help to describe some traits about this methodology:

1. *Qualitative data has the natural setting as the direct source of data, and the research is the key instrument.* The data are collected at the location of the study.

Table 8.1 Comparisons Between Qualitative and Quantitative Research

Qualitative	Phrases Associated With the Methodology		Quantitative
Case study	Naturalistic		Empirical
Documentary	Observation		Experimental
Ecological	Participant		Hard data
Ethnographic	Phenomenological		Positivist
Field work	Soft data		Social facts
Life history	Symbolic interaction		Statistical
Key Concepts Associated With the Methodology			
Common sense	Practical purposes		Hypothesis
Definition of situation	Process		Operationalize
Everyday life	Social construction		Reliability
Meaning	Understanding		Replication
Negotiated orders			Statistically significant
			Validity
			Variable
Academic Affiliation (beginnings)			
Anthropology			Economics
History			Political science
Sociology			Psychology
			Sociology
Goals			
Describe multiple realities			Establish the facts
Develop sensitizing concepts			Predict
Develop understanding			Provide statistical description
Test grounded theory			Show relationships between variables
			Test theory
Relationship with Subjects			
Empathy			Circumscribed
Emphasis entrust			Detached
Equalitarian			Distant
Intense contact			Short-term
Instruments and Tools			
Tape recorder			Computers
Transcriber			Indices
			Inventories
			Questionnaires
			Scales
			Test scores

Table 8.1 Comparisons Between Qualitative and Quantitative Research

Qualitative	Quantitative
Data Analysis	
Analytical induction Constant comparative method Induction Models, themes, concepts Ongoing	At end of data collection Deductive Statistical
Problems in Using the Approach	
Data reduction difficult Difficult to study large populations Procedures not standardized Reliability Time consuming	Controlling extraneous variables Obstrusiveness Validity
Design	
Design is a hunch as to how to proceed Flexible, evolving General	Design is a detailed plan of operation Specific Structured
Data	
Descriptive Field notes Official statistics Personal documents Photographs Subjects' own words	Counts, measures Operationalized values Quantifiable coding Quantitative data Statistical data
Sample	
Nonrepresentative Small Theoretical	Control for extraneous variables Control groups Large Precise Random selection Stratified
Methods	
Observation Open-ended interviewing Participant observation Reviewing of documents	Data sets Experiments Quasi-experiments Structural interviewing Structured observation Survey research

Source: Adapted from Bogdan, R., and Biklen, S. *Qualitative research in education* (5th ed.) Boston: Allyn & Bacon, 2007, pp. 44–46.

In our case study, Steven will go directly to the Wellness Center to study the actual goings-on firsthand. The researcher is really the instrument most readily used. Even if tape recorders or other equipment is employed, the researcher has the insight into where he or she should be and exactly how to collect the necessary data. The reason that qualitative researchers go to the location under study is that they are concerned with context and feel that situations can best be understood when they are directly observed. The setting has to be understood in the history context of the institution of which it is a part. When the data with which qualitative researchers are concerned are produced by subjects, as in the case of official records, the researchers want to know where, how, and under what circumstances the data came into being. Of what historical circumstances and movements are the records a part? Qualitative researchers believe that behavior is influenced by the setting and therefore always go to that location to collect the necessary data.

2. *Qualitative research is descriptive.* Numbers are not used to collect data, but rather words and pictures form the basic methods of data collection. The data include transcripts of in-depth interviews, field notes, photographs, tapes, memos, personal documents, and other official records. In our case study, Steven might ask the director of the Wellness Center for files pertaining to the goals and objectives stated when the center first opened.

Quotations are very often used in collecting qualitative data. In addition, a record is made of everything that occurs in certain situations. For example, when observing a conversation between two people, the researcher would probably describe the initiator (i.e., the person who did most of the talking and listening), the immediate surroundings (e.g., near a drinking fountain), and so on. The researcher is attempting to get a very comprehensive and deep understanding of the situation being studied. Therefore every detail must be described, and this is a very laborious task.

3. *Qualitative researchers are concerned with process rather than with outcomes or products.* The researcher is concerned with the natural history of the situation being studied. Questions related to how decisions are made in the context under study and to what becomes “common sense” are areas of concern. Qualitative studies tend to decipher exactly what goes on in an institution so that the expected outcomes are fulfilled. That is the **process** leading to the **outcomes**. In quantitative research, subjects are given tests (pretest and posttest) to determine the effectiveness of a program. The qualitative process discerns activities that would occur between the pretest and posttest and analyzes those events, with no concern for the outcome.
4. *Qualitative researchers tend to use inductive reasoning to analyze their data.* Qualitative investigators do not collect data to prove or disprove a prior hypothesis, but rather they collect the data first and then group them together. Glaser and Strauss (1967) describe a type of theory that builds from the bottom up as

grounded theory. The qualitative investigator puts together a theory after the data have been collected and after much time has been spent at the location with the subjects. Part of this process is to find out what the concerns are, as opposed to quantitative research in which investigators come into a situation with predetermined questions.

5. *Meaning is of essential concern to the qualitative approach*. Qualitative researchers are concerned with how different people live their lives and make sense of them; this is called *participant perspective*. For example, investigators might ask what people in a certain situation take for granted. In our case study, Steven may ask Wellness Center personnel for their perspectives on the efficacy of the center. He would ask other personnel in the hospital the same question. In other words, he would attempt to get the participants' perceptions about the Wellness Center and relate them to the perspectives of other people, looking for common ground.

Many situations in which qualitative research is conducted will not include all of the characteristics we have discussed, but they should include a majority. Investigators using this research method must have time and patience, as the study will undoubtedly demand much painstaking effort.

Theoretical Foundations

For any researcher to adequately collect and analyze data, he or she should be aware of and have an understanding of the theoretical foundation on which the research is based. The theoretical foundations of qualitative research are similar to those in anthropology and sociology, in which paradigms are used to guide research. A *paradigm* is a research perspective that holds views about how research is to be conducted and that has its own assumptions about how the world works and about what is important in that world. Most qualitative researchers use a **phenomenological approach**, which is the basis for most research in this area.

Phenomenological Perspective

Max Weber was the leading proponent of the phenomenological approach to research. The phenomenologist is concerned with attempting to understand human behavior through the eyes of the subjects in the study. This has been called *verstehen*, which is the interpretive understanding of human interaction. The phenomenological approach is used throughout most qualitative studies because of the importance of interviewing the subjects in a program or institution. Here the investigator has not made any presumptions about how the subjects view something and goes about conducting an informal interview without any structure. This perspective is ever present as a theoretical framework for qualitative researchers.

Symbolic Interaction

Symbolic interaction originated with George Herbert Mead in his book *Mind, Self, and Society* (1934). He viewed communication as the key to understanding the connection between intelligence (mind), self-consciousness (self), and the community (society). Gestures (verbal or not) made by people are symbols, taken to mean acts that stand for something else. Another aspect of Mead's work dealt with the fact that humans have a self-conscious awareness of themselves. Interaction between people depends on the degree to which it possesses a self-conscious quality. How others interpret the interactions depends on experience and history. In other words, symbolic interaction theory asserts that people's self-concepts are influenced by the way others respond to them.

People act not according to predetermined responses but as interpreting, defining, symbolic animals whose behavior the researcher can only understand by entering the defining process (Bogdan & Biklen, 2007). This is accomplished with a type of qualitative research called *participant observation*, which will be discussed later in this chapter. Defining is a shared event, and the people involved have usually developed congruent definitions of interpretations. As people see a need, they may change their definition of an interpretation, and this is where the qualitative investigator steps in to determine how definitions develop.

Culture

Culture anthropologists study other cultures, sometimes from a phenomenological perspective. *Ethnography* is the term used for the description of a particular culture. All anthropologists use the theoretical framework of culture in their research studies, and this organizes the ethnographic work. The ethnographer has few if any hypotheses, and there is no structured instrument with which to collect the data. The goal of the ethnographer is to describe in as much detail as possible the customs, religious ceremonies, mores, language, and other pertinent variables of a subculture or group. The best way to do this is for the investigator to become a participant observer and in so doing attempt to put aside his or her own culture.

Ethnomethodology

Harold Garfinkel (1967) coined the term *ethnomethodology* to refer to the study of how individuals create and understand life. It is the study of everyday, commonplace, routine social activity. Ethnomethodologists attempt to understand how people make order out of the complex world in which they live. A more complete discussion of ethnomethodology will appear later in this chapter because it is a type of qualitative research that has taken on importance in the last 40 years.

Methods of Qualitative Research

Qualitative methodologies are research procedures that enable the investigator to produce data. The methods that will be discussed include observation, **participant observation**, ethnomethodology, and document study. Although in-depth interviewing is also a qualitative research method, we chose to discuss it in Chapter 6, on survey research.

Observation

One of the primary methods of qualitative research is *observation*. It is a scientific technique if conducted under the proper circumstances. Observation must (1) serve a research purpose; (2) be planned systematically; (3) be recorded systematically; and (4) be subjected to checks and controls on validity and reliability (Bickman, 1981). The following section will describe the value and purposes of observation, methods of observation, what to observe, and observer training.

Value and Purposes of Observation Observational data are collected in a naturalistic setting in that the researcher does not manipulate or control people or other significant things related to the study. It is a discovery-oriented approach carried out in the field. Because the investigator is in the field, he or she can become very close to the situation and better understand the context within the program and its various complexities. Therefore, the value of observational data is that it enables those who asked for the information (the users) to understand the entire program through detailed and very descriptive information that is provided through the collection of observational data.

The collection of observational data may have three purposes: (1) to provide descriptions of behavior; (2) to record situational behavior; and (3) to study a topic that lends itself to this method. First, providing detailed descriptions of the behavior patterns of people is one of the purposes of health science research. The observational method of data collection enables researchers to accomplish this task. When observational data are recorded, it is done at the same time a behavior pattern is occurring. This allows investigators to get a true sense of individual and group behavior under real and accurate circumstances.

A second purpose for using observational data is that behavior can be recorded as it actually occurs. In our case study, Steven will use observation so that he and his staff can directly observe behaviors as they happen. In this manner, they will be able to observe how Wellness Center personnel interact with each other under varied circumstances and in several situations.

The remaining purpose of observational methods is that there are certain circumstances under which they are the *only* feasible method to collect the appropriate data. Infants and toddlers, for example, cannot be interviewed or given a survey to complete; hence, observation becomes the method used to collect data concerning these types of subjects. Another example would be a study of people with severe diseases (terminal cancer, schizophrenia), which is not possible except through observation.

There are several values or advantages of direct observation. Patton (2002) has best described the advantages of direct, personal observations:

1. By directly observing program observations and activities, the investigator is able to understand the *context* within which the program operates.
2. The firsthand experience with a program enables the experimenter to use the inductive approach.
3. The study personnel can observe things that are routine to those in the program.
4. The investigator can learn things about the program that cannot or will not be revealed in an interview or questionnaire.

5. The observers are able to present a comprehensive view of the program because they can move beyond the perception of the participants.
6. The investigator uses his or her knowledge and experience in terms of feelings, reflection, and introspection about a program.

Methods of Observation There are two major types of methods of observation: relatively unstructured and structured. In the former method, the investigator attempts to get directly involved in the situation and to describe it as nonselectively as possible. In structured methodologies, the investigator codes or categorizes the observed behaviors of the program participants.

Unstructured methods may involve being a participant observer, using a digital camcorder or recorder, using specimen records, and recording anecdotes. Because participant observer methods will be discussed later in this chapter, we will focus here on the unstructured methods.

By using a camcorder, one could ideally get a complete and accurate view of a program. However, is this really the goal of observation? Or is it to summarize, systematize, and simplify the event, rather than depict an exact replication (Bickman, 1981)? Even if one used a small camcorder, with an excellent and unobtrusive microphone, the result is not an exact reproduction because of biases introduced by the presence of the device.

Specimen records are descriptions of behavior over a brief continuous time period. They allow for extrapolations of one event to several or a series of events. Behaviors are noted with painstaking care, and the interventions of those observed are recorded so as to define a standing pattern of behavior. If these patterns of behavior can be observed under various environmental settings, a behavioral consistency can be determined. This is the major advantage of using specimen records.

Anecdotes are used widely by many people attempting to observe behavior. The observer selects places and particular events to observe before actually recording the anecdote. This is not true of specimen records or films. Anecdotal records are objective and usually written after the incident has occurred. This type of record can test hypotheses if proper sampling is used. In the previously mentioned methods, hypotheses are generated *after* the observations are made. Anecdotal records should not be interpretive, but merely descriptive and accurate.

Generally, unstructured methods lead to problems of reliability, observer bias, and memory distortion. Because these problems can damage any study, we suggest that unstructured methods be used to generate rather than to test hypotheses.

Structured methods are more formal methods used to observe behavior and to set up or test hypotheses. The investigator is able to select activities to observe before they occur and can plan a systematic recording of observations. There are several ways to record this type of information: duration, continuous, frequency-count, and interval.

Duration recording is used when the observer wishes to record the elapsed time during which the behavior occurs. In our case study, if Steven wanted to find out how long the coordinator of the Wellness Center talked during a staff meeting, he could use a stopwatch to accomplish this task.

Continuous recording occurs when the observer records all the behaviors of the subjects and thereby creates a *protocol*. A protocol is a narrative in chronological order of

everything that occurred in a given setting, such as the Wellness Center staff meeting. This is a very comprehensive method in that the observer must use a content analysis system to classify the observed behavior.*

When using *frequency-count recording*, an observer simply counts the number of times a particular behavior occurs. This is especially useful when behaviors occur at low frequency and observers can count several different behaviors at the same time.

Interval recording is used to study the sequence of behaviors of subjects. The observer records a specific behavior at specific intervals (e.g., every 10 seconds). If Steven were to record, at intervals, when the coordinator of the Wellness Center asked a rhetorical question, he could get an idea of the sequence of that behavior. In addition, if Steven had a frequency count of rhetorical questions and multiplied it by the interval, he could get the duration of that behavior, which could prove to be very important in diagnosing possible personnel problems.

What to observe. Many program aspects should be observed to get a comprehensive view of that program. We will discuss (1) program setting; (2) program activities and participant behaviors; (3) informal interactions and unplanned activities; (4) nonverbal communication; and (5) unobtrusive measures. Much of the following has been abstracted from Patton's *Qualitative Evaluative and Research Methods* (2002). These occasions can help an observer organize a methodology that will emphasize certain kinds of observations. These are called *sensitizing concepts*, and they provide a framework to enhance the importance of behaviors and events.

The *program setting* is the physical environment in which the research takes place. When the reader can visualize the setting through a complete and detailed description provided by the investigator, then the program setting is helpful. The researcher should avoid using interpretive words such as "very," "wonderful," and "lovely." Rather, words that actually describe the setting—colors, dimensions of space, or quotations of participants—should be used.

Program activities and participant behaviors are observed by asking questions such as: *What do the participants do? What is it like to be a participant? What do the observers see while the program is in progress?* Units of activity are generally regarded as organizers for the researcher. These units may include staff meetings, formal sessions, patient-client sessions, and the like. The investigator must focus the sequence of events in a chronological order. *When did the activity begin, who introduced it, and who is in charge?*

Gradually, the researcher attempts to observe each activity by asking questions that deal with statements made by staff and participants during the event, such as: *How did behaviors change over the duration of the activity? How did it feel to be engaged in that activity?* At the end of the activity, the observer asks: *What signals that the event has ended, what is said by whom, and what is the relationship of this particular activity to the other parts of the program?*

Observation of *informal interactions and unplanned activities* is just as valuable as viewing formal activities in a program. Investigators should ensure that time is allotted for this activity. It can occur during breaks or meals, before and after formal working

*For a more complete discussion of content analysis see Berg (2004).

hours, and during the workday. The researcher will probably overhear conversations or conduct face-to-face or small-group interviews. The way people interact or do not interact after a meeting or part of a program should also be observed. The fact that people do not interact is a part of the data and should be noted.

Nonverbal communication has received much attention by behavioral scientists, who of course include health scientists. When observing groups, noticing how people sit, what they do in their seats, how they dress, and how they space themselves in discussion groups (e.g., who sits next to whom and how often) enriches description of the process of a program. Investigators should describe the nonverbal cues of others, as well as their own reactions to those cues. And by watching for behavior patterns, they can learn about significant nonverbal behaviors.

Unobtrusive measures are helpful in obtaining data without the participants realizing that they are part of the study. Examples of this type of data collection include analyzing contents of wastepaper baskets and noting what is written on a laptop or memo pad. These unobtrusive measures are very helpful because, once people know that they are in a study, their reactions may become self-conscious and inhibited. Unobtrusive measures should not contaminate the way people respond. Additional ways to collect unobtrusive data include looking through directories, calendar diaries, and other such documents.

In our case study, Steven could observe reactions of people leaving a conference room, witness daily calendars to see the meetings scheduled, and look at internal memos. These would be unobtrusive measures and could possibly reinforce the results of other reactive data-gathering methodologies, such as surveys and interviews. The more varied the data-gathering techniques, the more congruence should appear among the results. This provides a more true and accurate picture of the program being observed.

The Observation Form There are so many types of observations and situations in which they take place that we encourage all investigators to prepare their own forms. Each time an observation takes place, a new form is required. Researchers should attempt to plan their observations so they can devise appropriate forms. Table 8.2 is an example of a form that Steven might utilize when observing a staff meeting of the Wellness Center.

Table 8.2 Observation Form

1. Every time the Wellness Center coordinator asks a question, place a check next to one of the following general categories that best describes the question:		
	Frequency	Total
a. Asks personnel for direct input	✓✓✓	3
b. Asks personnel to answer specific questions	✓✓✓✓	4
c. Asks for general questions	✓✓✓✓✓	5
d. Other	✓✓✓✓✓✓✓	7
Total		19

Source: Adapted from Gall, M., Borg, W., & Gall, J. (1996). *Educational research*. New York: Longman.

As is evident in the table, observation forms can be easily devised and, when utilized properly, can provide necessary and valuable information for the study.

Observer Training Many reviews on observer training (Berg 2004; Hartman & Wood, 1990; Thomas, 1993) provide helpful information on the training of observers. You should refer to these references for a more complete review. We will discuss the observation manual, observer orientation, training for the observational setting, and training in the observational setting.

The *observational manual* becomes the bible for all observers. The manual should discuss and explain ethical issues, dress codes, courtesy protocols, and all other matters that pertain to the collection of data by the observational method. The manual should clearly define all code categories, and positive and negative examples should be included to more explicitly depict the coding techniques. Health Data Analysts, the research consortium that employs Steven, would have a well-developed manual for use in most of the research projects the consortium conducts. Steven will have worked on this manual before contacting his staff.

In cases in which the observational method fits into the specific research project, *observer orientation* occurs when observers have been selected by the project director to be oriented to the purpose of the study. All observers must be encouraged to follow the coding system exactly and not allow other information to interfere with completion of the forms. The observers should not be told about any hypotheses, if any exist, because this knowledge might bias them when they are completing the necessary observation forms.

During orientation the observers must be informed about subjects' rights and the confidentiality of the study. Observers should not discuss their reactions with each other until the observations are complete.

Training for the observational setting includes having observers memorize the manual, especially the coding rules and definitions. This will eliminate confusion and disorganization when the actual observations occur. It might be worthwhile for Steven to ensure his staff's knowledge of the manual by having them practice and then demonstrate mastery on a test regarding the manual. Observers should be exposed to and trained in using the actual forms and any other equipment (digital recorders, etc.).

One way for training to occur is to place the trainees in settings that require them to use the materials that will be used in the actual study. This can be accomplished by having trainees watch videos and listen to digital recordings made by experienced observers. They enable the trainees to proceed from the simple to complex tasks, approximating an actual situation.

While the trainees are using the digital recorders, they should receive constant, consistent, and constructive feedback as to the accuracy of their responses. This can be done by comparing trainees' responses to experienced observers' responses and discussing the discrepancies. In addition, this method can be useful in reconstructing forms and/or categories of behavior on the forms.

After the training is completed and the trainees have demonstrated almost perfect accuracy in their response to these tools, they should begin training in the observational

setting. This is the final phase of the training and is conducted under supervision. Steven would take one of his staff to a very straightforward observational setting (behind a one-way mirror in an office) and encourage the staff member to begin making observational recordings. Afterward, Steven would review the staff member's forms to check for accuracy. Each staff member would go through the same type of procedure before attending larger, more complicated observational situations.

Participant Observation

Participant observation involves the collection of data in the field that combines document analysis; interviewing of respondents and informants; and direct participation, observation, and introspection (Denzin, 1989). The field of anthropology is best known for using participant observation; more recently, sociology, education, and health science researchers have used this method of qualitative data collection. The aforementioned disciplines have used participant observation in natural settings such as schools, hospitals, and clinics. Our case study would lend itself to participant observation as a methodology.

The participant observer becomes part of the setting and "goes native." The observation may take several forms and may vary as to the degree of the researcher's participation, how much is disclosed to the subjects in the study, and the degree to which the activities and subjects are directly observed by the investigators (Glesne, 2005). Researchers can become totally involved in the setting (e.g., become a staff member of the Wellness Center in our case study) or be a partial participant in a tribe or religious group without becoming an actual member of that group. When an observer "goes native," he or she wants to understand the values and experiences of that group. The participant observer must disregard her or his own values because they might cause the participant observer to become emotionally involved with the group. This becomes a difficult dilemma for the participant observer. He or she must share experiences of the group but cannot become totally involved because some sort of detachment must be retained in order to accurately report the observations.

How much is told to the subjects can vary from everything to nothing at all. At times it is necessary to conceal the fact that there is a study being conducted. This may result in some ethical problems. However, compromises are usually achieved by using partial disclosure. Here, only a few select people are informed about the participant observer. Junker (1960, pp. 35–38) has described four types of participant-observation situations:

1. **Complete participant:** In this role, the observer's activities are entirely concealed. The observer is a complete member of an in-group, thus sharing secret information guarded from outsiders. The observer's freedom to observe outside the in-group system of relationships is severely limited. Such a role tends to block perception of the workings of the reciprocal relations between the in-group and the larger social system and makes it difficult to switch from this to another role to observe the details of the larger group.
2. **Participant as observer:** Here, the observer's activities are *not* entirely concealed but are kept "under wraps" or subordinated to activities as participant. This role may limit access to some kinds of information, perhaps especially at the secret level.

3. **Observer as participant:** This is the situation in which the observer's activities as such are made publicly known at the outset, are more or less publicly sponsored by people in the situation studied, and are not kept under wraps. This role may provide access to a wide range of information, and even secrets may be given to the participant observer.
4. **Complete observer:** This describes a range of roles where, at one extreme, the observer is behind a one-way mirror and, at the other extreme, the observer's activities are completely public.

Participant observation must involve direct observation and is usually supplemented by other data collection methodologies. One of the most common complementary sources of data collection is the use of informants. *Informants* are a group of members who are in a position to reveal worthwhile information or who are wholly representative of the group under study.

Advantages and Disadvantages of Participant Observation The first advantage of participant observation is the ability of the process to *explore* a theory or a type of measurement. In addition, it allows for hypothesis formation where it is impossible to do so before the beginning of the study. A further point of exploration lies in the investigator being able to research a new area as a participant observer.

A second advantage to participant observation is that it allows investigators to gain *access* to subjects or data where it might not otherwise be feasible. Organizations that might feel threatened, such as in our case study, would be prime candidates for participant observation. This methodology would enable a researcher to gain valuable data as an insider of the organization. A further consideration is that in some instances subjects are unable to recall events or may not view events as important, thereby not giving the investigator accurate, reportable information. Additionally, participant observation becomes advantageous when subjects cannot self-report data, as in the case of very young children, impaired persons in hospitals, or those who are afraid to self-report data (prisoners, gang members, etc.).

The third major advantage to participant observation is that it offers the possibility of obtaining a *richness* of data. While other methods provide hard data in terms of numbers and statistics, participant observation enables the researcher to see variables within the context of the natural setting. Because subjects are off-guard, the descriptions garnered are accurate as to what exactly occurred.

As with any methodology, there are both pros and cons; there are several disadvantages with participant observation. The most serious of these is the ethical element, especially when no one under study is apprised of the investigation. The problem then is one of deception and of the reactions of subjects who were, in effect, duped. Another disadvantage of participant observation is the possibility that the participant observer will become too emotionally involved, lose objectivity in reporting, and then later provide personal interpretation to the data. A third disadvantage lies in the reliance on the participant observer's memory to recall all aspects of the events that occurred. This can be a slow and arduous process because the observer must covertly write or dictate notes whenever feasible.

Being a Participant Observer Many of the considerations previously discussed concerning validity, reliability, sampling, and subject selection must be adhered to in participant

observation. However, as Walizer and Wienir (1978) discuss, some special concerns pertain to participant observation.

Selecting the problem is a consideration for participant observation. Investigators may have relied on previous research to pose the problem or even have used participant observation to find the problem. There are some major things to look for when observing a particular group that guide the problem selection. The following has been adapted from Walizer and Wienir (1978, p. 338):

1. How is the institution or program organized?
2. What is the nature of the social relationships that exist within the program?
3. What types of technology are utilized to make the work environment plausible?
4. What are the relationships between management and staff?
5. What activities do employees do together? Apart?

These questions are just a few examples of how participant observers begin to construct a framework for their study. These questions naturally lead to others and thus can readily focus the problem to be studied. There are times when participant observers are given the problem before embarking upon the research.

A second consideration is *choosing the setting* for participant observation. If the problem has previously been delineated, then the participant observer must choose an appropriate site that would contain that problem. For example, if one wanted to study the relationship between secretaries and middle managers, the chosen site must include both these characters, and both must have expressed the desire to participate in the study. Another aspect to consider in choosing the setting is to ensure that the participant observer will be comfortable in the setting. If you were asked to be a participant observer in a nuclear power plant and did not feel at ease in the situation, the study would not benefit from your participation.

Another consideration is *establishing social relationships* with the subjects. This is the most important part of the design because the entire concept of participant observation relies on acceptance of the observer by group subjects. Before the study commences, it would be wise to obtain the necessary approval from high-level employers, presidents of corporations, chieftains of tribes, and so on. In our case study, Steven would get permission from the hospital director and clinic president if he wanted to conduct part of the research by using participant observation.

Once the study is in progress and the necessary permissions have been granted, the participant observer should begin to become associated and acclimated to the program. In this regard, the observer should attempt to remain in the background and not attract attention. This prohibits people from being too curious about the "new person on the block."

A fourth consideration in attempting to become a participant observer is *finding informants*. These persons are used to observe for the participant observer and to suggest to and inform the participant observer about the program and its problem or problems to be studied. The observer must make sure that the informant is reliable and relaying truths about what he or she sees and hears. The informant should be tested by the participant observer to verify his or her comments and perceptions.

A further consideration for the participant observer is *establishing rapport* with the subjects. The goal of the investigator is to blend in with the program and act as natural as possible. Bogdewic (1999) has delineated ways in which a participant observer may establish rapport:

- **Be unobtrusive:** You should be more observer than participant. Your behavior and attire should not draw attention, as your goal is to learn what it takes to fit in.
- **Be honest:** People in the setting you are studying will have a limited understanding of why you are there. Questions about your interests and what you hope to find should be dealt with in an open and direct manner. You should assure people that their participation is voluntary and that their identities will remain anonymous.
- **Be unassuming:** Try not to threaten the subjects by your expertise in technical and professional matters. You can downplay your expertise.
- **Be a reflective listener:** This communication skill helps build rapport and is an excellent way to learn the language of the participants. Particular words in a social situation have significance and meaning to the subjects.
- **Be self-revealing:** Participants will have some degree of curiosity about you, and your willingness to discuss common interests and life experiences can open the door to a more trusting relationship.

Participant observation is absolutely demanding and can take a great deal of time, patience, and effort. While the observer is spending the day with the subjects in a group, the other waking hours are spent recording, collating, and analyzing data. Hence, participant observation can be a very time-consuming activity, especially if it persists for a long period of time. However, the rewards are great in that participant observation, as a qualitative research method, gives the richness and completeness needed for obtaining information and drawing conclusions concerning some research problems.

Ethnomethodology

Ethnomethodology is the study of methods used in everyday, commonplace, and routine social activities. Garfinkel (1967) defines ethnomethodology as an organizational study of a person's knowledge of his or her ordinary affairs, or of his or her own organized enterprises, where that knowledge is treated by investigators as part of the same setting that it also makes orderable. It should be made clear here that ethnomethodology is not an alternative methodology aimed at a more effective solution of traditionally formulated problems. Focusing upon the complicated character of action scenes, ethnomethodology necessarily develops a style of research responsive to its subject matter. In other words, ethnomethodology is *not* a research method per se, but rather it is a method to attempt to find out how people make sense out of ordinary situations in which they live. Ethnomethodologists, then, examine common sense in an attempt to understand how people see, describe, and explain order in the world in which they live.

As an example, Wieder (1974) explored how narcotic addicts in a halfway house used a "convict code" (e.g., "do not snitch," "help other residents") to explain and account for their behavior. He illustrated the way in which residents "tell the code"

by applying maxims to specific situations when they are called upon to account for their behavior. He wrote: "The code, then, is much more a method of moral persuasion and justification than it is a substantive account of an organized way of life. It is a way, or set of ways, of causing activities to be seen as morally, repetitively, and constrainedly organized" (p. 158). This is an example of how ethnomethodologists suspend their own commonsense assumptions to study how common sense is used in everyday life.

Advantages and Disadvantages of Ethnomethodology The *advantages* of ethnomethodology are many. The first is that it studies nonverbal as well as verbal behavior. Second, it is longitudinal because it is ongoing, and changes in behavior can be viewed over a long period of time. Third, this type of study can provide insight into what and why people think about commonplace activities and behaviors, thus enabling health scientists to make better order of why people behave the way they do in areas of health.

One of the *disadvantages* of ethnomethodology in relation to the health sciences is that this type of study involves investigating the *process* of how something occurs, rather than the *product* of that occurrence. As an example, in our case study Steven would not use ethnomethodology to ascertain the attitudes of Wellness Center personnel toward centralized management, but would use ethnomethodology to study the process of how those attitudes were formed. Another disadvantage is that ethnomethodology does not lend itself to large-scale studies but is better for process-oriented, smaller investigations.

An interesting point to include here is that ethnomethodology actually studies all the previously discussed methods as a means to garner knowledge about the process of how people make sense of their commonplace lives. In this manner, researchers can gain valuable insights into questionnaire construction and coding of other survey materials.

Indexical Expressions Indexical expressions are situation-specific words and/or phrases whose meanings change from situation to situation and may depend upon who is uttering the word or to whom the remarks are directed. Garfinkel and Sacks (1970) listed the following indexical words: *she, we, he, you, here, there, no, this, that, it, I, then, soon, today, and tomorrow*. These words have varying meanings, dependent upon the context. The indexicals have to be interpreted by an individual who is participating in the interaction before the meanings of the words are clear. The ethnomethodologist does not want to convert these indexical expressions into objective, nonindexical expressions, as a traditional researcher might approach this situation. Instead, the ethnomethodologist wants to study the rules people set to make sense of these indexicals in everyday conversation. This can be a very important aspect of a research project: the interpretation of important words.

Another aspect of ethnomethodology is that conversation and interaction are regulated by rules or norms. Ethnomethodologists discover how sense is made out of the structuring and ordering of indexicals. They can put meaning to indexicals that are made clear through a situationally specific process in which the context may be problematic and differ from place to place or time to time. However, the rules by which meanings are

explicated remain objective, constant, and nonproblematic (Bailey, 1994). Topics that are usually studied by ethnomethodologists are:

1. Formulating, or the process by which one conversationalist interprets or explains a part of a conversation
2. Sequencing of a conversation
3. Terminating of a conversation

All of these and many other topics provide invaluable information to the health scientist, especially when conducting a qualitative research study. Experimental and survey research results may tell us what the subject knows or thinks, but ethnomethodology takes it one step further in an attempt to find out why the respondent answers in particular ways. Ethnomethodologists are indeed interested in human behavior in that they seek to find the rules that govern behavior.

Document Study

We shall discuss the study of documents with specific reference to **nonpersonal documents**. Personal documents will be discussed in a later section of this chapter, under *Techniques of Collecting Qualitative Data*. A very valuable source of information is retained in a program's or institution's records and documents. The investigator will have a better understanding and increased knowledge about a program once he or she reads its documents. At the outset of the project in our case study, Steven should negotiate receiving at least the following types of documents: routine client records, correspondence from and to staff, financial charts, and official or unofficial documents generated by or for the program (Patton, 2002). These documents can avail the investigator of basic sources of information regarding the activities and processes of the organization, and they can enable the researcher to view other questions not previously considered to follow up on observations, participant observation, or ethnomethodological research.

Types of Documents There are many official documents that the researcher should attempt to obtain: minutes of meetings, memos, newsletters, policy documents, code of ethics, philosophy statements, and the like. The qualitative researcher is looking for how the organization or program is defined by those who are involved in that program. Therefore, a review of these documents will prove beneficial.

Internal documents include memos and other communications that abound in any organization. In a hospital, such as our Wellness Center, the amount of paper that flows from the top to the bottom is immense. Of course, some flows in the opposite direction as well. These documents can reveal the true chain of command, the interoffice fighting and subsequent negotiation, and the rules and regulations. In addition, leadership styles might emerge from these documents. It is advised, as we stated in the beginning of this section, that the researcher ensure that he or she will have access to this information before the project actually commences.

External communication includes those materials that are circulated outside the organization. This would include newsletters, public philosophic statements, news releases, marketing advertisements, and public access programs (e.g., health fairs, open

houses). These documents can provide official points of view and insight into administrative hierarchy. Many organizations, especially hospitals, have hired public relations firms. This makes it a little difficult to ascertain who wrote what. In any case, the researcher should obtain the necessary information regarding the public relations organization, such as whether all external documents get reviewed in-house previous to publication. These documents are readily accessible because they are produced for outside consumption. In many instances, files will be kept of these types of documents, so they will be very easy for the researcher to obtain.

Personnel records can provide valuable information about hiring and firing practices, promotion and reward systems, and administrative policies regarding personnel. In addition, these files also can provide information about the people (e.g., personnel managers, supervisors) who add paper to the file. Access to these files may be granted if researchers agree not to identify people but instead use a coding system to describe the content of the files.

Document study can provide a well-rounded view of the organization or program that the researcher is studying. The aforementioned methods of qualitative research—observation, participant observation, ethnomethodology, and document study—all enhance the researcher's ability to gain an understanding of the events and activities of a particular program or an organization. A good qualitative study would most likely put to use the majority of these methods because they will prove to be beneficial. The next part of this chapter will discuss how, using various techniques, Steven may actually collect the necessary data concerning the efficacy of the hospital's Wellness Center.

Techniques of Collecting Qualitative Data

There are several techniques to use when you are collecting qualitative data. These include field notes, subjects' written words, photography, and official statistics. The following sections briefly describe these techniques.

Field Notes

Field notes are the most important part of collecting data in qualitative research. They are used in *all* methodologies: observation, participant observation, ethnomethodology, and document study. Field notes contain everything and anything that the observer feels is worth noting. Any information that will enable the observer to gain a better understanding of the program must be noted down immediately. If one leaves observation to memory, one is leaving much to chance.

Descriptions are the basics of field notes. Basic information as to who was there, what was happening, and where the observation took place should be included in the notes. In addition, a description of the event, what took place, and the interactions between people should be recorded. Field notes also should contain quotations of what people said. In addition, the notes should reflect the observer's feelings and reactions to the experience, as well as the meaning and significance of the event. Furthermore, the notes should contain the observer's insights, interpretations, beginning analyses,

and working hypotheses regarding the situation (Patton, 2002). These comments are noted by using observer's comments.

Lofland, Shown Anderson, and Lofland, (2006) have some suggestions for those writing field notes:

1. Record the notes as soon as possible after the observation.
2. Discipline yourself to write notes quickly, and reconcile yourself to the fact that the recording of the notes can be expected to take as long as the actual observation.
3. Dictating rather than writing is acceptable, but writing may have the advantage of stimulating thought.
4. Typing field notes is preferable to handwriting because it is faster and easier to read.
5. Make at least two copies of the field notes. One original copy is retained for reference, and other copies can be used as rough drafts to be cut up, rewritten, and reorganized.

Content of Field Notes Two types of materials are included in the field notes: descriptive and reflective (of the observer's ideas and concerns). The *descriptive* part of the field notes includes the following (Bogdan & Biklen, 2007):

1. **Portrait of the subjects:** Include their physical appearance, dress, style of talking, and mannerisms.
2. **Reconstruction of dialogue:** The conversations between subjects, as well as what subjects might say to the observer, are recorded. Use direct quotations, especially when they are unique to the setting.
3. **Description of the physical setting:** The observer should draw a diagram of the furniture arrangements and where people are sitting. Note also any blackboard writing and what may be on bulletin boards.
4. **Accounts of particular events:** Note who was involved in the event, in what manner, and the nature of the action.
5. **Depiction of activities:** Include detailed descriptions of behaviors.

As we discussed previously, the *reflective* part of the field notes should be recorded as well as descriptions of events, activities, and behaviors. These notes should be designated by "OC" (**observer's comments**). Bogdan and Biklen (2007) offer the following comments on what should be included in the reflective part of the field notes:

1. **Reflections on analysis:** Speculate about what the observer is learning, emerging themes, patterns, and additional ideas.
2. **Reflections on method:** Include comments on the study design, accomplishments, and plans for what to do next.
3. **Reflection on observer's frame of mind:** Observers may have preconceived notions about the subjects. When these are changed or reinforced, they should be noted.
4. **Points of clarification:** Make additional notes to add or clarify a previous notation.

Field Note Form Observers should adopt a standard form for their field notes. They should include at least the following:

1. **Title page:** This should include the date, time, and place of observation, as well as when the notes were recorded. The title of the event may also be included here.
2. **Diagram of setting:** As mentioned previously, a diagram of the activity or event should be included at the beginning of the notes.
3. **Wide margins:** They are necessary because the observer or someone else might want to make appropriate comments on a hard copy.
4. **Paragraphs:** New paragraphs should be formed very frequently to correspond with each new person speaking or with every change in the setting.
5. **Quotation marks:** They should be used as often as possible. Even though the observer may not quote exactly, if the notation is very close to an exact replica, enter the quotation marks.

Field Note Techniques There are various ways to record field notes, but the observer must find the technique most comfortable for herself or himself. Most people use laptop computers, MP3 players, or digital recorders. Using the latter two, the researcher can place the information from the equipment directly onto a computer. Be sure that the equipment the observer utilizes does not interfere with the natural workings of the subjects within the group. In other words, the equipment must be unobtrusive. Using a pen and paper on site and then rewriting or embellishing the notes later on is a bit more labor intensive but may be less intrusive, depending on the situation.

Another mechanism for recording field notes is called the Stenomask or Sylencer. It is a sound-shielded microphone that can be attached to any recording device (Patton, 2002). The goal of the Stenomask is to allow a person to speak and not be heard by other people, and it also keeps the background noise away from the microphone. The Stenomask can be plugged into any recorder or computer and any other device that has a microphone.

Field notes provide the basis from which qualitative data is recorded and then analyzed. Recording field notes is a tedious task that requires patience and long hours after the observations have been completed. It is necessary that notes be organized and clear so that all involved in the project may benefit from the effort.

Personal Documents

Much information can be obtained from studying people's personal effects. This may include their clothing and the furnishings in their home or office. What may be of greater significance to qualitative researchers is examination of documents that indicate how people lead their everyday lives. These include calendars, diaries, letters (personal and business), autobiographies, scrapbooks, books read, and poetry written. Thomas (1923) was an early proponent of utilizing personal documents to make inferences about people's lives. He believed that autobiographies, letters, and the like were an important source of data because they were capable of presenting life as a connected whole and of

showing the interplay of influences on individuals. Thomas's research focused upon immigrants at the beginning and later analyzed their diaries, letters, and other personal documents. From these analyses, he was able to depict some central themes of these people: individualization, demoralization, and deregulation.

The use of personal documents usually does not allow for a large sample size, although the Thomas study, as just described, must be considered an exception. Using personal documents enables the qualitative researcher to make a generalization about a subject and then use excerpts from the documents to illustrate them. In our case study, if Steven believed that the coordinator of the Wellness Center was autocratic, he would search the personal documents of that coordinator to look for a sense of autocratic personality.

How to Obtain Personal Documents Obtaining personal documents such as diaries, autobiographies, and letters is not an easy task. One way that was used by Thomas and Znaniecki (1927) was to advertise in a newspaper to find the appropriate materials. Some people might be willing to share diaries and letters. In our case study, organizational files would be very helpful. These personnel files should contain references to people in the organization, indicating how people define themselves in their respective positions.

The written words of subjects can be of invaluable assistance to the qualitative researcher who is attempting to gain an understanding of how and why a system works. Although obtaining these personnel documents may be difficult, it allows the investigator to exercise her or his imagination.

Photography

Qualitative research can be greatly enhanced by photography. The following section is taken from Bogdan and Biklen's (2007) excellent work on the benefits of photography in collecting qualitative data.

Social scientists have used photography since the nineteenth century to depict social documentaries on how people live (Riis, 1980; Stott, 1973; Thomson & Smith, 1877). The adage "a picture is worth a thousand words" has been adhered to by social scientists utilizing photography (Becker, 1978; Wagner, 1979). The advent of photography has enabled researchers to study aspects of life that cannot be researched through other approaches—images are more telling than words. There are two categories of photographs that qualitative researchers may use: found photographs (pictures others have taken) and those that the researcher has produced.

Found Photographs Many organizations have archives of photographs depicting groundbreaking ceremonies, outings, and other events germane to an organization. Newspapers usually have **photo libraries** as well as book libraries. In addition, county offices will have aerial photographs of land.

Photographs can reveal factual information that may shed light onto organizational structure. Parties may have been photographed, thus depicting who was there, seating arrangement, mistress or master of ceremonies, and general ambiance.

A photograph is like all other forms of qualitative data: to use it, the investigator should place it within the proper context and understand what it is capable of telling.

Photographs can represent the photographer's point of view, a superior's orders, or even the subject's demands. This can present valuable information in that when photographs are studied, clues are ascertained as to what people value by the images they prefer. These add to the evidence that the other qualitative techniques have enabled the investigator to gather.

Photographs may present anomalies: images that do not fit the theoretical constructs that the investigator has been forming. This can enhance the researcher's analysis and insights and even alter preconceived notions about the subject. Researchers may utilize photographs to discern how people define their world: what people take for granted, what they assume is unquestionable, and what organizational assumptions exist.

Researcher-Produced Photographs Investigators can collect factual information by using photographs to depict utilization of facilities. The technique requires the use of hidden cameras, but it is a technique that is acceptable. Participant observers take photographs in the course of the study so they can more accurately recall events, activities, people, what they were wearing, seating arrangements, and so on.

There are the usual advantages, disadvantages, and cautions when researchers are about to use photography as a data-collection technique. In certain instances picture taking may inhibit the establishment of rapport (the researcher appears to be an outsider), and there are other occasions when it is advantageous to developing rapport (other cultures may feel a sense of pride at being asked to pose—this serves as a discussion tool). We recommend that you do not take pictures at the very beginning of a project but wait until you have established rapport; you might even wait until subjects themselves start snapping—thus giving you a chance to join in the photography session.

Photographs, whether found or taken by the investigator, can serve as a tool for enabling the researcher to better understand the values and inner workings of the organization or program being studied. Photography can be a researcher's tool as well as a cultural product.

Official Statistics

Quantitative data that have already been collected can serve to help qualitative researchers in that the data may suggest trends (e.g., the Wellness Center has had five coordinators in 4 years) and also provide descriptive statistics (e.g., age, sex, race, and socioeconomic status of Wellness Center staff). In addition, hypotheses may be broadened and/or delineated dependent upon what the official statistics delineate.

Qualitative researchers tend to be critical of quantitative statistics because they are asking questions that cannot easily be answered with numbers: "Rather than relying upon quantitative data as an avenue to accurately describe reality, qualitative researchers are concerned with how enumeration is used by subjects in constructing reality. They are interested in how statistics reveal subjects' common-sense understanding" (Bogdan & Biklen, 2007, p. 154).

Types of Official Statistics There are several different kinds of **official documents** that provide statistics for researchers. They include census documents; health statistics provided by insurance companies; statistics provided by voluntary health organizations; and

personality inventories provided by hospitals, schools, clinics, personnel departments, and social service agencies. For the purpose of brevity, we will discuss census documents and other directories that can provide data for the qualitative research.

Census surveys are conducted every 10 years in the United States. They supposedly enumerate every person living in the country. (The first modern U.S. census was taken in 1790.) These massive undertakings are computed to make sure that apportionment of government representatives remains accurate and truly representative and to provide data for a wide variety of research interests. The census asks general questions regarding age, sex, socioeconomic standing, number of people living in a household, and the like. It should be noted here that not all households are asked the same questions because the census, at times, uses sampling to gather some of its information.

Directories are another source of vital statistics that can be valuable to the investigator. Telephone directories and city directories usually have information that is needed and useful when conducting a study. City directories usually list the name, address, and occupation of a person, but these are not representative of the population because people have to agree to be listed in the directory. There are also professional directories, such as the *American Association of Sex Educators, Counselors, and Therapists Directory*. These are compendiums that may include brief biographies, which will be helpful to the researcher.

Occupational directories, such as the *Dental Association Directory*, can provide information that allows construction of indexes to compare cities. This has been especially useful in depicting utilization and placement of physicians. There are also a vast array of *Who's Who* directories, many with regional and professional classifications. In addition, county and state governments keep records that provide information concerning financial transactions. These include automobile registration, pet ownership, collected fines, and residential taxes.

Problems in Using Statistical Records There are several problems inherent in amassing voluminous amounts of data, which the qualitative researcher should take into account.

1. **Data collection methods:** For example, the reporting of deaths in a region is created locally and may not mean the same thing to the entire universe. Undercounting is another problem inherent in data collection.
2. **Ambiguous terms:** Categorical definitions that are used by governmental agencies are not always used in the same way by nongovernmental agencies. Boundaries of districts and regions may change as the population changes. In addition, technical definitions (for example, the U.S. census defines an urban area as a town of at least 2,500 people) do not always coincide with common usage of these terms.
3. **Bias:** Lists for directories are not assembled for research purposes and are not usually completed with the care of accuracy necessary for a research project. Professional directories might be incomplete because a fee was charged for inclusion, which causes a bias in the reporting.

Official records and statistics can be an invaluable aid for the qualitative researcher. Easy access to a great deal of information may enable an investigator to get a good picture or sense of an organization or of the place where that organization exists. As with

any compilation of large amounts of data, cautions must be observed in the use of these compendiums. While we have reported on only a few types of censuslike compilations, the U.S. government garners many special subtopic reports from the census and other mass surveys.

Analyzing Qualitative Data

After collecting reams of data, most of it containing voluminous words, the researcher asks, *How can I make sense of this mess?* What needs to be done then is to make coherent sense out of these many pages of words. The qualitative investigator will look in the data for **themes** and **patterns** that might help construct and/or support hypotheses.

Recognizing Themes and Constructing Hypotheses

Taylor and Bogdan (1998) offer the following suggestions for constructing hypotheses and recognizing themes:

1. **Read the notes carefully:** Read everything, even notes of minor incidents, very carefully. Write margin notes recording possible patterns and trends leading to themes and hypotheses.
2. **Construct typologies:** These are classification schemes that can be useful when formulating hypotheses. They are formed by denoting how your subjects classify people and behavior and the differences between and among subjects that allow them to be classified.
3. **Read the relevant literature:** Consult the professional literature to compare the literature findings with what is beginning to appear in your data. In addition, utilize the concepts, models, and paradigms of others.
4. **Code important conversation topics:** Code those conversation topics that keep recurring. (Coding will be discussed below.)

It is very important that the researcher be able to determine common threads and themes through analysis of the data. This process then leads to hypotheses formulation and support for those hypotheses. To accomplish what at times seems like an insurmountable task, the investigator must develop a coding mechanism by which to organize and assemble the data.

Coding the Data Once the notes have been reread, typologies constructed, and literature reviewed, an elaborate coding system must be devised. This system will serve to organize and assemble the mounds of data that have been collected. When the review of the notes commences, the investigator makes notes and begins to code as described in the previous section. There are several types of codes, including (1) descriptive and (2) explanatory. *Descriptive* codes do not require interpretation but indicate a class of phenomenon in the notes. As an example, in our case study, Steven would note “PERS” in the margin to denote “personality” wherever appropriate. This will enable him to quickly note where all relevant personality notes appear in the text.

Explanatory codes indicate where patterns or themes have emerged. Steven could use “PAT” (pattern) or “TH” (theme) to indicate an administrative style that he has seen as evident in the data. Codes, no matter which kind, tend to pull the data together and make sense of the field notes.

Creating the Code There are several methods to help the investigator create the code. Miles and Huberman (1994) suggest devising a *start list* previous to doing fieldwork. The list develops from the conceptual framework—the list of research questions, problem areas, and key variables that were determined at the beginning of the study. Usually, a master code is developed. For example, in our case study, “ADMST” might mean “administrative style,” and “AUT” (for “autocratic”) might be a subcode under ADMST. This list may have as many as 90 codes and should be kept on one piece of paper for easy reference. Table 8.3 is an example of a section of a code sheet. The first column denotes a descriptive level for the general categories, the second column indicates the code, and the third column keys the code to the research question from which it derives.

Table 8.3 Section of a Coding Sheet

Administrative Hierarchy (AH)	AH (DUB) (END)	2.1
Style = ST	AH-ST	2.1.2
Role = RO	AH-RO	2.1.3
Demeanor = DR	AH-DR	2.1.4
Subject Use = SU	AH-SU	2.1.5

Bogdan and Biklen (2007) suggest other items to be considered when determining coding subsections:

1. **Setting/context:** information on surroundings
2. **Definition of the situation:** how people define the setting
3. **Perspectives of subjects:** ways of thinking, orientation
4. **Ways of thinking about people and objects:** subjects’ understanding of each other, of outsiders, and objects that are included in their world
5. **Process:** categorizing events, changes over time, and flow
6. **Activities:** regularly occurring kinds of behavior
7. **Events:** specific activities
8. **Strategies:** tactics, methods, techniques, plays, and other conscious ways subjects accomplish things
9. **Relationship and social structure:** behaviors not officially defined by the organization
10. **Methods:** material pertinent to research-related issues

These **coding patterns** can enable the investigator to think about the categories for which codes are to be developed. This method has its advantages and disadvantages over

the start list. However, it is recommended that the researcher find the coding method that is easiest to work with and that will provide the most complete listing of codes to organize the data.

Data Organization Now that the investigator has developed a coding system, the mechanics of actually going through the data and organizing them becomes the main task. The first step is to number all the pages of data so that the location of sources is easily accessible. Numbering the data in chronological order of their occurrence is a good way to sequence the data.

A second step is to begin coding the data using any of the methods previously discussed or even one that the investigator may construct on his or her own. After the coding categories have been designed, the third step is to go through the data and mark each paragraph (or sentence) using the appropriate coding category. The fourth step is to sort the data.

The analysis of qualitative data can be a taxing and tedious task—and, at the same time, the most rewarding. Hypothesis construction and the recognition of themes are the primary reasons for analyzing the data. Coding of the data enables the researcher to organize the substantial amount of data collected. Once the data have been coded, they then have to be sorted so that themes, patterns, categories, and trends can become evident.

Using Computers in Data Analysis

With computer programs available to assist in analyzing qualitatively collected data, such tasks as those listed below can be made easier (Miles & Huberman, 1994, p. 44):

- Making notes in the field
- Writing up or transcribing field notes
- Editing: correcting, extending, or revising field notes
- Coding: attaching key words or tags to segments of text to permit later retrieval
- Storage: keeping text in an organized database
- Search and retrieval: locating relevant segments of text and making them available for inspection
- Data “linking”: connecting relevant data segments to each other, forming categories, clusters, or networks of information
- Memoing: writing reflective commentaries on some aspect of the data as a basis for deeper analysis
- Content analysis: counting frequencies, sequences, or locations of words and phrases
- Data display: placing selected or reduced data in a condensed, organized format, such as a matrix or network, for inspection
- Conclusion drawing and verification: aiding the analyst in interpreting displayed data and in testing or confirming findings

- Theory-building: developing systematic, conceptually coherent explanations of findings; testing hypotheses
- Graphic mapping: creating diagrams that depict findings or theories
- Preparing interim and final reports

The programs used in data analysis can be categorized into five major types (Weitzman & Miles, 1995):

1. Text retrievers (examples: Metamorph, Orbis, Sonar Professional, The Text Collector, Word Cruncher, ZyINDEX)
2. Textbase managers (examples: ask Sam, Folio VIEWS, MAX, Tabletop)
3. Code-and-retrieve programs (examples: HyperQual2, Kwalitan, Martin, QUALPRO, the Ethnograph, and HyperRESEARCH)
4. Code-based theory-builders (examples: AQUAD, ATLAS/ti, HyperRESEARCH, NUD.IST, QCA, and NVIVO)
5. Conceptual network-builders (examples: Inspiration, MECA, Meta Design, ScmNet)

Choosing the appropriate computer program is an important part of the data analysis. You should consult with professionals who have utilized these types of programs to determine which program or programs will best suit your research needs.

Case Discussion

Steven, the project director of the study to determine the efficacy of a hospital's Wellness Center, has a highly skilled team of professionals who are familiar with qualitative methodologies. They decided to utilize computer programs to assist them in coding and retrieving their data. The program they chose was Ethnograph, as they believed this program was most appropriate for their project. The project team utilized a variety of mechanisms for collecting the data as discussed throughout the chapter.

SUMMARY

Qualitative research is an approach utilized to collect data and report the findings. It differs from quantifiable methods by having as its main goals the description of multiple realities; development of sensitizing concepts; and understanding of a particular program, organization, or setting. The designs used are generally flexible, and they evolve as the study progresses. Data come in different forms: field notes, official statistics, personal documents, photographs, and subjects' written words. Methodologies used are observation, participant observation, ethnomethodology, and document study. The samples utilized are usually nonrepresentative and small. Analysis of the data is very time consuming because it cannot be simply added up and provide a definite answer; it must be analyzed by the researcher in a thoughtful and thorough manner.

There are several problems in using qualitative research methodologies. These include the fact that (1) the data reduction is difficult; (2) large populations are not easily studied using this approach; (3) the procedures are not stabilized; (4) reliability is subjective; and (5) the method is very time consuming. However, qualitative approaches are very useful when a researcher wants to study naturalistic settings such as schools, hospitals, organizations, and the like. In our case study, the consortium Health Data Analysis decided quite appropriately to utilize qualitative methods to study the efficacy of the hospital's Wellness Center.

CRITICAL THINKING QUESTIONS

1. Distinguish between qualitative and quantitative research methodologies.
2. Describe the importance of having theoretical foundations when utilizing qualitative methods.
3. Identify the advantages of direct observation. Are there any disadvantages? If so, what are they?
4. What are the four types of participant observation? Which one could best fit your research project?
5. Define ethnomethodology. Explain its advantages and disadvantages.
6. What are the documents you would most likely utilize in your research project?
7. Differentiate between descriptive and reflective field notes.
8. Do you believe photographs are useful in research studies? Defend your answer.
9. How can you construct themes and patterns when using qualitative research methods?

SUGGESTED ACTIVITIES

1. Devise a list of topics that could be studied utilizing the qualitative approach.
2. Conduct a literature search on one of the topics in your answer to Activity 1 and attempt to set up a coding system for those citations.
3. Describe, in detail, a situation wherein you could be a participant observer. What are the steps you would take, from the conception of the idea to the completion of the data analysis?
4. One of your assignments in your health evaluation class is to collect qualitative data about pharmaceuticals related to the relief of pain from arthritis. You are to use only the Web. Detail how you would go about collecting the appropriate data, and list the websites that were most beneficial.
5. Utilizing your library's online capabilities, create an annotated bibliography of the most recent (2008–present) sources concerning qualitative data collection for the health sciences. Which sources would you recommend? Why?

REFERENCES

- Bailey, K. (1994). *Methods of social research* (2nd ed.). New York: Free Press.
- Becker, H. (1978). Do photographs tell the truth? *After Image*, 5, 9–13.
- Berg, B. (2004). *Qualitative research methods for the social sciences*. Boston: Allyn & Bacon.
- Bickman, L. (1981). Observational methods. In C. Selitz, L. Wrightsman, & S. Cook (Eds.), *Research methods in social relations*. New York: Holt, Rinehart and Winston.
- Bogdan, R., & Biklen S. (2007). *Qualitative research in education* (5th ed.). Boston: Allyn & Bacon.
- Bogdewic, S. (1999). Participant observation: In B. Crabtree & M. Miller (Eds.), *Doing qualitative research* (2nd ed.). Newbury Park, CA: Sage Publications.
- Denzin, N. (1989). *The research act: A theoretical introduction to sociological methods*. Englewood Cliffs, NJ: Prentice-Hall.
- Gall, M., Borg, W., & Gall, J. (1996). *Educational research* (6th ed.). New York: Longman.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Garfinkel, H., & Sacks, H. (1970). On formal structure of practical action. In J. McKinney & E. Teryakian (Eds.), *Theoretical sociology* (pp. 338–366). New York: Appleton-Century-Crofts.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Hartman, D., & Wood, D. (1990). Observational methods. In A. Bellack, M. Harsen, & A. Kazdin (Eds.), *International handbook of behavior modification and therapy*. New York: Plenum.
- Junker, B. (1960). *Field work: An introduction to the social sciences*. Chicago: University of Chicago Press.
- Lofland, J., Snow, D., Anderson, L. & Lofland, L. (2006) *Analyzing social settings: A guide to qualitative observation and analysis*. Belmont, CA: Wadsworth.
- Mead, G. H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.
- Miles, M., & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage Publications.
- Patton, M. (2002) *Qualitative evaluation and research methods*. Newbury Park, CA: Sage Publications.
- Riis, J. (1980). *How the other half lives*. New York: C. Scribner's Sons.
- Stott, W. (1973). *Documentary expression and thirties America*. New York: Oxford University Press.
- Taylor, S., & Bogdan, R. (1998). *Introduction to qualitative research methods*. New York: Wiley.
- Thomas, J. (1993). *Doing critical ethnography*. Newbury Park, CA: Sage.
- Thomson, J., & Smith, A. (1877). *Street life in London*. London: Sampson Low, Murston, Searle and Rurington.
- Thomas, W. (1923). *The unadjusted girl*. New York: Harper & Row.
- Thomas, W., & Znaniecki, F. (1927). *The Polish peasant in Europe and America*. New York: Alfred A. Knopf.
- Wagner, J. (Ed.). (1979). *Image of information*. Beverly Hills: Sage Publications.
- Walizer, M., & Wienir, P. (1978). *Research methods and analysis*. New York: Harper & Row.
- Weitzman, E., & Miles, M. (1995). *Computer programs for qualitative data analysis*. Thousand Oaks, CA: Sage Publications.
- Wieder, D. (1974). Telling the code. In R. Turner (Ed.), *Ethnomethodology*. Baltimore: Penguin.

Evaluation Research

KEY TERMS

art criticism	evaluation research	practical benefits
bias-free approach	evaluation team	regression-discontinuity design
criterion-referenced tests	external evaluator	reliability
decision makers	external experts	subjective approach
ethnomethodologists	goal-based approach	validity
evaluation contract	health risk appraisal	
evaluation process	outcome data	

Case Study

Rosewood Hospital has had a cardiac rehabilitation program for two years. A new chief of staff is searching for cost-cutting measures and has asked for an evaluation of the cardiac rehabilitation program. The chief of staff sought proposals from agencies, universities, and other groups to evaluate the program. The evaluation contract was eventually awarded to the Department of Community Health at the University of Dover. The department's evaluation team was headed by Sarah, a professor with experience in hospital-based evaluations.

Introduction

Evaluation research is yet another type of research that health scientists utilize for a myriad of reasons. Usually in evaluation research, scientists seek to determine if a program's goals and objectives have been achieved. Evaluation research has been defined as:

1. The application of scientific principles, methods, and theories to identify, describe, conceptualize, measure, predict, change, and control those factors or variables important to the development of effective human service delivery systems (Streuning & Brewer, 1983, p. 211).

2. A social science activity directed at collecting, analyzing, interpreting, and communicating information about the workings and effectiveness of social programs (Rossi, Lesley, & Freeman, 2004, p. 2).
3. An evaluation using an experimental or quasi-experimental design conducted to establish the efficacy or effectiveness—internal and/or external validity—and cost effectiveness or cost benefit of an intervention among a defined population at risk for a specific impact or outcome rate during a defined period of time (Windsor, Baranowski, Clark, Boyd, & Goodman, 2003).

These definitions give us a sense of what evaluation research is about: *a process of determining the effectiveness of a program*. We emphasize *process* as opposed to product. Once the process is completed, decisions are then made about the program being evaluated. For instance, in our case study Sarah may provide information about the cardiac rehabilitation program concerning its length, duration, and cost benefits, which can provide valuable information to the chief of staff. The chief of staff may decide to improve the program's weaknesses, terminate the program, or add more funds to the program.

Purposes of Evaluation Research

As we have stated, there may be many and varied reasons for a research evaluation project to be undertaken. We can look at the purposes for evaluation research from different vantage points: (1) the program evaluator; (2) the administrator; (3) the consumer or public; and (4) the organization or, in our case study, the hospital.

Sarah's team of evaluators from the University of Dover may wish to contribute to the knowledge and the discipline of evaluation research. In addition, the team may view this as a means for advancement in their area of expertise. An altruistic purpose for Sarah's group may be a belief that they can help the discipline of health science by conducting this evaluation.

The chief of staff of Rosewood Hospital may have other purposes for the evaluation, including (1) gaining control over the program, (2) bringing attention to the cardiac rehabilitation program, or (3) bringing the program to the attention of the hospital (Shortell & Richardson, 1978).

The Rosewood Hospital administration may want to be able to justify the money it spends on the program. Additionally, they may want to determine whether the program is indeed worth the expenditures and plan for its expansion. The government that helps fund the hospital may want the evaluation to ensure that its dollars are well spent. Special interest groups, such as cardiac rehabilitation specialists, may have gotten community support for the program and therefore need to prove its worthiness.

Chelimsky (1978) describes the following purposes for evaluation research: management and administrative reasons, assessment of the appropriateness of program changes, identification of ways to improve the delivery of interventions, or accountability to funding agencies. In addition, evaluations may be completed to fulfill planning and

policy purposes, to test innovative ideas, to decide if programs should be curtailed or expanded, or to support one program in lieu of another.

Whatever the purpose that underlies evaluation research, those involved in the process must be sure to accomplish the task by using appropriate methodologies, processes, and analyses. In other words, the research project must be able to be replicated by other evaluation groups.

What Can Be Evaluated?

Almost anything can be evaluated, and this usually is done in some manner or form. Listed here are some examples of things that have been or could be evaluated:

- Programs
- Drugs
- Health
- Databases
- Lifestyles (single, married, etc.)
- Software programs
- Teaching styles
- Workshops
- Staff and personnel
- Management systems
- Organizations
- Needs-assessment strategies

Even evaluation models and schemes can be evaluated, and some have been so scrutinized. The next section will discuss some evaluation models and attempt to critique them.

Steps in Conducting Evaluation

Once Sarah was appointed to conduct an evaluation of the cardiac rehabilitation program, she had to determine a plan of action. We will use this portion of the chapter to describe the step-by-step process of setting up an evaluation of a program. Two major sources were utilized in the writing of this chapter: Rutman's *Evaluation Research Methods* (1984) and Shortell and Richardson's *Health Program Evaluation* (1978). Figure 9.1 depicts the **evaluation process**. From the inception of the project to the final report, there must be interaction between the evaluator(s) and the person or group who will be affected by the evaluation: the decision makers.

Although we have depicted the steps as separate entities, in reality some may be working at the same time. In other words, they are interactive.

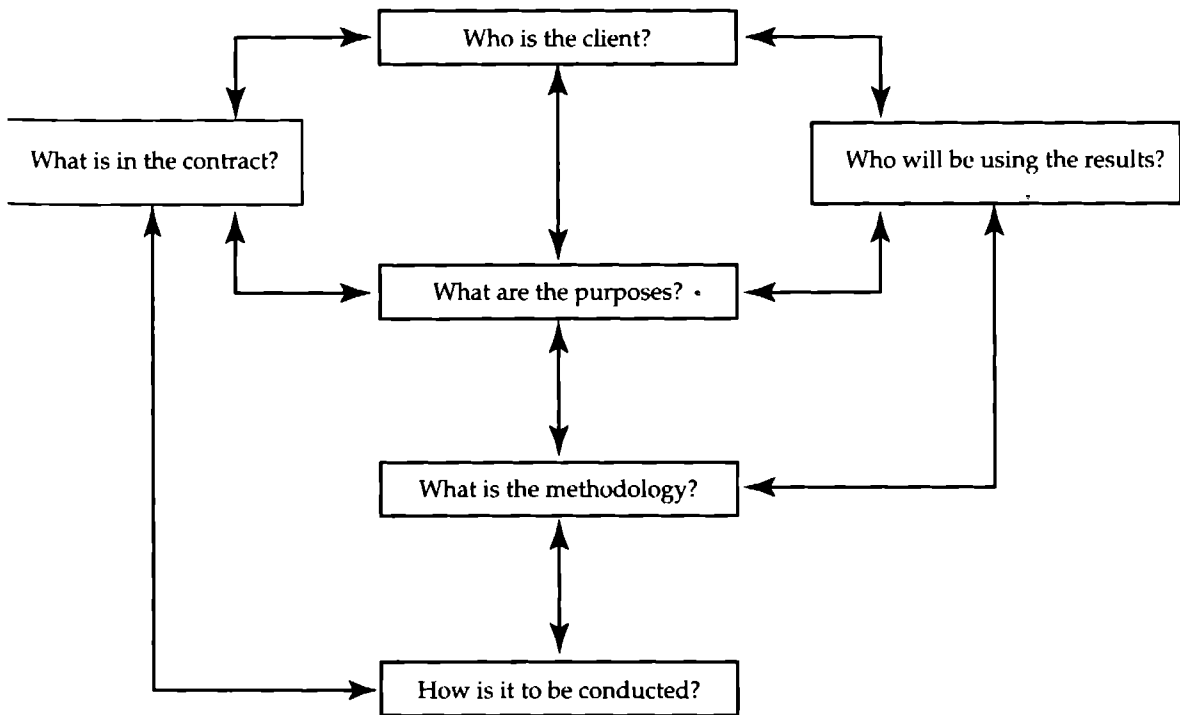


Figure 9.1 The Evaluation Process

Who Is the Client?

In our case study, Sarah has been hired by the chief of staff. Is the chief of staff the primary client, or is the hospital board the primary client? Sarah must determine the power structure of the hospital board and those persons who will be instrumental in making decisions with the added information provided by her evaluation group. She might interview the board members or ask the chief of staff or informed health care workers to ascertain who the decision makers are in Rosewood Hospital. This can become a cumbersome and complicated process, but it is very worthwhile. If it does appear that the entire hospital board is the actual client, then the evaluation group must take this into consideration at every phase of the evaluation process.

What Are the Purposes of the Evaluation?

Sarah should determine the exact purposes for which the evaluation is to be undertaken. If possible, she should ascertain the covert as well as the overt reasons for conducting the study. Earlier in this chapter, we discussed the major purposes of conducting an evaluation and refer you to that section for a brief review. In our case study, let us assume that the members of the hospital board and the chief of staff disagree about their evaluation goals. The board wants to know if the cardiac rehabilitation program is cost-effective and if they should keep it in lieu of other programs. Knowing the main purpose of the

evaluation, the team will have the necessary insight as to what specific methodologies and techniques should be utilized to conduct the study.

What Is the Methodology to Be Utilized?

When determining the methodology to be used, the evaluator must be familiar with the purposes of the evaluation as we have discussed. We will discuss the various methods to use later on in this chapter. However, briefly, they are as follows: experimental and quasi-experimental designs, correlation, surveys, personnel assessment, expert judgment, and case study. Once the methodology is determined, it is then necessary to consider the feasibility of implementing that method. The following characteristics must be considered when determining the feasibility of the method: available funds; time schedule; availability of data; and the legal, political, ethical, and administrative constraints that usually occur when conducting an evaluation.

What Is Included in the Contract?

When Rosewood Hospital sent its request for proposals (RFPs) across the nation, it specified funding levels, time frames, purpose of the study, and other information. Once the hospital contracted with the University of Dover to perform the evaluation, Sarah had to develop a more inclusive contract between Rosewood Hospital and the evaluation team. The **evaluation contract** that they will develop should determine responsibilities of the evaluation team and should include:

1. A line-item budget
2. The entire scope of the evaluation
3. Details concerning the research design and data collection procedures
4. Levels of cooperation of the Rosewood Hospital board
5. Controls to ensure adherence to the evaluation plan
6. Determinations as to the publicity of the final report

How Is the Evaluation to Be Conducted?

As Rutman (1984) has so adequately explained, conducting a program evaluation entails three tasks: (1) measurement, (2) use of a particular research design, and (3) analysis of the data. Each of these areas must receive the necessary attention so that the evaluation process will produce usable results.

Measurement The **evaluation team** decides on the type and amount of information that are necessary to answer the questions posed at the beginning of the study. Evaluations may obtain four categories of information: (1) program, (2) objectives and effects, (3) antecedent conditions, and (4) intervening conditions.

Program information is collected on how the program is run. Information on the process of the program can determine how the program was implemented: if it was

implemented as designed; how the implementation process affected the results of the program; and whether the program was cost-effective.

Objectives and effects are the central point of evaluating programs. The extent to which a program reaches its goals and objectives can be measured in addition to the effects it has produced.

Antecedent conditions refer to the context within which the program operates, the characteristics of the client, and the background of program personnel. Antecedent information helps to interpret the findings, which enables the evaluator to determine what is most beneficial to the client; what type of personnel benefit the program best; and whether the context in which the program operates is best for meeting the program's objectives.

Intervening conditions are unplanned occurrences during the program's activities. For example, the head teacher of the gifted health science program may leave that program, thus causing a disruption to the organization. Measurement of the impact of these intervening conditions can enable the evaluators to determine which factors help the program reach its objectives.

Another consideration is determining the methods by which the information will be collected. Data may be collected through questionnaires, interviews, observations, program documents, and official statistics. These methodologies have been discussed in detail elsewhere in this text; therefore we will not discuss them here. Validity and reliability of instruments become important considerations. **Validity** refers to the extent to which a procedure measures what it is supposed to measure. There are several types of validity:

1. **Face validity:** On the face of it, it is obvious that the instrument measures what it purports to measure.
2. **Content validity:** The instrument will produce a reasonable sample of all possible responses, attitudes, and behaviors.
3. **Construct validity:** The extent to which scores on a proposed instrument permit inferences about underlying traits, attitudes, and behaviors.
4. **Predictive validity:** The instrument can accurately predict a future occurrence.

Reliability refers to the stability of the instrument. In other words, the same results will be consistently reproduced in subsequent administrations of the instrument. To avoid unreliability, evaluators must ensure that instruments are properly worded and administered in a consistent manner.

Research Design Evaluation research treats research design as other types of research do—very importantly. A good research design enables researchers to ask the pertinent questions and determine the answers to those questions. In this regard Cook and Campbell (1994) state the importance of *attribution* and *generalizability* in the choice of a research design. Designs concerned about attribution (the fact that the program has produced the measured results) use several measurements, pretests and posttests, control groups, and random assignment to groups.

To ensure generalizability, the sample must be comprised of people who are representative of the population. In addition, the experiment should be replicable by other investigators.

Data Analysis The final task in conducting a program evaluation is concerned with analyzing the data. For inferences to be made about the program, the correct statistical techniques need to be employed. Statistical methodologies are usually determined by the hypotheses (questions asked) and the research design. Conclusions and recommendations about the program are based on the data analysis; therefore, inappropriate techniques can lead to false or erroneous statements concerning the program.

Who Will Use the Results of the Evaluation?

At the beginning of this section we described the importance of determining who the decision maker(s) will be or who the client is. The **decision makers** will utilize the study's findings for a multiplicity of reasons. Therefore, it is imperative that the decision makers are involved in the evaluation so they will feel a commitment to use the findings of the project. It is anticipated that if the many groups that have a stake in the program (in our case study, members of the hospital board, chief of staff, health care workers, patients, and administrators) receive information about the process and outcome of the program, then more appropriate decisions can be made concerning the program.

The steps in conducting an evaluation have been discussed with the idea that they are interactive with each other. The evaluation team should determine, in advance, a well-conceived plan of action to carry out the evaluation. Communication with pertinent personnel is of utmost importance when these procedures are drawn up. Planning is imperative, no matter the type of evaluation research.

Evaluation Models

Evaluation research dates as far back as 2200 BCE, when the Chinese emperor insisted that all of his public officials pass proficiency tests. Would that be possible today? Testing seems to have stimulated the evaluation movement when Alfred Binet was asked by France's public instruction minister to devise a test to screen for mentally disabled children in the classrooms. This test became the basis for the IQ (intelligence quotient) tests currently in use in the United States. With time, evaluation and measurement took a firm foothold in science. Individual differences were characterized by evaluation studies; and, toward the middle of the twentieth century, evaluation strategies began to concentrate on groups, organizations, and curricula. The following discussion depicts several evaluation models and explains advantages and disadvantages of each model.

The Behavioral Objectives Model

The Behavioral Objectives Model, advanced by Tyler (1949), Mager (1962), Bloom (1984), Suchman (1976), and Popham (1993), starts with a program's goals and collects data to determine if these goals were met. The program's success is measured by the outcomes of the program in relationship to the stated goals. The field of education has had much to do with this model, because Tyler (1949) defined educational goals in terms of student behaviors. These student behaviors were then assessed to determine if

modifications or refinements in the curriculum were needed. The behavioral objectives approach, as proposed by Tyler, included the following steps:

1. Derive a pool of candidates by examining learner studies and by soliciting suggestions from content specialists.
2. Pass the pool of candidates through a series of three screens: philosophical, psychological, and experimental.
3. Put the candidates who survive the screening process into a matrix whose rows depict the content areas and whose columns depict the student behaviors expected in relation to the content areas. The individual matrix cells represent individual objectives.
4. Identify situations in which students can express the behaviors mentioned in the objectives.
5. Develop instruments that can test each objective.
6. Apply the instrument, in pretest and posttest paradigms, so that behavioral changes assigned to the curriculum can be measured.
7. Examine the results to determine the strengths and weaknesses of the curriculum.
8. Develop the hypotheses that account for the pattern of the strengths and weaknesses.
9. Modify the curriculum, and begin the process once more.

Tyler's approach led the way for program and organizational evaluations, in addition to the individual evaluations that were being conducted before Tyler came upon the evaluation research scene. In the health sciences, this model has been used in the area of dental health by the American Dental Association. More recently, his model has taken on additional refinements in the form of **criterion-referenced tests**, in which objectives are first set, and then the tests are based on that set of objectives. In addition, competency testing has become popular in many states. Here, objectives that are the basis for a test serve as the minimal objectives for a particular grade level. In other words, students must pass the competency test to go on to the next level or grade.

Education is not the only field in which the behavioral objective model has been used. Businesses and other organizations have utilized the *management-by-objectives* (MBO) approach, whereby personnel determine their own objectives and then are evaluated on the basis of how well they met those predetermined objectives. In the public health arena, Suchman (1976) wrote a book using a **goal-based approach** in which he identified goal activity and put it into operation. The reader was to assess the effect of the goal operation, form a value about that goal, set the goals and objectives, and finally measure those goals. Suchman said, "The most identifying feature of evaluative research is the presence of some goal or objective whose measure of attainment constitutes the main focus of the research problem" (p. 37).

This model has its *weaknesses* in that questions usually arise concerning who sets the goals and objectives, whose interests they represent, whether the goals really are a complete set of the desired behaviors, how these goals can be measured, and whether important outcomes are reflected by the respecification of objectives (House, 1999).

These questions or inherent problems of this approach must be carefully thought out before one embarks upon a behavioral objectives model toward evaluation.

A *strength* of the behavioral objectives approach is that it provides validity. By applying this model, the objectives should be directly linked to the measured outcomes of the program. A second advantage is that this model enables the evaluator to have a set plan of action by establishing predetermined steps to be taken.

The Systems Analysis Model

In the Systems Analysis Model, a few output measures are determined, and then the differences in programs or their policies are measured as they vary against the predetermined indicators. This model had its roots in the Department of Defense in what was once called the Department of Health, Education, and Welfare and is now the Department of Health and Human Services.

When the U.S. government began utilizing this model, it was to evaluate Title I, a program that provides funds for disadvantaged youth. There were 30,000 projects to be evaluated by using test scores as the measure of success. The results were to be reported in a normal curve and aggregated at the state and national levels. These evaluations led the government to act concerning certain programs. What developed was a systems approach to a cost-benefit analysis that would compare programs.

Cost-benefit analysis is a distinct and difficult methodology to undertake, even for a governmental organization. The goals are predetermined, and then the outcomes of the program are measured. Once the outcomes have been measured, the programs then must be compared on a cost level, which should determine the best outcome for the least amount of money.

Systems analysis can be used for evaluating management and planning procedures, policy structure, and budget procedures. Rossi et al. (2004) ask the following questions when preparing an evaluation:

1. Does the intervention reach the target population?
2. Is the intervention implemented in the manner specified?
3. Is the intervention effective?
4. How much does the intervention cost?
5. What are the intervention's costs relative to its effectiveness?

These types of evaluations must be very objective and leave no room for the common-sense type of qualitative research discussed previously. It is obvious that the social scientist's method of conducting this type of evaluation is relied upon very heavily. These evaluations usually can produce realistic evidence, which should be able to be duplicated by other researchers. A comprehensive evaluation can illuminate facts about program planning, program monitoring, impact assessment, and economic efficiency (House, 1999).

This model has been used with great success by many economists, especially within the governmental sector of our nation. Evaluators utilizing the systems analysis model have a preconceived notion as to the function of the program under scrutiny. The role

and function of that program have been well delineated by the government, thus enabling the evaluative researcher to answer concrete questions relative to that program. This model has recently begun to be used in the health science area. Investigators attempt to determine the cost-effectiveness of a program, such as a hospital's patient-education program. In our case study, Sarah would most certainly choose this method as one of many to evaluate the program.

The major *disadvantage* of this model is that, because it is utilized so often by governmental agencies, it excludes the interest and concerns of the program participants. In an excellent critique of this model, House (1999) claims:

At its worst, the systems analysis approach leads to scientism—the view that the only way to the truth is through certain methodologies. Objectivity is equated with reliability, with producing information from only certain types of instruments. Impartiality and validity are sacrificed. In reducing everything to a few indicators so that one can demonstrate reliability, do cost-benefit analysis, and discover the most efficient programs (thus maximize utilities), the outcomes of complex social programs are narrowed to a few quantitative measures. (pp. 226–227)

At times, the systems analysis model takes an almost hard, cold approach to evaluation, and seems to not consider the human/social context of the program or organization.

There are *advantages* to this model, especially when the evaluator is looking for simple cause-and-effect relationships. In this case, a few indicators may indeed be enough to determine those relationships. For example, the systems analysis approach may be used to determine physician shortages in certain areas of the country. A simple count would suffice, here using one indicator. However, the evaluators must ensure not to extrapolate a possible shortage of physicians to indicate insufficient health care delivery.

The Decision-Making Model

Evaluation research made perhaps one of its largest contributions in the field of decision making. Evaluators believed that they could enable decision makers (e.g., program leaders, administrators) to make more informed decisions based on the data they collected. Stufflebeam (1971) developed the CIPP model to enable evaluation to contribute to the decision-making process in program development. (CIPP is an acronym for the four types of educational evaluation processes in the model: context, input, process, and product.) The model also includes three settings in which decisions are made: homeostasis, incrementalism, and neomobolism. There are additionally four types of decisions that can be made: planning, structuring, implementing, and recycling. Stufflebeam also delineated three steps in the process of evaluation: delineating, obtaining, and providing. We have chosen to discuss the types of evaluation Stufflebeam described because they have wide implications for health science evaluation research.

Context evaluation involves analyzing a program's problems and needs (a discrepancy between a desired condition and an existing one). In our case study, Sarah might find there are 200 cardiac rehabilitation patients in the program, but she has determined that there is a need, by consensus, to reduce that number to 150. Hence a *need* has emerged. Once the needs have been determined, program objectives that will alleviate that need are proposed or determined.

Input evaluation considers the resources and strategies of the program. The data that are collected during this stage of the research project enable the evaluator to make decisions regarding (1) whether equipment or resources might be too expensive; (2) whether a particular strategy could be effective in achieving the program's goals; (3) whether strategies are legally or morally acceptable; and (4) how to best utilize staff as resources.

Process evaluation begins once the program is underway. For example, Sarah will begin to observe and collect data relative to the cardiac rehabilitation program. If she were to find that program patients did not attend the sessions and had higher absentee rates than other patients, the program decision makers may take some action based upon this information. For example, they might encourage patients to attend cardiac rehabilitation programs in other hospitals. In addition, process evaluation methods serve to keep records of the events that occur during the program. These records should be kept over a relatively long time period so that peaks and valleys can be ascertained.

Product evaluation determines the extent to which the program's goals and objectives have been achieved. The evaluators develop measures of the goals and then administer them to the proper audiences. The results enable program administrators to make decisions concerning modifying or ending the program. There is another approach to the decision-making model, as suggested by Patton (2002). He suggests that for evaluations to be successful and relevant, the decision makers must be identified so that the gathered information gets to the people who can implement change. A second step in conducting this type of evaluation is to identify and focus the relevant questions. The program administrators also may specify how they would use the answers to questions and thus provide much needed input into the research study.

Decision-making approaches primarily use survey methods such as questionnaires, face-to-face interviews, and small group discussions. The evaluator goes to the setting (for example, Sarah would go to Rosewood Hospital) but does not necessarily set up an experiment. The program administrator, or decision maker, plays a very important role in this type of evaluation model.

The major *advantage* of this model is that the methodology selects the information that would be most useful to the evaluator. Criterion measures are predetermined so that the evaluation is actually focused for the evaluation team. In our case study, Sarah would meet with the administrator of the cardiac rehabilitation program to help determine the questions to be asked during the evaluation process. At this point, some might believe that the decision maker (the administrator) may have too much to say concerning the evaluation and thus not be objective in the approach.

This leads to some *disadvantages* of the decision-making model: the administrator is preferentially treated, the evaluator becomes very close to the managers of the program, and the evaluation can become undemocratic and unfair. In addition, the identification and specification of important decisions are difficult tasks for the evaluator. Who should be included in this process? How should these people be utilized before, during, and after the evaluation?

The Goal-Free Model

The Goal-Free Model was developed by Scriven (1973) when he was employed as one of a group of advisors to the Educational Testing Service to screen candidates for a list of

innovations and developments that had been funded by the federal government and whose evaluations proved that they were worth giving to schools. During this experience he delineated intended and unintended effects, or side effects, of a particular program. If one knew the goals of a program, this might serve as a contaminating effect. Therefore, Scriven decided that evaluations should be goal free and that the organizational framework of the process becomes the effects of the program, rather than the goals.

Goal-free evaluation is not widely utilized because of two factors: (1) Scriven has not given much guidance as to the procedures of the model; and (2) evaluators cannot easily find criteria to judge the program if they are not to use the administrators' or the developers' goals and objectives. It is apparent that social science areas may find it difficult to utilize this model, but one good example of its use is offered by Consumers' Union. This group typically evaluates all sorts of products using what the members believe are criteria that consumers would prefer. Goal-free evaluation uses the reference group of the consumers of programs rather than the producers or administrators.

Scriven has developed the concept of consumers' need to be analyzed and does so through needs assessments of a particular group of people. He has insisted upon a **bias-free approach** when conducting the research so the evaluator can remain as objective as possible. The techniques that are utilized include double-blind experiments, in which neither the subject nor the evaluator knows which is the treatment and which is not. Scriven describes goal-free evaluation as a triple-blind experiment in which neither of the treatments is known.

The goal-free model, though not utilized very frequently, does serve to point evaluators to the important aspect of the unintended or side effects a program may offer. Evaluators have become sensitized to the not-so-evident effects and therefore have added a richness to the field of evaluation research.

Scriven's goal-free model is implicit in reducing bias in the evaluation. Because the evaluator has no contact with program personnel and does not have an inkling as to the goals of the program, a supposedly unbiased evaluation can occur. Unfortunately, there have not been many goal-free evaluations, assumably because of the unwillingness of administrators to allow strangers to act as investigators, or "hunters." The major *disadvantages* of the goal-free model include (1) the lack of a clear methodology as to how to approach this model; and (2) the lack of social interaction between normally social people: evaluators and administrators of human service programs.

The *advantages* of the goal-free model lie in its ability to be completely bias-free from stated or predetermined goals. Consumers' Union does evaluate products in this manner. However, we must remember that products are not processes, and it is process that concerns most evaluations in the health sciences. It appears that goal-free evaluations, in combination with other models, make a contribution to program or organizational evaluations.

The Connoisseurship Model

Connoisseurship is the art of perception that makes the appreciation of complex educational practices possible (Eisner, 2002). In this model, a judge is used to determine the effects of a program, curriculum, or educational system. Just as an art critic might

write about a painting, so will the connoisseur write of program evaluations. Connoisseurship consists of recognizing and appreciating the qualities of a program, and the connoisseur writes about that program, or, in other words, renders a critique or criticism of that particular program. Sarah, in our case study, would be the connoisseur.

The connoisseur, or critic, must be an experienced evaluator who will point out the significant aspects of a program through writing about his or her feelings about the program. With such an approach there are many methodological considerations to take into account: How can one know whether a critic is to be trusted? How can one be sure that the critic is not imagining the events? How can one know what confidence to place in the critic's description, interpretation, and evaluation of a program?

Eisner (2002) does suggest that, to obtain validity for this model, one should look for instances of structural corroboration, where separate pieces of evidence validate each other if they fit together to form a coherent, persuasive whole. Another way to validate the critic's findings is to ensure that the language used is referentially adequate. This can be accomplished by having other critics view the program through analysis of videotapes and audiotapes to assess if the first critic's interpretation of the material was valid.

The critic must be experienced and trained in utilizing the connoisseurship model. The evaluator becomes totally involved in the program and becomes as familiar with it as is possible. The criticism that comes about through this model will usually lead the program to improve its standards and probably perform better. Although this model is not in the traditional mode of collecting and analyzing data, it does serve to supplement other approaches and to provide additional insight to what is being evaluated.

This model, because it is very closely aligned with **art criticism**, provides a very impressive and expressive idea for completing an evaluation. Because this type of method has worked very well in the arts (drama, music, painting, etc.), it should prove valuable to the science of evaluation. Connoisseurship has the *advantage* of adding another way to accomplish and augment an evaluation. It provides an expert review that, although biased, can detect weaknesses, flaws, and strengths of programs and/or organizations.

However, as with all of our evaluation models, there are *disadvantages* in utilizing the model. One is that the connoisseur, who is really an expert, must decide on criteria in making judgements about programs. But how does the evaluator decide on these criteria, and how are these criteria justified to the administrator of the program? A second disadvantage is that Eisner (2002), the creator of the connoisseurship model, maintains that the evidence for the criticisms is the behavior of the teachers of programs being evaluated. If this is so, can or will teachers readily accept the criticisms? Have they?

The Case Study Model

In this model, the *process* of a program or an organization is the focus of the study. The evaluator attempts to depict the program to those involved in the program by presenting a case study. This is usually accomplished through interviews that serve the same purpose as those that **ethnomethodologists** conduct. You will often hear the term

naturalistic applied to these methodologies. The major proponent of the case study model has been Stake (1991), who stated that

case studies will often be the preferred method of research because they may be epistemologically in harmony with the reader's experience and thus to that person a natural basis for generalization. . . . If the readers of our reports are the persons who populate our houses, schools, governments, and industries, and if we are to help them understand social problems and social programs, we must perceive and communicate in a way that accommodates their present understandings. Those people have arrived at their understandings mostly through direct and vicarious experience. (p. 5)

It is evident that the case study model enables people to gain an understanding of their world through complex descriptions of that world. The methodologies utilized to collect the data are interviews, as we stated previously. However, the writing becomes very informal, with illustrations and allusions. This evaluation model has led Stake (1991) to develop and implement responsive evaluation, which will be discussed later in this chapter.

The case study model comes from the qualitative research methodologies in which participant observation, interviews, document study, field notes, and subjects' own words are used to collect and record the data. Although the case study model is similar to the connoisseurship model, there is a difference: The case study evaluator determines the perceptions of other people, whereas the connoisseur relies on only his or her own experience and values. Sarah would utilize these methodologies and act as the sole case study evaluator.

The case study model is a **subjective approach** to evaluation and has met with criticism as compared to the more scientific methods of evaluation. Because of the nature of the methodology of the case study, different observers will emphasize different parts of a program. This leads to the *disadvantage* of inconsistency. Another disadvantage is the interpretation made by the evaluator. The evaluator must observe all parts of a program and then draw conclusions. Whose values are the basis of these conclusions?

There are many *advantages* to the case study model. As House (1999) has stated, these include (1) obtaining rich and persuasive information from program participants and other people removed from the program; (2) representing diverse points of view and different interests; and (3) potentially being persuasive, accurate, and coherent.

Accurately depicting a case study and trying to avoid becoming personally involved can be a difficult task for any evaluator. Although this model has many advantages, it suffers from the same weakness that other models do: not having written methods and procedures.

The Accreditation Model

Many professional associations conduct evaluations of their respective professional training programs. In the area of health sciences, one of these accrediting bodies is the Council for Education in Public Health (CEPH), which accredits schools of public health and community health programs outside schools of public health. Other examples of professional reviews involve lawyers, physicians, educators, social workers, and speech and hearing therapists.

The CEPH accrediting process is similar to that of many professional organizations, so we have chosen to use it as an example. Typically, the accrediting body sets up criteria

Table 9.1 Evaluation Models

Model	Audience or Reference Group	Methodology
Behavioral objectives	Managers, social scientists	Behavioral objectives, achievement tests
Systems analysis	Economists, managers	Cost-benefit analysis, planned variation
Decision-making	Administrators	Surveys, interviews
Goal-free	Consumers	Bias control
Connoisseurship	Consumers, connoisseurs	Critical review
Case study	Client, practitioners	Case studies, interviews, observations
Accreditation	Professionals, public	Self-study review by panel

Source: From *Evaluating with Validity* by E. House, 1999, Beverly Hills: Sage. Copyright 1980 by Sage. Adapted by permission.

that are organized into several sections concerned with the professional training program. The approach that CEPH utilizes involves a self-study evaluation of the program. The program personnel have several months to prepare an in-depth analysis of the program and submit it to the CEPH staff.

A subcommittee of CEPH is appointed as the visiting on-site group, and they review the self-study report previous to the on-campus visit. Once on the site, the committee interviews staff, faculty, and students associated with the program being evaluated. The committee prepares a brief report and informally gives its findings to the program director. Typically, a few months later, a formal report is forwarded to the director with comments regarding the strengths and weaknesses of the program. The program director is then given an opportunity to correct any perceived weaknesses or devise a plan to do so. Program representatives are then invited to a meeting of the full council for a final disposition. The results of this process are then communicated to the appropriate program personnel.

Accreditations are useful in that they provide a mechanism for self-evaluation as well as simultaneous peer evaluation. Programs that have been accredited are believed to have met criteria and standards set by the profession. The accreditation model has grown as professional groups come under increased pressure to evaluate their own programs. The *advantages* of this model include (1) the self-evaluation mechanism; (2) setting of criteria and standards for a profession; (3) a continual effort at evaluation; and (4) accountability to the public. Although these advantages place a great burden on the profession in question, it seems apparent that a review of methods and procedures of professions can only help to improve that profession.

Outcome	Typical Questions
Productivity, accountability	Is the program achieving the objectives? Is the program producing results?
Efficiency	Are the expected effects being achieved? Can the effects be achieved more economically? What are the <i>most</i> efficient programs?
Effectiveness, quality control	Is the program effective? What parts of the program are effective?
Consumer choice, social utility	What are <i>all</i> the effects?
Improved standards, heightened awareness	Would an unbiased critic approve this program? Is the audience's appreciation increased?
Understanding, diversity	What does the program look like to different people?
Professional acceptance	How would professionals rate this program?

The accreditation model also has its *disadvantages* in that the public has challenged the model and the process, which can bring about disharmony within the profession itself. A second limitation is the procedure that professional organizations utilize in their evaluations. The accrediting teams may not always have fair and competent people, and each team may vary when conducting a site visit. This can lead to uneven evaluations among member programs and create a system of review that does not promote equality.

Table 9.1 summarizes the evaluation models we have discussed. These models can be utilized alone or in any combination. We have attempted to give a clear description of each model with a sense of its advantages and disadvantages so that evaluators may be able to decide which model is best suited for each evaluation situation. In our case study, Sarah might appropriately choose a combination of the decision-making and case study models. Can you suggest another model that Sarah could utilize to evaluate the cardiac rehabilitation program?

Types of Evaluation Research

There are several types of evaluation research, some of which have developed from or have been precursors to the models previously discussed. We have chosen to discuss four types of evaluation research: (1) needs assessment; (2) formative; (3) summative; and (4) responsive. These types of evaluation research can be utilized in the health sciences because they have broad application in the social sciences.

Needs Assessment

Usually, a needs assessment occurs when someone at the decision-making level feels there is a discrepancy between an acceptable condition and the existent condition. In our case study, the needs assessment would have been conducted *previous* to the implementation of the cardiac rehabilitation program. Often, needs assessments are conducted at the community level: Hospitals begin wellness centers, voluntary associations set up screening programs, and public health departments determine the *need* for free services.

Public support of a program should be demonstrated before that program becomes an actuality. Recently we have seen strong support for child and partner abuse education programs, both in hospitals and other community groups. This has resulted because of the exposure of several child abuse cases involving adults who were in charge of preschools and parents who were abusing their own children, as well as attention given to partner abuse through the media. Even though there appears to be widespread support for child and partner abuse education programs, it is best to carry out a needs assessment to determine if a *specific group* would support such a program.

Needs assessment has been called front-end analysis (Harless, 1973), which means that the assessment of the needs of a program, evaluation of the program's conception, a cost estimate, determination of feasibility, and projections of demand and support are researched before the program commences. Two factors that might determine if a program is needed are frequency and intensity. If many people have a need for a program (frequency), then public support can be elicited, as in the case of the child abuse education program. In addition, if the need is seen as intense or grave, the program will receive support as well. There are some questions that researchers should consider when they are about to undertake a needs assessment (Anderson & Ball, 1978, p. 20):

1. What made us think that there were needs requiring investigation?
2. Whose needs are we talking about?
3. How can we find out whether the needs are frequent or intense enough to justify intervention?
4. How much frequency or intensity is sufficient, and how much discrepancy between acceptable state and observed state is undesirable?

Once researchers have determined to undertake a needs assessment, it becomes a fairly straightforward matter. The need has been determined; there is a discrepancy between an acceptable condition and one that now exists. Let us assume that our case study began at the needs assessment stage. Sarah was asked to evaluate a cardiac rehabilitation program. The major design would be to survey the important figures (as previously mentioned) and perhaps interview a smaller, select group. The survey would be designed so that as many issues relevant to the program would be discerned to elicit an analysis, which would compare the status of the current program with the desired program.

The final report that Sarah submits to Rosewood Hospital's board of directors and chief of staff should enable the decision makers to be better informed about the cardiac rehabilitation program. Many specific objectives and goals of the program can be determined by the outcome of the analysis of the results.

Formative Evaluation

Scriven (1967) distinguished between two forms of evaluation: formative and summative. Formative evaluation occurs when data are being collected. This process provides information so that revisions and improvements can be made for the program. Formative evaluation can occur when instructional programs or curricula are to be evaluated. The cardiac rehabilitation program would have been a good target for formative evaluation. Industry also utilizes this type of evaluation when evaluating products during particular stages of development.

Baker (1974) describes four stages of concern when an evaluation of a new program (formative evaluation) is undertaken:

1. Determine the results of the program.
2. Diagnose the weak areas.
3. Limit the number of subjects exposed to the new, unproven program.
4. Limit the costs of the program.

Keeping these guiding principles in mind, the evaluator should then decide on what kinds of data are to be collected. **Outcome data** is the first category of data collection. Here, the researcher is concerned with how the program affects those people directly involved in the program. A second type of data that is to be collected is that of program implementation. How does the program operate? Is the sequence adequate? What about the presentation of the format? At this stage, weaknesses of the program may be enlightened and thus serve as targets for revision. For the revisions to have a high probability of success, it is suggested that the data collected on program effects should be correlated with those concerning its implementation. The data collected are usually in the form of observations, questionnaires, and interviews.

Formative evaluations are usually conducted by an in-house or internal evaluator. Formative evaluators are concerned with several important questions when conducting the review (Baker, 1974). The first to consider is: *Are the instructional materials accurate?* This leads evaluators to examine the curriculum materials, instructional guides, plans, and activities to ensure that the concepts are properly presented and are accurate. For example, in the cardiac rehabilitation program, all the materials and subsequent examples would be scrutinized.

A second major concern is: *Does the content reflect an appropriate range?* The researchers would look for a broad range of materials rather than a set of materials that represents only one point of view. For example, in our case study, it would be inappropriate for the materials to contain an overabundance of information about hydrotherapy if that took away from a well-rounded program.

A third consideration is: *Is the product well designed instructionally?* The evaluation team has to determine if the prescribed goals and objectives of the program are being met. In addition, if the program planners had designed the program to utilize a certain learning theory, were the principles of that theory ascribed to in the delivery process of the program?

A fourth question to be asked is: *Does instruction account for all planned outcomes?* In this situation, the evaluator is faced with determining whether the outcomes,

or goals, as stated by the program developers have been accomplished. In health science, we must ensure that all domains—cognitive, affective, behavioral, skills, and decision-making—are taken into account when developing and evaluating a program.

A fifth consideration is: *What is the instruction's level of quality?* Here, the program activities are judged to see if they are effective. By assessing the participants' levels of activity and consequent outcomes, an evaluator can determine if learning opportunities meet their objectives.

A final question in the internal review of a program is: *Do things fit together?* The evaluator must ask when reading through a program: *Do references match, are page numbers accurate, are the objectives clearly met, are there typographical errors, and can the program be easily utilized by the intended user?* Once the internal review has been completed and is communicated to the program planners, it is up to them to institute the changes or make revisions.

Preliminary Field Test After the internal review process has been completed, the program is tried out on a small sample of potential program participants. Two areas of concern are of importance here: the *effects* of the program on the participants and the *implementation process* of the program. At this phase of the evaluation, questionnaires, observations, and tests are used to collect the data.

Participant tests are utilized to determine the outcomes of the program. These can be used in a pretest-posttest manner or in any other experimental design that is appropriate. The use of standardized tests is not recommended here because those tests would not necessarily reflect the anticipated outcomes of a particular program. Tests also can be in the form of observing behavior and attitudes or observing skills in decision making.

To obtain data concerning the implementation process of the program, the evaluator can collect items completed by the participants during the program. These materials can be in the form of assignments, memos, letters, fieldwork situations, and so on. The data can be related to the participant's postprogram scores to determine any weaknesses in the program or any areas that need revision. Observational data become very useful at this stage because the evaluator can pinpoint when certain activities work, are helpful, provide enrichment, and the like. The data collected during the preliminary field test enable the evaluator to gather information regarding the program participants' reactions and to detect unplanned outcomes of the program (Baker, 1974). The evaluators then prepare a report for the program planners so that they can debug the program and get it ready for an operational field test. In this context, the program is then placed in the situation for which it was intended. At this stage, the evaluation is not as concerned with meeting program goals and objectives, but rather attention is directed toward issues of program utility, integration, and access. In our case study, Sarah must ask if the cardiac rehabilitation administrators and health care workers would become actively involved in the evaluation process.

Summative Evaluation

Once the program has been evaluated in the formative manner, summative evaluation can take place. The purpose of this type of evaluation is to assess the overall effectiveness of a program and the extent to which the program is worthwhile in comparison to

other similar programs. The results of the evaluation assist decision makers in terms of whether to adopt a particular program, utilize a product, or implement a procedure. An **external evaluator** usually is employed to conduct the summative evaluation. This person should be as impartial as possible and have no connection to the program whatsoever. In our case study, Sarah is conducting a summative evaluation in that she is concerned with evaluation of the cardiac rehabilitation program after it has been in effect for 2 years. Once Sarah completes the evaluation process, administrators and the Rosewood Hospital board will be able to determine if the program is worthwhile and if it is cost-effective for the hospital. Thus, summative evaluation can provide the consumer with an independent assessment of a program.

Summative evaluation is concerned with the impact of a program, and, because of this, the researcher has to ask: *impact with respect to what?* Many evaluators set about conducting a summative evaluation with a behavioral objectives approach, and yet others advocate employing a goal-free approach. Once the criteria against which the program's effectiveness will be measured are chosen, the evaluator then chooses measures that reflect the chosen criteria. The researcher also is concerned with the target population, sampling, and the study design. An evaluation report usually is rendered to those who called for the evaluation. This report should indicate the successes and weaknesses of the program and should be able to point to those program features that influence successes or failures (Anderson, Ball, Murphy, & Associates, 1975).

Planning Summative Evaluations There have been many summative evaluations in the health sciences, especially in the area of program evaluation. There also have been summative evaluations in the medical field in which a product or procedure has been evaluated for its effectiveness. No matter the context of the evaluation, planning is a very important and vital aspect in summative evaluation. Because external evaluators are utilized, the process can be costly and very time consuming.

Purposes of Summative Evaluation Anderson and Ball (1978) have delineated the various purposes of summative evaluation, which include *monitoring the continuing needs for the program*. At times, it can be shown that the program has met its original intent and should be disbanded. This is true in some health maintenance programs, for example, in which hypertension is under control, smoking has ceased, or weight loss has been achieved. This purpose can be achieved through collecting survey data, conducting personal assessments of the participants, and gathering expert judgments about the program.

A second purpose for summative evaluation is the *assessment of the cost-effectiveness of the program*. Sometimes it appears that it may be impossible to evaluate human effectiveness in terms of costs; for example, if a medical procedure saves one life at a high cost, is this cost-effective? An example is the use of heart transplants in very ill people. If the evaluator were to ask the recipients' family about the cost-effectiveness and then survey the general public, disagreement about the outcome would certainly become apparent. In terms of our case study, the evaluators of the cardiac rehabilitation program would address one of the following decisions:

1. Determine the least expensive means of achieving a specified level of performance.
2. Determine the greatest level of performance that can be obtained for a specified cost.

In cost-effectiveness analysis, it is easier to determine the cost-effectiveness of a program when it is evaluated against another similar program than it is to evaluate a program as an absolute. Sarah will have to determine which approach to take when determining the cost-effectiveness of the cardiac rehabilitation program. Methods to utilize in determining these factors include surveys and correlational status studies.

A third purpose of summative evaluation is to determine the *global effectiveness of the program in meeting the goals and objectives of that program*. Most programs have predetermined goals, objectives, and, as discussed previously, criteria and measures by which to determine the success of the programs. Factors such as long- and short-term effects also should be considered. If a program was instituted as a precursor to a more advanced program (e.g., Phase I before Phase II), the effects of the first program should be evaluated at the end of the first program. On the other hand, if one of the aims of the Phase I program was to teach patients to utilize cardiac rehabilitation programs, this cannot be evaluated until the patients have a chance to choose their type of health care delivery. Long-term effects are much more difficult to evaluate because of the impracticality of following subjects over a long period of time. Methodologies that are utilized for this purpose include experimental and quasi-experimental studies.

The fourth purpose of summative evaluation involves *determining the possible side effects of a program*. Although program objectives may have been clearly delineated at the outset, other positive and negative effects may occur because of that program. For example, the cardiac rehabilitation program may have had a positive effect in keeping patients from high rates of absenteeism from the program because of their interest in the program. On the other hand, a negative, and unintended, side effect may be that some patients enrolled in the program resented it because of the time commitment they were supposed to give. Side effects can be determined by using the case study, experimental, or quasi-experimental approaches.

Responsive Evaluation

Responsive evaluation research focuses on the issues and concerns of the persons who have a stake in the evaluation, or the *stakeholders*. The evaluator *responds* to what different audiences wish to know. Stake (1998) was the first to use the term *responsive evaluation* and says the following about the method:

An educational evaluation is responsive evaluation if it orients more directly to program activities than to program intents; responds to audience requirements for information; and if the different value perspectives present are referred to in reporting the success and failure of the program. To do a responsive evaluation, the evaluator conceives of a plan of observations and negotiations. He arranges for various persons to observe the program, and with their help prepares brief narratives, portrayals, product displays, graphs, etc. He finds out what is of value to his audiences and gathers expressions of worth from various individuals whose points of view differ. Of course, he checks the quality of his records, he gets program personnel to react to the accuracy of his portrayals; and audience members to the relevance of his findings.

He does most of this informally—iterating and keeping a record of action and reaction. He chooses media accessible to his audiences to increase the likelihood and fidelity of communication. He might prepare a final written report, he might not—depending on what he and his clients have agreed on. (p. 16)

What organizes an evaluation utilizing the responsive approach? As Stake has alluded, it is the issues and concerns of those persons with whom the evaluator has had conversations that will receive the attention. He has referred to other evaluation approaches as *preordinate*. The following are suggested steps that Stake (1991) enumerated. They are to be used in progression but also may interchange as the evaluation progresses:

1. The evaluator talks with program staff, clients, and anyone else involved in the program to gain a sense of their posture with respect to the purposes of the evaluation.
2. The evaluator then places limits on the scope of the program. In addition, documents and official records will have been reviewed to help set these limits.
3. The evaluator personally observes the program in action.
4. As a result of Steps 1 through 3, the evaluator begins to discover, on the one hand, the stated and real purposes of the program, and on the other hand, the concerns that various audiences may have with it and/or the evaluation.
5. The evaluator begins to conceptualize the issues and problems that the evaluation should address.
6. Once issues and problems have been identified, the design takes some form. (This occurs very late, as is not true of other evaluation approaches, and is called an *emergent design*.)
7. The evaluator selects methods for gathering data. Stake considers the instrument as humans, in that they will be observers.
8. The data collection procedures are accomplished.
9. Once the data have been collected and processed, the evaluator goes to an information reporting mode. The information is organized into themes, and the evaluator prepares portrayals designed to communicate naturally and provide as much direct personal experience as possible. Portrayals, in this sense, include case studies, plays, artifacts, and videotapes.
10. The evaluator matches issues and concerns to audiences in deciding what form the report will take.
11. The format of the report must be decided when reporting to each audience. The reports may be in the form of written statements, discussions, newspaper articles, and films.
12. The final step is completed when the evaluator assembles any formal reports.

It is obvious that this type of evaluation is continuous and interactive, in that at any point the evaluator could begin again because of the results found so far in the process. There may be new issues and concerns that have been discovered through

Table 9.2 Comparison of Preordinate and Responsive Evaluation Models

Comparison Item	Type of Evaluation	
	Preordinate	Responsive
Orientation	Formal	Informal
Value perspective	Singular; consensual	Pluralistic; possibility of conflict
Basis for evaluation design (organizer)	Program intents, objectives, goals, hypotheses; evaluator preconceptions such as performance, mastery, ability, aptitude, measurable outcomes; the instrumental values of education	Audience concerns and issues; program activities; reactions, motivations, or problems of persons in and around the evaluation
Design completed when?	At beginning of evaluation	Never—continuously evolving
Evaluator role	Stimulator of subjects with a view to testing critical performance	Stimulated by subjects and activities
Methods	Objective; “taking readings” (for example, testing)	Subjective; negotiations and interactions (e.g., observations and interviews)
Communication	Formal; reports; typically one stage	Informal; portrayals; often two stage
Feedback	At discrete intervals; often only once at end	Informal; continuously evolving as needed by audiences
Form of feedback	Written report, identifying variables and depicting the relationships among them; symbolic interpretation	Narrative-type depiction, often oral (if that is what the audience prefers), modeling what the program is like, providing vicarious experience, “holistic communication”
Paradigm	Experimental psychology	Anthropology, journalism, poetry

Source: From *Fourth-Generation Evaluation* by E. Guba & Y. Lincoln, 1999, Newbury Park, CA: Sage. Copyright © 1989 by Sage. Adapted by permission.

the evaluation process. Table 9.2 summarizes the preordinate evaluation approaches in comparison to those of responsive evaluation.

Methodological Approaches in Evaluation Research

Although we have discussed the various methodologies used in other types of research, it is important that methodological approaches be explained with specific respect to those used in evaluation research. The methods that will be discussed include: experimental and quasi-experimental designs, correlation, surveys, personnel assessment, expert judgment, and case study.

Experimental and Quasi-Experimental Designs

True experimental designs will enable evaluators to assess absolute answers to questions they have conceived. If, for example, a curriculum is being evaluated for teaching students the cognitive skills in anatomy, then a true experimental design should be utilized. Randomization, an issue discussed at great length elsewhere in this text, also should be employed. The evaluator should decide to use the classroom, client, program, or school as the unit of randomization because this provides a much stronger design than a non-randomized study.

Quasi-experiments are usually used when randomization is not feasible. As an alternative to randomizing subjects, matching them or using a statistical technique (analysis of covariance) may suffice. Evaluators may want to generalize their results to other, similar groups, such as the success (or failure) of a preschool program. Therefore, in this instance, generalizability is very important and should be carefully considered when setting up the study design. Quasi-experimental designs include time series and the pretest-posttest non-equivalent control group design. The **regression-discontinuity design** (Anderson et al., 1975) is frequently utilized in evaluation research. Here, comparison groups that differ from the treatment group in one significant and continuous dimension are chosen. An example of this would be a control group that differed significantly in family income.

Correlation

Correlational methods are utilized frequently by evaluators and, sometimes without justification, for the reported results. Correlations do not imply cause and effect, and researchers must take precautions not to make such inferences. Regression analyses, however, have proved to be helpful in determining the formative evaluation aspects of a study in which the positive and negative aspects of the program can be compared. Correlations can be utilized in the typical experimental control group design, in which measures are compared to the pretest scores. In addition, correlations between costs and program effectiveness indices across several programs, or even parts of a program, could be utilized to determine the continuation or modification of a program (Anderson & Ball, 1978).

Surveys

Surveys, which are a major type of evaluation research, are the primary instruments utilized to collect the necessary data in a needs assessment. Surveys can include interviewing (telephone, individual, group), observations, questionnaires, and analysis of program records. There are some general facets to interviewing, questionnaire design, and observations that should be considered by evaluation researchers:

1. Provide the appropriate reading level in the questionnaire.
2. Avoid sensitive or ambiguous questions.
3. Match the interviewer with the interviewee.
4. Ensure that the questions are relevant, and remain on task as to the purpose of the survey.

These are discussed in greater detail in Chapter 6, which addresses survey research.

Personnel Assessment

Personnel assessment is especially useful for determining the context of the program being evaluated. The administrative structure and procedures of the program can be ascertained by surveying (as above) the personnel involved in the practice of administration. Data about staff roles, relationships, responsibilities, in-service training, hiring and firing procedures, policies for internal evaluation for remuneration, and the like can be gathered using personnel assessment.

The actual collection of this type of data varies from program to program and within the context of that program. There are many cognitive, affective, behavioral, psychological, and physiological instruments available for use. However, even though these tests may be standardized, we recommend them with caution and ascribe to the tenet that most instruments for program evaluation should be tailored for each program.

Expert Judgment

Expert judgment can come in three or more forms, including: (1) the evaluators as experts; (2) the program staff as experts; and (3) outside panels of experts. The evaluation researcher and the evaluation team can and should be considered experts. They will have to make important decisions concerning design, data collection, interpretation, and the like. With their experience and education, they should be, and usually are, accountable and responsible for the decisions they render.

The program staff that is being evaluated can serve as experts as well. They can give firsthand observations, thoughts, and beliefs about the program. Although they at times may appear to be too close to the program to be objective, recurrent themes can be deduced from the many data sources. These themes will provide invaluable information to the evaluation team.

External experts can be very helpful to the evaluation process. They can be especially useful when the goal is to estimate cost-effectiveness of programs. In these instances, economists or other social scientists familiar with cost-effectiveness determinations can be called upon for their expert advice and judgment. In addition, if documented support for the program is necessary, outside experts, political and professional, may be called upon to provide documentation of the merits and the worth of the program.

Case Study

The case study approach has been discussed in detail elsewhere. In evaluation research, as in other types of research, the case study is very valuable in determining the effectiveness, merit, and worth of programs or organizations. Usually, the evaluator or the team will determine how the case study should be conducted, when it should occur, and which elements are necessary for inclusion in this methodological approach.

A case study can help determine if a program should be initiated, continued, or expanded. It can also help diagnose weaknesses and strengths of a program so that modifications can be instituted. And finally, this methodology can enable the evaluator to establish the process of how and why the program operates.

The methodological approaches discussed in the preceding chapter were illuminated in light of evaluation research. While these approaches are useful and important in other types of research, their uniqueness is apparent to evaluation research. It should be noted here that several methods may be used in any one evaluation project, as they each contribute to the total process.

Methodological Approaches

Evaluation researchers use several methodologies in their studies. These include simulation modeling, cost analysis, and contextual evaluation. While these methodologies may have been utilized in other types of research, evaluation researchers have seen the advantages of these methods.

Simulation Modeling A Simulation Model enables one to understand how and why an intervention works or should work. Several characteristics make simulation modeling a valuable tool for evaluation research (Cooper & Huss, 1981). The first is that simulation modeling requires the investigator to make explicit assumptions regarding the system to be studied. Simulation modeling assists this effort by simulating alternatives or variations to raise assumptions. The second characteristic is that the model can create conditions that would otherwise be difficult, hazardous, or unlikely to occur. These types of models are frequently used in nuclear reactor coding systems. A third use of simulation modeling is that events not under the control of social scientists can be studied. The barriers may be costs or ethics that are beyond the control of the investigator. And finally, the Simulation Model can be utilized to forecast future conditions, guide future allocations of funding, select optional populations for the program, or evaluate additional program alternatives before implementation. Simulation modeling usually begins from theory, builds up the causal links, and then produces results that can be compared with other data to test the reasonableness of the modeled theory. An example of Simulation Models that have been utilized in the health sciences is the health risk appraisal. The **health risk appraisal** is a tool that can describe an individual's chance of becoming ill or dying from some cause over a period of time. Most health risk appraisals are statements of probability of a disease, rather than detection or diagnosis. These kinds of appraisals have been in existence since the eighteenth century, when health professionals began associating specific illnesses with certain occupations. These appraisals did not have a scientific base, as they were made by observing patients. Since that time, the creators of the health risk appraisals have improved the scientific base of the tools in demonstrating the relationship between certain risk factors and specific causes of death or disability. A health risk appraisal is:

1. A method or tool that describes an individual's chance of becoming ill or dying from a select cause over a specific period of time, as compared to either (a) the population as a whole; or (b) some similar subset of the population, such as those people of the same age, race, and sex
2. A technique that is in a relatively early stage of development and most useful for middle-class white people

3. Most often intended to:
 - a. Raise an individual's and/or group's level of awareness and knowledge of personal risk factors and potential health outcomes.
 - b. Serve as a vehicle for health education and counseling in order to promote voluntary health-related behavior change.
 - c. Serve as a group needs-assessment instrument for planning health education/health promotion programs.

There are two major types of health risk appraisals: self-scored and computer-analyzed. The computerized health risk appraisal gives an estimate of risk based upon physiological (e.g., blood pressure), biochemical (e.g., cholesterol), and health habit (e.g., smoking) data. It can provide an estimate of the risk of dying and/or having a serious adverse health effect from a certain cause within a specified time period. The appraisal also compares risk with the "average" risk for the same age and sex and provides an appraised age as well as an achievable age. In addition, it also can provide group analysis based on participants in the appraisal process.

Cooper and Huss (1981) recommend that an evaluator develop a Simulation Model with the help of a systems analyst if the evaluator is not familiar with advanced computer capabilities. The steps in the construction and utilization of a simulation model for evaluation purposes include:

1. Determine that a Simulation Model is an appropriate device. Where the system is simple and all interactions are clear, other techniques may be more cost-benefit effective. With a large number of observations and few variables, standard causal analyses are a possible preliminary or alternative approach.
2. Define the boundaries of the simulation. The scope must draw the line between events to be simulated in the model and those assumed to be externalities or inputs to the model.
3. Specify the key interactive components. These may be the individual program participants, the service units, or any decision maker. If the reaction of a group as a group is a key event, then that group is a necessary component.
4. Define the actions (outputs) of each component as a list of possible "events." These become the effect of previous system events and the cause of other system events.
5. Specify which components will be affected immediately and directly by an event. An example from our model is that the event of gasoline purchase fills the automobile tank, lowers the tank level for the station, and alters the space available for the station to pump gas.
6. Define the effect on a component of the events that component "receives." This frequently requires an assumption or working hypothesis. If a number of outcomes is possible, then assign a probability to each and have the program select at random to match the probabilities.
7. Begin simulating to see how the model operates. This involves making systematic changes in variables and any constants to see how the model responds. This process is sometimes referred to as a sensitivity analysis. (Cooper & Huss, 1981, pp. 28–29).

Cost Analysis Statistical significance of program benefits have long sufficed for determining the effectiveness of a program. Evaluation researchers began to realize that the public and the funding sources were not satisfied with numbers of statistical importance. Instead, what was wanted was a demonstration of the **practical benefits** of the program. To accomplish this, researchers showed that the benefits exceeded all adverse program effects, including budgetary costs; hence the term *benefit-cost analysis*. Thompson, Rothrock, Strain, and Palmer (1981) found that there was a problem using benefit-cost analysis in the realm of evaluation research. This problem concerned the evaluation of main program effects. For example, evaluators in hypertension control programs cannot monetarily value statistical decreases to control rates. In a situation such as this, researchers rely on the principle of cost-effectiveness; even if we do not know the value of achieving an objective, we do know that we wish to achieve the objective in a way that minimizes costs.

Cost-effectiveness analysts determine the ratio of monetary to nonmonetary program effects (e.g., the amount in dollars required per hypertensive patient who achieves control). These ratios do not determine if the programs should be implemented but instead are used comparatively by decision makers who would choose the least expensive program that produces the same result.

The cost-effectiveness ratio for social programs is expressed as follows (Thompson et al., 1981):

$$\frac{ML - MG}{B - R}$$

ML = monetary losses (program costs plus some induced costs)

MG = monetary gains (monetary costs that are averted)

B = benefits

R = risks (adverse side effects of the program)

For example, in a mammography screening program, the costs (*ML*) could be \$200,000; the monetary gains (*MG*) could be \$20,000 in savings of hospitalizations for patients; 15 could be the number of breast cancer deaths prevented (*B*); and 2 could be the mortality number for those dying from the chemotherapy and radiation treatments (*R*). The ratio then is

$$\frac{\$200,000 - \$20,000}{15 - 2} = \$13,846 \text{ per life saved due to the mammography screening program}$$

In cost analysis, the numerator of the ratio is the focus of attention. Determining the *costs* of a program is extremely important and can have important implications for program evaluation. In determining the costs of a program, the following must be considered: *direct costs*, which include budgetary and nonbudgetary direct costs; *indirect costs*; and *opportunity costs*. (For a complete discussion of these main categories of costs, see Thompson et al., 1981.)

Contextual Evaluation We have chosen to discuss ethnography here within the context of program evaluation. To reiterate the difference between experimental and contextual evaluations, *experimental evaluations* validate simplified causes and effects through controlled comparisons, and *contextual evaluations* attempt to understand the complexity of a program.

Contextual evaluations assume that social interventions have several facets and relationships that relate to multiple outcomes (Britan, 1981). These evaluations seek to determine how programs work, how they fit into settings, how they achieve results, and how they can be improved. Contextual evaluations are inductive and describe program implementation—how treatments occur, and how program activities relate to formal rules, informal goals, participant understandings, and environmental pressures. The major methodology for a contextual evaluation is ethnography.

Ethnography encompasses the use of observations and interviews and relies very heavily on participant observation. The hallmark of ethnography is the observation, recording, and analysis of behavior in context. It includes systematic descriptions of social systems that look for interrelationships among particular behaviors, customs, rituals, beliefs, and values in terms of broader patterns of cultural understanding, social structure, and environment (Britan, 1981). The usual format for an ethnographer is a case study in which an analysis of program implementation and impact is conducted.

There are certain advantages to using experimental evaluations: in cases in which program goals are clearly stated, treatments are straightforward, and theoretical constructs are unambiguous. On the other hand, contextual evaluations are more useful when the goals are broadly stated, treatments are complex, and theoretical constructs are vague. Therefore, the choice of an evaluation methodology may depend upon the type and setting of the program.

Case Discussion

The cardiac rehabilitation program was judged by Sarah's evaluation team from the University of Dover to be a cost-effective venture for the hospital. Their recommendation to the chief of staff of Rosewood Hospital was to retain the program. This decision was based upon a variety of material gathered by the study team. The major approach used by the Dover team was conducting a case study enhanced by survey data, meeting with hospital personnel, conducting patient interviews, and conducting a cost analysis of the cardiac rehabilitation program.

SUMMARY

This chapter has attempted to provide a condensed version of the field of evaluation research. This area of research generates much discussion and experimentation in the social sciences, with specific emphasis in educational evaluation. The case study depicted

such a project in which a program for cardiac rehabilitation patients was to be evaluated. A broad, general definition of evaluation research was established as the process of determining the effectiveness of a program.

The purposes of evaluation vary, depending upon the vantage point: administrator, consumer, evaluator, or organization. Almost anything can be evaluated, especially if there is support from one of the aforementioned groups.

There are many evaluation models, including behavioral objectives, systems analysis, decision-making, goal-free, connoisseurship, case study, and accreditation. A discussion of the advantages and disadvantages of each model led us to the conclusion that perhaps a combination of models might be necessary for many program evaluations. The following steps were enumerated to aid the evaluator in conducting an evaluation: determine the client, the purposes of the evaluation, the methodology to be utilized, the nature of the contact, how the evaluation should be conducted, and who will use the results of the evaluation.

The major types of evaluation research used in the health sciences include needs assessment, formative, summative, and responsive. Each type has its inherent strengths, weaknesses, and best area of suitability. It is generally up to the contractor to determine the type of evaluation, and this is accomplished by the nature of the evaluation research contract and the question that needs to be answered by the investigator. Methodological approaches utilized in evaluation research are experimental, quasi-experimental, correlation, surveys, personnel assessment, expert judgment, and case study. Though these approaches are used in other research paradigms, they are discussed with specific reference to evaluation research. Innovative research methodologies were discussed, including simulation, cost analysis, and contextual evaluation.

Evaluation research is a valuable tool for the health sciences. Programs have been asked to be accountable, and through a well-designed evaluation, that accountability can be assessed.

CRITICAL THINKING QUESTIONS

1. Delineate the major purposes for evaluating a program.
2. Who do you believe should be part of the evaluation team?
3. What should a program evaluation accomplish?
4. Choose one of the evaluation models, and explain why you would utilize it in your program evaluation.
5. Distinguish between formative and summative evaluation.
6. Explain the role of stakeholders in the evaluation process.
7. Using the case study presented in the chapter, describe how you would evaluate who the decision makers are before beginning evaluation research.

SUGGESTED ACTIVITIES

1. Refer to the list in the section *What Can Be Evaluated?* and add to the evaluation possibilities.
2. From the list in Activity 1, choose one topic and briefly describe the steps you would take in the evaluation process.
3. Interview a program administrator to assess that program's evaluation status (i.e., when was it last evaluated, by whom, why, and so on).
4. You have been asked by your employer to locate the most recent references with regard to health program evaluation. Utilizing the Web, determine at least 10 references that would be beneficial to you in this assignment. Annotate each reference, then rank-order them with a view to their importance and helpfulness.
5. Utilizing the Web, evaluate any health-related program that is complete. The program should be of some interest to you and somewhat interactive. In doing your evaluation, state what type of evaluation methodology you will use, describe the evaluation plan, and predetermine the possible outcome of your evaluation.

REFERENCES

- Anderson, S., & Ball, S. (1978). *The profession and practice of program evaluation*. San Francisco: Jossey-Bass.
- Anderson, S., Ball, S., Murphy, R., & Associates. (1975). *Encyclopedia of education evaluation*. San Francisco: Jossey-Bass.
- Baker, F. (1974). Formative evaluation of instruction. In W. Popham (Ed.), *Evaluation in education*. Berkeley: McCutchan.
- Bloom, B. (1984). *Taxonomy of educational objectives*. New York: Longman.
- Britan, G. (1981). Contextual evaluation: An ethnographic approach to program assessment. In R. Conner (Ed.), *Methodological advances in evaluation research*. Beverly Hills: Sage Publications.
- Chelmsky, E. (1978). Differing perspectives of evaluation. In C. Rentz & R. Rentz (Eds.), *Evaluating federally sponsored programs: New directions for program evaluation* (Vol. 2, Summer). San Francisco: Jossey-Bass.
- Cook, T., & Campbell, D. (1994). The design and conduct of quasi-experiments in field settings. In M. Dunnette (Ed.), *Organizational psychology* (2nd ed.). Chicago: Rand McNally.
- Cooper, R., & Huss, W. (1981). Simulation as an evaluation tool. In R. Conner (Ed.), *Methodological advances in evaluation research*. Beverly Hills: Sage Publications.
- Eisner, E. (2002). *The educational imagination* (4th ed.). New Jersey: Prentice Hall.
- Guba, E., & Lincoln, Y. (1994). *Fourth-generation evaluation* (2nd ed.). Newbury Park, CA: Sage Publications.
- Harless, J. (1973). An analysis of trend analysis. *Improving Human Performance: A Research Quarterly*, 3, 229–244.
- House, E. (1999). *Evaluating with validity*. Beverly Hills: Sage Publications.
- Mager, R. (1962). *Preparing objectives for programmed instruction*. San Francisco: Fearon.
- Patton, M. (2002). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage Publications.

- Popham, W. (1993). *Educational evaluation*. Boston: Allyn & Bacon.
- Rossi, P. H., Lesley, M. W., & Freeman, H. E. (2004). *A systematic approach evaluation* (7th ed.). Thousand Oaks, CA: Sage Publications.
- Rutman, L. (Ed.). (1984). *Evaluation research methods*. Beverly Hills: Sage Publications.
- Scriven, M. (1967). *The methodology of evaluation* (AERA Monograph Series on Curriculum Evaluation No. 1). Chicago: Rand McNally.
- Scriven, M. (1973). Goal free evaluation. In E. House (Ed.), *School evaluation: The politics and process*. Berkeley: McCutchan.
- Shortell, S., & Richardson, W. (1978). *Health program evaluation*. St. Louis: C. V. Mosby.
- Stake, R. (1991). Responsive evaluation and qualitative methods. In W. Shadish, T. Cook, & L. Leviston (Eds.), *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage Publications.
- Stake, R. (1998). *Evaluating art programs in community centers*. Stamford, CT: JAI Press.
- Streuning, F., & Brewer, M. (Eds.). (1983). *Handbook of evaluation research*. Beverly Hills: Sage Publications.
- Stufflebeam, D. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.
- Suchman, E. (1976). *Evaluative research*. New York: Russell Sage Foundation.
- Thompson, E., Rothrock, J., Strain, R., & Palmer, R. (1981). Cost analysis for program evaluation. In R. Conner (Ed.), *Methodological advances in evaluation research*. Beverly Hills: Sage.
- Tyler, R. (1949). *Basic principles of curriculum instruction*. Chicago: University of Chicago Press.
- Windsor, R., Baranowski, T., Clark, N., Boyd, R., & Goodman, R. (2003). *Evaluation of health promotion, health education, and disease prevention programs*. New York, NY: McGraw Hill.

Analytical Epidemiologic Studies

KEY TERMS

cohort	incidence	prospective studies
case control investigations	incidence rate	random variation
confounding	information bias	retrospective cohort study
cross-sectional study	odds ratio	selection bias
current cohort study	person-time	
epidemiology	prevalence	

Case Study A

Henry, a physician, had just completed his medical fellowship in geriatric medicine and had accepted a position with the state's regional geriatric center. During his training, he became intrigued by the differences and similarities among adult children who had to place a parent in an assisted living facility. During the 5 years of his residency and fellowship experience, he learned that some adult children not only coped very well but actually appeared to have personal growth. In contrast, others seemed to dwell in guilt, grief, personal sadness, and depression. He knew that being an adult child of a parent with Alzheimer's disease or related dementia could contribute to the differences among these adults. In his new position, he decided to investigate this event by getting baseline measurements on demographics of the adult children: amount of time spent caring/visiting, the economic costs, coping skills, depression, self-rated health status, nature and type of support for them, in addition to information about the patient (to include functional physical and cognitive status, severity of the dementia, and length of time since diagnosis).

Case Study B

As an epidemiologist for the state department of health, Cortni was asked to investigate a case in which an employee who had previously worked in a factory was now suffering

from acute myeloid leukemia (AML). Inquiry by the patient's physician revealed that four friends who had worked in the factory with him had all died from AML during the past 2 or so years. In addition, although the patient had not been in touch with many of those who had worked there for a year or two during that same time frame, he had heard that at least one of them was also suffering from AML. The patient noted that some strong chemicals were used as part of the production process in the factory. He began work there about 15 years ago and remained an employee for 6 years. He had saved money and left that position to begin a small realty company. The patient and his physician believed that exposure to the chemicals, even though it had been years ago, was the cause of his AML. Cortni was given the task to determine the veracity of their belief.

The Nature of Epidemiology

As a discipline, epidemiology knows few, if any, boundaries. This is to be expected because health is multidimensional. Therefore, epidemiologic investigations are connected to the biomedical sciences, behavioral sciences, computer sciences, engineering, and other disciplines. Rockett (1994) has embraced this concept by stating that **epidemiology** is "the study of our collective health" (p. 2). As such, it comprises two facets, one of which is descriptive, involving the identifying of patterns, trends, and disease and injury differentials. The second facet moves beyond the descriptive approach to embrace causation or etiology. This latter approach is called *analytic epidemiology*.

Analytical Methodologies in Epidemiology

There are two major analytic methods used in epidemiologic investigations: cohort and case control. The purpose of each is to test the hypothesized cause-and-effect relationship between a suspected risk factor and a disease, injury, or even a social condition such as welfare status.

Cohort studies, sometimes called **prospective studies**, work from a postulated cause to an effect. Therefore, the general design is to begin with a group of people (a cohort) and observe them over a period of time. Selection for the group can be based upon the presence or absence of a characteristic (e.g., diabetes, spinal cord injury, welfare) or at random. The individuals within the group, as to be expected, will vary in exposure to one or more of the factors being observed. The epidemiologist watches to determine the differences in the rate at which the characteristic occurs (disease, injury, behavioral problem) in relation to exposure to the factor(s).

Case-control investigations, or retrospective studies, start with a group of people who already have the characteristic (health problem) and compare them with people who do not have that health problem (characteristic). The people who have the health problem are referred to as "cases," while those in the other group are termed "controls." In searching for the "cause," the case-control method determines if the two groups differ in the degree of exposure to different factors. Figure 10.1 illustrates the timing differences between cohort and case-control studies.

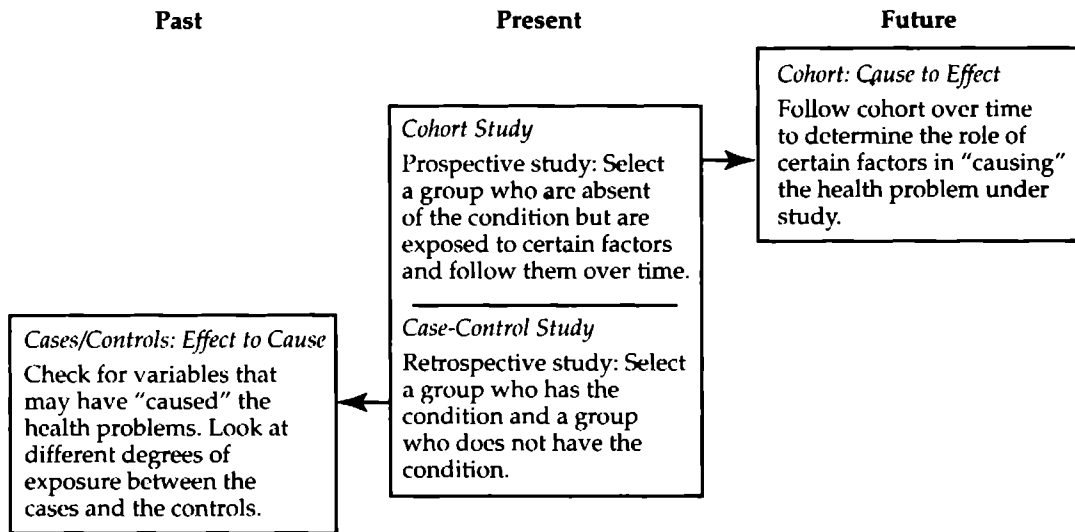


Figure 10.1 Timing Schematic for Cohort and Case-Control Prospective and Retrospective Investigations

Another way to compare the similarities and differences between these two analytic methods is shown in Figure 10.2. Herein, a 2 × 2 table shows how groups can be categorized as well as the timing direction of the particular method. Once again, note how the cohort study moves forward (cause-to-effect), similar to an experimental design. The case-control approach, sometimes referred to as *ex post facto*, moves backward in time (effect-to-cause).

Figure 10.2 shows that when a positive relationship exists between the health problem (behavioral, medical, public) and the exposure factor(s), those who are exposed (Group *a*) will be greater in number than those who are not exposed (Group *d*). Logically there should be a larger concentration of subjects in Cells *a* and *d* than in Cells *b* and *c*. This holds true with both the cohort and case-control methods.

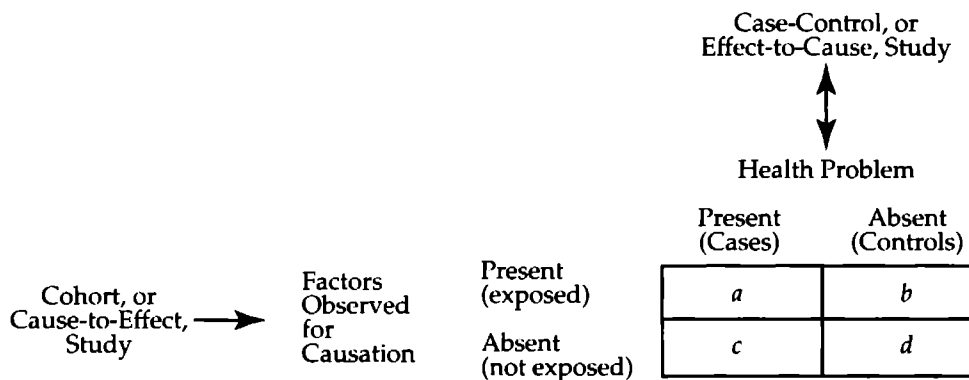


Figure 10.2 A 2 × 2 Table Illustration of Cohort and Case-Control Methodologies

Both Figures 10.1 and 10.2 are basic models. Complexity can arise when the health problem is subdivided into categories or when the risk factors are analyzed in degrees of exposure rather than all or none. Examples of subdividing health problems include drug abuse, which can be subdivided into alcohol, cocaine, tobacco, and so forth. Similarly, coronary heart disease (CHD) can be broken into angina pectoris, myocardial infarction, and sudden death subcategories.

Cohort Investigations

A **cohort** is simply a group of people with a common experience over a defined period of time. For example, a marital cohort consists of all people married within a certain time period. A disease (e.g., juvenile diabetes) cohort may be all those patients with juvenile diabetes who presented at Hospital A during a particular time frame. All members of the cohort are assumed to be free of the health problem at the commencement of the investigation. They are divided into two or more groups, according to exposure to the risk factors under investigation. Using vasectomy as a risk factor, Jamieson and colleagues (2004) examined the pregnancy rates of women whose husbands had a vasectomy by tracking the women over a 2-year period. Simon, Peters, Christiansen, and Fletcher (2000) investigated the effect on patient satisfaction of medical student participation in care and the presence of medical student teaching. One group of patients were involved with medical students, while another group was not. In another educational study, Chok and Gomez (2000) examined the effect of a neck care talk to patients receiving physiotherapy at a neck pain clinic. Two groups were used, with one receiving the educational intervention.

In our case study with Henry, we find that his cohort group is all the adult children who place a parent suffering from Alzheimer's disease in an assisted living facility. He knows that he can collect information for the length of time the parent resides in the facility. However, he can set his follow-up period for a different length of time, such as 4 years. In epidemiologic terms, he would conduct a **current cohort study**, which means the subjects are selected at the beginning of the study (present time) and followed over a period of time (future time). The cohort model can be modified by making it a retrospective in nature rather than a prospective one (Hulley et al., 2007). In this scenario, called a **retrospective cohort study**, the cohort assembly, baseline data, and follow-up took place in the past. Kauffman and colleagues (2008), in a study of metabolic differences among phenotypic expressions of polycystic ovary syndrome, used this approach, as did the study by Chen and colleagues (2008) examining the association between teenage pregnancy and neonatal and postnatal mortality. Of course this type of study demands adequate data about risk factors and outcomes on the cohort of subjects, which can be problematic. Current cohort studies have the advantage of greater research control over data collection.

There are several ways to select a cohort. One way is simply by accessibility, as in Henry's case, while other ways include history of exposure or availability of medical records. A researcher can take a random sample of exposed and unexposed individuals. The cohort is then divided according to their exposure to the risk factor(s). This sample is random in that, among exposed and unexposed participants, there is an equal chance of studying subjects

Table 10.1 Advantages and Disadvantages of the Cohort Strategy

<p><i>Advantages</i></p> <ol style="list-style-type: none"> 1. The cohort is categorized by exposure to the risk factor(s) before the health problem occurs. 2. Provides for the study of multiple potential effects of exposure, both positive and negative. 3. Permits calculation of incidence rates among those exposed and those not exposed. Specifically, exposure-specific incidence rate, attributable risk rate, relative risk, and odds ratio can be determined. 4. Provides flexibility in variable selection. 5. Allows for a large degree of quality control in the research effort. <p><i>Disadvantages</i></p> <ol style="list-style-type: none"> 1. Large numbers are generally required, which can increase expense and time requirements. 2. Study duration is usually long. 3. Lengthy duration can present problems with attrition and change in subject status as it relates to the risk variables (e.g., smoking, residence). 4. Follow-up can be difficult as the length of the study increases. 5. Difficult to control extraneous variables. 6. Almost impossible to study detailed mechanisms.
--

who will develop the health problem and subjects who will not develop the health problem. In our case study, Henry could use his accessible population and group them according to exposure to such risk factors as family support, cost, and severity of disease.

At the conclusion of the study, the researcher compares the two groups for the **incidence rate** of the health problem. In a cohort study, estimations can be made for exposure-specific incidence rates, attributable risk, relative risk, and the odds ratio. These will be discussed in a later section. The advantages and disadvantages of the cohort method are outlined in Table 10.1.

Case-Control Studies

A case-control study requires the investigator to work backwards in time (Newman, Browner, Cummings, & Hulley, 2007). The sample is drawn from a population of patients with the outcome (the cases). In Cortni's case study, she would draw from the population of patients who worked at the factory and suffered from AML (AML is the outcome). She would also select another sample from a population who does not have the outcome (the controls). The investigator then compares the predictor variables to determine which ones contributed to the outcome. Cortni would look at such predictor variables (risk factors) as smoking, exposure to chemicals or radiation, inherited syndromes, and blood disorders. Case control investigations are generally much less expensive than cohort investigations and cross-sectional studies that will be addressed later in this section. In comparison, most case control studies require fewer subjects as noted by Newman and colleagues (2007). Some investigators reached definitive conclusions with only seven cases (Herbst, Ulfelder, & Paskanz 1971).

As a general rule, it is best to select the case group from newly diagnosed cases that meet the criteria within a designated time frame. Sources for case groups can be records of physicians and employers, HMO groups, hospital records, death certificates, and so forth. Of course, information can be gathered from subjects themselves or even from proxies for the subjects or patients. Oftentimes a combination of records and actual subjects are used (Landmark & Abdelnoor, 2000). In Cortni's case, as in many case-control studies, selection of cases is often simple in that the number of sources is very limited.

In regard to control groups, the obvious point is to procure a sample from a population at risk for the outcome. Frequently investigators may seek controls from hospitals or clinics. While this may be feasible, depending upon the study, the researcher must be careful not to allow bias to enter the equation since many hospital/clinic patients are already afflicted with some malady. The bottom line is that controls should be representative of the general population in terms of probability of exposure to the risk factor(s). Preferably, they should have had the same opportunity to be exposed. As you would logically expect, it is important that sources and methods of data collection are similar for both cases and controls.

The population of controls for Cortni should be similar to her cases in exposure to the risk factors for AMI. While she would make every attempt to find cases similar to

Table 10.2 Advantages and Disadvantages of the Case-Control Method

Advantages

1. The number of subjects can be small. This is because the study starts with identification of cases and a like number of controls.
2. A study can be planned quickly, carried out in an expedient manner, and analyzed in a short period of time.
3. More than one risk factor can be identified in the same data set.
4. Requires few subjects, making it excellent for studying rare diseases.
5. There is no risk to subjects.
6. Subjects do not need to volunteer.
7. Minimal ethical problems.
8. Relatively inexpensive.

Disadvantages

1. The information necessary to the investigation may be unavailable or inaccurate.
 2. May be difficult or impossible to validate some records.
 3. Recall by a subject or informant for the person with the health problem may be quite different than that of controls.
 4. Selection of an appropriate control group can be difficult. For example, selection from hospital or clinic records can introduce bias, depending upon the risk factors and the nature of the controls.
 5. Rates of disease in exposed and nonexposed subjects cannot be determined.
 6. The odds ratio is the only calculation available to case-control studies. In cases of rare diseases (health problems), the relative risk can be calculated because in these instances the odds ratio approximates relative risk.
 7. Difficult to control extraneous variables.
 8. Difficult, if not impossible, to study detailed mechanisms.
-

the patient presented to her (factory worker who worked during the same time period under similar or identical conditions and contracted AML), she also would need to seek a sample of controls that is comparable. She may even elect to use matching* (discussed in detail later) on principal risks that are related to AML but not necessarily of great importance to her, such as age and gender. Another approach is to use two or more control groups (Newman, et al., 2007; Hurwitz, et al., 1987).

In comparison to the case-control study, the epidemiologic **cross-sectional study** has the investigator procuring all the measurements on a single occasion or at least within a very short period of time (Newman, et al., 2007). Generally speaking cross-sectional studies are descriptive in nature. An example would be the National Health and Nutrition Examination Survey (NHANES). Table 10.2 summarizes the advantages and disadvantages of the case-control method.

Establishing Causation

As noted at the outset of this chapter, analytic epidemiology attempts to establish causation between the health problem and selected risk factors. In the absence of experimentation, the following can be of value in determining causation (Rockett, 1994; Schlesselman, 1982):

Temporal sequence: Causative factors occur before effects (events). This seems obvious but is not always readily identifiable. For example, is endometriosis the cause or consequence of the disease process that leads to infertility?

Consistency: If the association of the health problem and risk factor(s) is evidenced in several different conditions of the study, it gives credence to causation.

Strength of association: Generally, the larger the association between risk exposure and the health problem, the less likely the relationship is spurious. In other words, the larger the value of relative risk, the more likely is causation.

Specificity of effect: This is rather simplistic in that the effect should be specific to the risk factor(s). If the risk factor is removed, the effect should be also. Of course, most events do not stem from a singular cause, which makes this approach weak, but it may be of some value when combined with other approaches to causation.

Biologic gradient: A dose-response relationship makes an excellent argument for causation.

Existing data and theory: When the relationship between risk factor(s) and the health problem is consistent with previous research efforts or perhaps theoretical constructs, then causation is more likely to be established.

Although causation may never be fully proven in the experimental sense (e.g., smoking and lung cancer), analytic epidemiology can assess the likelihood of causation. In a real world with many research constraints, we often cannot ask for any more.

Problems of Error

Like other research approaches, analytic investigations can succumb to error. Rockett (1994) and Newman et al. (2007) describe three principal types of error: bias, random variation, and random misclassification. This section addresses these errors along with suggested ways to prevent them.

Bias

Bias refers to any error in design, conduct, or analysis of a study that makes the estimate of an exposure's effect on the risk of disease inaccurate. One way this can happen is through **selection bias**. In this instance, poor research design brings about a difference in comparison groups (e.g., diseased and nondiseased, exposed and not exposed) that misrepresents true results. As you would expect, bias is a threat to internal validity. To prevent this source of error, cohort groups should be homogeneous in as many factors as possible, except for the risk factors under investigation. In case-control studies, the only difference between the two groups should be their disease status or health problem. Matching is one technique to assure similarity of several factors and will be discussed in a later section.

A second form of bias is **information bias**, which occurs in cohort studies when information about disease outcome is not collected uniformly across groups. In case-control studies, information bias occurs when exposure information is not collected uniformly for cases and controls. The prevention step is obvious—collect information uniformly. This means that you must monitor those people who are helping you collect data. Another way to reduce information bias is to “blind” the study to both the subjects and the data collectors. In other words, do not reveal any more about the study than you have to for both ethical and data collection reasons.

A third form of bias is termed **confounding**. Confounding variables are extraneous variables that are (1) risk factors for the health problem under investigation; and (2) factors that are associated with the exposure of interest but that are not a consequence of exposure. For example, in Cortni's case-control study, gender could be a confounding factor since AML is more common in males, although no one is sure why. Further, if Cortni were just to look at AML and exposure to chemicals in the factory setting, she may find that cases actually smoked much more than controls. That would be problematic since as many as 1 in 5 cases of AML is caused by smoking. The two principal ways to control for confounding variables are by matching and by using statistical techniques. These are discussed in the section titled *Control in Epidemiologic Research*.

Random Variation

As the name implies, **random variation** means chance differences between groups. This source of error affects the ability to generalize to a larger population—external validity. It generally occurs because of unrepresentativeness in the comparison groups. Henry and Cortni can minimize random variation in their studies by increasing the size of their study and comparison groups.

Random Misclassification

Errors can be made when determining a subject's exposure status or disease (health problem) status. In other words, a subject could be misclassified. In Henry's cohort investigation, he may misclassify an adult child on any one of the variables or risk factors under consideration. He may even misallocate the adult child whose parent was admitted for Alzheimer's disease by mistaking them to be well adjusted when in fact they are not. It is less likely that Cortni will misclassify participants in her study since the controls will not

have AML, although, if she selects workers who underwent the risk and do not have AML, there is the possibility that AML could be diagnosed at a later date. If this happened in either investigation, the results are likely to underestimate the real association between exposure and effect. The possibility of finding a strong causal-like relationship between risk factor and health problem is greatly reduced, if not eliminated, in the presence of random misclassification.

Control in Epidemiological Research

While the preceding section addressed some ways to control for error in analytic epidemiologic research, this section expands on some of the more complex methods.

Matching

One of the most common methods to reduce error is to match on a subject-to-subject basis. For example, in Cortni's case-control study, each case individual can be matched with each control individual on several factors. For example, a 50-year-old White male with AML could be matched with a 50-year-old White male who does not have AML. Similarly, matching could occur on the smoking risk factor. Matching should be done on relevant risk factors as demonstrated by Niccolai et al. (2007), who looked at the methodological issues in design and analysis of a matched case-control study of vaccine effectiveness.

If subject-to-subject matching is not feasible, group matching may be the next best choice. For example, Cortni could go with a mean age or perhaps with age categories (e.g., 10-year age groups). Group matching at least keeps cases and controls in similar proportions. Sometimes more than one control individual needs to be matched to a case individual. When matching multiple controls to each case, a constant ratio should be used, such as 1:2 or 1:3 (Schlesselman, 1982). Like many aspects of research, however, this looks great on paper but may not be functional in the real world, as seen by studies that attempt this and end up with incomplete matching.

While the intent of matching is to eliminate biased comparisons, researchers must be careful not to overmatch. Overmatching reduces validity or statistical efficiency. Remember that a matched factor cannot be evaluated for its etiological role. Another problem with overmatching is cost—depending on the factor, matches can be difficult to locate.

Logically, if matching is part of the overall design, then the analysis should reflect that design. In other words, matching has two steps: (1) matching the design; and (2) matching the analysis. If Step 2 is not carried out, the estimated relative risk is likely to be biased.

Homogenous Grouping

Another procedure is to use groups that are as homogeneous as possible. In this approach, you would select a sample in which all the subjects would be homogeneous on the variable(s) in question. For example, if level of education were a confounding variable, its

effect could be controlled by the use of subjects who all have the same education (e.g., high school). In this manner, the effects that are found can be justified as stemming from the independent variable more readily (Ary, 2006).

The use of homogeneous groups is not always possible, however, because some of the extraneous variables either may not be identified or may not be present in large enough numbers to carry out the research. Further, if control is accomplished on only one variable, such as level of education, you cannot generalize findings to other levels of education.

Stratified Sampling

This strategy is similar to homogeneous grouping in that subgroups are formed by separating the ranges of selected variables and sampling a predetermined number of cases and controls within cells made from the multiple cross-classification. Using the education example from above, you could look at education as a whole and subdivide it into three groups: elementary school, high school, and college. It would then be a matter of determining how many should be in each educational cell, or subgroup. Controls should be in a constant ratio to the case subgroup—1:1, 1:2, and so forth. The actual numbers may vary across the strata, but the case-control ratio would be the same.

Poststratification

Matching, homogeneous grouping, and stratified sampling demand that you identify variables to control before commencing the study. Poststratification, as well as regression analysis, avoids this requirement. Simply, poststratification means that you classify unmatched cases and controls by their values on one or more variables determined during the study. This strategy is similar to stratified sampling except that (1) the variables used for grouping do not have to be selected ahead of time; (2) the case-control ratio does not have to be identical across strata; and (3) not all subgroups will contain both cases and controls, making comparison impossible in some instances. Needless to say, this would be a definite drawback for some research efforts.

Multivariate Analysis

Like poststratification, multivariate analysis is conducted after the data have been collected. Research has shown that we can rarely rely on a two-variable study to explain, predict, or control relationships or variables. Instead, we must deal with several variables simultaneously. To do so, we must use multivariate procedures. While several of these are discussed in detail in Chapter 12, on inferential analysis, the outline below describes their usage.

Analysis of Covariance The dependent variable is continuous in nature, whereas the independent variables are a mixture of nominal and continuous variables. The continuous variables are used as control variables.

Multiple Regression Analysis The dependent variable is generally continuous, as are the independent variables, although in the research world this technique is used with any type of independent variable. The principal objection to linear regression is that the parameters have a limited range of validity and interpretation (Schlesselman, 1982).

Logistic Regression Also called logit analysis, logistic regression analyzes relationships between several independent variables and a dependent variable that is nominal. Remember, a nominal variable is categorical (e.g., gender). It is an either/or classification system with no ranking. Logistic regression uses maximum likelihood estimation for analysis, thereby transforming the likelihood (probability) of an event occurring into its odds.

Linear Structural Relation Analysis (LISREL) This is a complex statistical procedure that, like logistic regression analysis, is based on maximum likelihood estimation. It is a versatile approach and is often used much like path analysis. It is often a better approach than path analysis, however. Since the data collected by health researchers usually contain some type of error, LISREL is preferred over least-squares regression in doing path analysis. Further, least-squares regression operates on the assumption that residuals in the various regression equations are not correlated, an assumption that is difficult to accept. Finally, path analysis usually assumes a unidirectional flow in causation, and we know that in the area of health, cause and effect are often iterative or reciprocal.

Proportional Hazards Regression In many health investigations, the end point of interest is survival. This is particularly so when studying chronic disease. Educators and researchers alike use estimated survival curves for teaching or explaining survival data. The Kaplan-Meier estimate of a survival curve, referred to as an *actuarial estimate*, has long been used. However, specific regression models have been developed to consider the simultaneous influence of several explanatory variables on survival time. Matthews and Farewell (2007) provide a detailed usage of proportional hazards regression with an example of two lymphoma groups. One group presented with clinical symptoms, while the other was asymptomatic. The regression is used to determine why the asymptomatic group has a longer survival rate. Specifically, what factors played a role to create this difference? These authors also discuss this regression technique in terms of time-dependent covariates.

Analysis of Results in Analytic Epidemiology

As discussed at the outset of this chapter, the purpose of both cohort and case-control studies is to test the hypothesized cause-effect relationship between a suspected risk factor(s) and a disease, injury, or social/health condition. This section presents information about measures of association for each type of investigation.

It is important to understand some basic terms prior to discussing the analyses.

Prevalence

Prevalence is a measure of the proportion of individuals within a population who have a specific health problem, disease, injury, or other health event at a particular point in time. It can be expressed as:

$$\text{prevalence} = \frac{\text{number of people who have the health problem at a point in time}}{\text{number of people in the study program}}$$

Incidence

A subcategory of prevalence is **incidence**, in that it is the number of new cases in a population at risk during a specified time period (e.g., a calendar year or duration of a study). Rosner (2006) interprets the incidence rate as “the probability an individual with no prior disease will develop a new case of the disease over some specified time period” (p. 63). The incidence rate is expressed as:

$$\text{incidence} = \frac{\text{number of new cases over a period of time}}{\text{population at risk}}$$

The incidence rate formula is sometimes modified to fit the nature of the study. For example, the numerator could be the number of conditions rather than the number of people. This would reflect the fact that a person may get the condition more than once during the specified time period. For example:

$$\text{incidence} = \frac{\text{number of colds in a 6-month period}}{\text{number of people at risk}}$$

Similarly, for some cohort (prospective) investigations, the denominator can be changed to reflect the problem of attrition. For example, in Henry’s study some individuals will move away, drop out of the program, or somehow get lost for follow-up. Also, other individuals will enter the study after it is initiated. Both these situations mean that there will be an unequal period of observation time for subjects. As a result, subjects will contribute unequally to the calculation of population at risk. To overcome this dilemma, epidemiologists use the idea of “**person-time**.” This is often, but not always, shown as person-years and is used in the denominator of the incidence formula. Table 10.3 is a hypothetical example of person-years calculation for incidence in Henry’s study.

From Table 10.3 we find that 15 people were followed for a period of 53 person-years. If 2 of the 15 had personal growth, the incidence rate would be 2 in 53 person-years, or 0.0377 per person-year. This can be expressed as 3.77 per 100 person-years of observation. The calculation is as follows:

$$\begin{aligned} \text{incidence} &= \frac{\text{number of adult children who had personal growth over 5 years}}{\text{person-time at risk}} \times 100 \\ &= \frac{2}{53} \times 100 = 3.77 \end{aligned}$$

Of course, the multiplier could be 1,000 or 100,000 rather than 100 person-years.

The use of person-years rather than population units should bring greater accuracy to the measurement. However, in using person-years, the following three conditions must be met. First, the risk of the condition must be constant throughout the period of study. Second, the rate of the condition for those who are lost to follow-up should be about the same as for those who remain in the study. Third, if the condition under study is a rapidly fatal disease, the rate could be artificially high if certain subjects are studied for less than the full period of time.

Table 10.3 Person-Years Calculation for Incidence

Adult Child	Number of Years Observed
1	4
2	5
3	1
4	4
5	2
6	5
7	5
8	2
9	3
10	5
11	4
12	3
13	5
14	4
15	1
Total Person-Years	53

Analyses for Cohort Investigations

In cohort studies, investigators measure the strength of the association between exposure and the disease or health outcome by means of the rate ratio or relative risk. Relative risk (RR) is defined as the ratio of the incidence rate for persons exposed to a risk factor to the incidence rate for those not exposed to that same risk factor. It can be expressed as:

$$\text{relative risk} = \frac{\text{incidence rate for those exposed to risk factor}}{\text{incidence rate for those unexposed to risk factor}}$$

Table 10.4 illustrates a typical 2×2 table for a cohort study using population units rather than person-time units.

Using the cells in Table 10.4, relative risk would be:

$$RR = \frac{a / (a + b)}{c / (c + d)}$$

Table 10.4 A 2×2 Cohort Table for Determining Relative Risk Based on Population Units

Risk Factor	Disease Present	Disease Absent	Total	Incidence Rate
Exposed	<i>a</i>	<i>b</i>	<i>a + b</i>	<i>a/a + b</i>
Not exposed	<i>c</i>	<i>d</i>	<i>c + d</i>	<i>c/c + d</i>

RR, or risk ratio, shows the extent to which it is more (or less) likely that a health problem or condition will occur in the exposed group as compared to the unexposed group. If RR equals 1, there is no relationship between exposure and the condition or health problem. If RR is greater than 1, a positive association exists, meaning that those exposed are so many times more likely to contract the health problem. In contrast, if RR is less than unity, there is a negative association, implying that the exposed group is less likely to contract the health problem. •

If person-time units rather than population units are employed in the study, the 2×2 table would need to reflect this change, and the relative risk formula would be modified (Rockert, 1994). Table 10.5 shows the change.

Using the cells in Table 10.5, a relative risk would be:

$$RR = \frac{a/e}{c/f}$$

To illustrate this further, Table 10.6 presents hypothetical data for Henry's cohort study on the well-being of adult children of Alzheimer's patients. In other words, imagine that after 5 years of investigation, his results would be those in Table 10.6. Here, "Exposed" refers to having a parent with Alzheimer's disease and "Not exposed" refers to not having a parent with Alzheimer's.

Note that person-time units in years are used to develop the incidence rates, which in turn are multiplied by 100,000 to develop the incidence rates per 100,000 person-years. The RR or rate ratio, using our formula for person-years, is:

$$RR = \frac{466}{77} = 6.05 \text{ per } 100,000 \text{ person-years}$$

This is interpreted as a positive association between the risk factor, such as severity of the parent/patient's disease, and well-being. While it cannot be said that this risk factor "causes" well-being, Henry can say that those exposed to the risk factor are six times more likely to experience poor well-being than those who were not exposed to that risk.

Another rate that can be used in cohort studies is the attributable risk rate, which is simply the difference between incidence rates. Specifically, it is the incidence rate for the exposed group minus the incidence rate for the unexposed group. This rate, also called the risk difference, indicates the magnitude of the absolute change brought about by exposure. Suppose we have two people who are at equal risk for the health

Table 10.5 A 2×2 Cohort Table for Determining Relative Risk Based on Person-Time Units

Risk Factor	Disease Present	Disease Absent	Person-Time Unit	Incidence Rate
Exposed	a	b	e	a/e
Not exposed	c	d	f	c/f

Note: Person-time unit replaces population unit in the denominator for incidence rates.

Table 10.6 Hypothetical Results from a Cohort Study of Well-Being

Risk Factor	Well-Being Status		Person-Time Units	Incidence Rate	Incidence Rate per 100,000
	Yes	No			
Exposed	10	560	2,148	0.00465	466
Not exposed	2	610	2,590	0.00077	77

Note: Relative risk = 6.05 per 100,000 person-years.

problem except for the presence or absence of exposure to the risk factor. The risk of disease for the exposed person is the incidence rate for exposure, while the incidence rate for nonexposure is the rate for our other individual. If the exposure did not exist, then the risk for our first person would be the same as for the second. Subsequently, the difference between the two incidence rates represents the increase (or decrease) in risk owing to the exposure. In our example with Henry's data in Table 10.6, we would have the following:

$$\begin{aligned}
 \text{attributable risk} &= \text{incidence rate for exposed group} \\
 &\quad - \text{incidence rate for unexposed group} \\
 &= a/e - c/f \\
 &= 466 - 77 \\
 &= 389 \text{ per } 100,000 \text{ person-years}
 \end{aligned}$$

These two methods of calculating rate ratio or relative risk need to be modified if matching was used in the epidemiologic study design (Fleiss, Levin, & Paik, 2003). As noted in the section on matching, in such studies the unit of analysis is the matched pair. Table 10.7 illustrates the possible outcomes for a matched cohort investigation.

Table 10.7 A 2 × 2 Table for Relative Risk in a Matched Cohort Investigation

Exposed Subjects	Not Exposed Subjects	
	Diseased	Disease Free
Diseased	<i>a</i>	<i>b</i>
Disease Free	<i>c</i>	<i>d</i>

where

Cell *a* = exposed who get the disease + nonexposed who get the disease
 Cell *b* = exposed who get the disease + nonexposed who are disease free
 Cell *c* = exposed who are disease free + nonexposed who get the disease
 Cell *d* = exposed who are disease free + nonexposed who are disease free

Table 10.8 A 2 × 2 Table for a Case-Control Study

Exposure to Risk Factor	Disease Status	
	Diseased (Cases)	Disease Free (Controls)
Exposed	<i>a</i>	<i>b</i>
Not exposed	<i>c</i>	<i>d</i>
Proportion exposed	$a/a + c$	$b/b + d$

In Table 10.7, those in Cells *a* and *d* are concordant pairs, whereas those in Cells *b* and *c* are discordant. Only the latter are used for the relative risk ratio, which becomes Cell *b* divided by Cell *c*, or, in other words, *b/c*. McNemar's test is generally used to test significance.

Analyses for Case-Control Investigations

The **odds ratio** is the most common measure of association between exposure and the health outcome in case-control or retrospective investigations. The odds ratio can be seen as an estimate of the relative risk. Specifically, the odds ratio is the ratio of the odds of disease in exposed individuals relative to the odds of disease in unexposed individuals. Table 10.8 illustrates a typical 2 × 2 table for a case-control study.

The odds ratio, signified by the sign ψ , is calculated as:

$$\text{odds ratio } (\psi) = \frac{ad}{bc}$$

Interpretation is similar to relative risk in a cohort study. That is, an odds ratio of 1 implies no association between outcome and exposure. However, when the odds ratio is greater than unity, the exposure is positively related to the outcome (thereby contributing to it). On the other hand, a negative relationship between the exposure and the outcome exists when the odds ratio is less than 1. Let's use some hypothetical data from Cortni's investigation on AML. Table 10.9 illustrates her hypothetical data.

Table 10.9 Hypothetical Data from a Case-Control Study of Acute Myeloid Leukemia (AML)

Exposure to Risk Factor (Benzene)	Disease Status	
	Diseased (Cases)	Disease Free (Controls)
Exposed	40	20
Not exposed	15	33
Total	55	53

Note: Odds ratio = 4.4

The risk factor selected is exposure to the chemical benzene that was used in the factory. Using the formula, the odds ratio is:

$$\text{odds ratio } (\psi) = \frac{40 \times 33}{15 \times 20} = 4.4$$

This is greater than unity, so Cortni can say that those workers exposed to benzene have four times the odds of getting AML as compared to those who were not exposed to the chemical.

As with cohort investigations, this formula must be modified if matching has occurred. Once again, the analysis unit is the matched pair, and the odds ratio becomes Cell *b* divided by Cell *c*, which represent discordant pairs. Table 10.10 illustrates a 2 × 2 table for the odds ratio in a matched case-control (retrospective) study.

In Table 10.10, those in Cells *a* and *d* are concordant pairs and do not contribute to the results. McNemar's test is generally used to test significance (Fleiss, 1981).

Confidence Intervals

Statistical inference can be in the form of hypothesis testing or estimation of parameters. While hypothesis testing is more pronounced in the literature, parameter estimation can play a major role. The two forms of estimation are point estimation and interval estimation. The former is single statistic (or point) to estimate the population parameter. For example, if you calculated the mean score of the national examination of the Council on Resident Education in Obstetrics and Gynecology (CREOG) for a sample of 64 ob/gyn residents and found it to be 510, that mean would be the point estimate of the population mean (μ). Similarly, relative risk and odds ratios are point estimates. The problem with all point estimates is that they fail to convey the accuracy of the estimation. Just how accurate is the relative risk of 6.05 per 100,000 person-years in Henry's study or the odds ratio of 4.4 in Cortni's investigation?

A confidence interval (*CI*), or interval estimation, provides a range of values within which the population parameter has a specified probability of falling. That is, a *CI* sets a

Table 10.10 A 2 × 2 Table for Odds Ratio in a Matched Case-Control Study

Cases	Controls	
	Exposed	Not Exposed
Exposed	<i>a</i>	<i>b</i>
Not exposed	<i>c</i>	<i>d</i>

where

Cell *a* = cases exposed + controls exposed
 Cell *b* = cases exposed + controls not exposed
 Cell *c* = cases not exposed + controls exposed
 Cell *d* = cases not exposed + controls not exposed

range or interval of numbers in which we have a designated degree (usually 95% or 99%) of assurance that the population parameter lies within the upper and lower limits of the *CI*. The degree of probability is established by the researcher.

Using our CREOG example, we can calculate the *CI* for our sample mean of 510. The calculations involve both the standard error of the mean (SE_M) and the principles of the normal curve or distribution. Suppose that our sample of 64 ob/gyn residents had a standard deviation of 80. This being the case, the SE_M of our sample would be:

$$SE_M = \text{standard deviation} = \frac{80}{\sqrt{64}} = 10$$

Given this information, you can determine the 95% *CI* as follows:

$$\begin{aligned} 95\% \text{ CI} &= (510 \pm (1.96 \times 10)) \\ &= (510 \pm (19.6)) \\ &= (490.4 \leq \mu \leq 529.6) \end{aligned}$$

Given this result, you can say that you are 95% confident that the population mean is greater than or equal to 490.4, but equal to or less than 529.6. In other words, if the study were repeated 100 times, with a *CI* for each, you can be assured that 95% of the *CI*s calculated would contain the population mean (or whatever population parameter was being presented).

There are several methods for determining the *CI* for an odds ratio, and they are beyond the scope of this text. (Schlesselman, 1982, explains six techniques.) However, the same principles apply as in our illustration with the sample mean, and it is important to understand them in order to interpret epidemiological results. If we have a 95% *CI* for an odds ratio, we can be assured that there is a 95% chance that the true exposure-disease parameter (true measure of association) is contained within the confidence limits (upper and lower ends of the confidence interval). A general rule of thumb when reading the literature is that "the wider the confidence interval, the larger is the variability of the point estimate and the less likely that the point estimate is accurate" (Peterson & Kleinbaum, 1991, p. 715).

In reading epidemiologic studies, you will find that the *CI* is used to ascertain statistical significance. This is a two-tailed significance test allowing the epidemiologist to recognize that the exposure either increases or decreases the risk of disease. The investigator establishes the significance level; for example, $\alpha = .05$. The null hypothesis is rejected when the $100(1-\alpha)\%$ *CI* does not overlap the null value being tested. Remember that for RR and odds ratio, the null value equals 1. Consequently, if a 95% *CI* for the RR or odds ratio fails to overlap 1.0, we reject the null hypothesis that there is no association between the risk factor and the health outcome. It is rejected at the .05 significance level. In this manner the confidence level is used in place of the *p* value to test the hypothesis.

Suppose that in Cortni's study, with an odds ratio of 4.4, the 95% *CI* is 1.8–6.2. This does not overlap with 1.0; therefore, she would reject at the .05 significance level the null hypothesis that there is no relationship between benzene and AML. She would

arrive at the same result if $p \leq .05$. In contrast, if her 95% *CI* were 0.75–6.2 she could not reject the null hypothesis because it overlaps 1.0. She would have the same conclusion if $p > .05$.

Computer programs are available to establish RR, odds ratio, and *CI*s. While it would be good to have that knowledge, as a minimum you should be able to interpret the results of analytic epidemiological investigations, including *CI*s and levels of significance.

Case Discussion of Case Study A

In this study, Henry followed the adult children who placed a parent suffering from Alzheimer's disease or related dementia into an assisted living facility for a period of 5 years—a current cohort investigation. Cohort selection was through accessibility. He determined the risk factors to be amount of time spent caring/visiting, the economic costs, coping skills, depression, nature and type of support for them, and physical health status, in addition to information about the patient to include functional physical and cognitive status, severity of the dementia, and length of time since diagnosis. Henry's null hypothesis would be there is no association between these risk factors and the well-being of the adult children. Since the adult children who placed a parent in this facility come from the same geographic area, it is likely that they will be homogeneous in several respects. This helps control for several variables. It is likely that he will use multiple regression for analysis since it deals with continuous variables. The results, as discussed in the chapter, should be based on person-years because he will have subjects exiting and entering at different points over the 5-year investigation. To solidify his relative rate finding of 6.05 per 100,000 person-years, Henry should establish a *CI* and use it to test for significance.

Case Discussion of Case Study B

In her retrospective study, Cortni wanted to review the relationship between selected risk factors and AML for workers at a particular factory during a specified time frame. Based on the factory experience for risk factors, her null hypothesis could be: There is no relationship between exposure to the chemical benzene and AML.

Her choice of cases came from physician records and patients who each had suffered AML and had worked at the factory during the same time as the original patient. In so far as her controls, Cortni selected factory workers who were there during the same time frame and matched on several characteristics. She would look at the relationship using several risk factors such as smoking, gender, radiation exposure, prior cancer episodes, other prior blood disorders, and inherited syndromes in addition to exposure to chemicals. Cortni could use logistic regression for analysis. As in the cohort study, she should determine the *CI*'s surrounding her odds ratio of 4.4 and check for significance at the .05 level.

SUMMARY

The two principal directions for epidemiologic studies are descriptive and analytic. The focus of this chapter was analytical epidemiologic investigations. One of type of analytic methodology is the prospective or cohort study, while the other is the retrospective or case-control investigation. The task for both is to test the hypothesized cause-and-effect relationship between a suspected risk factor and a health condition. The principal difference between the two approaches is the commencement or timing of the study (going forward or backward in time). The advantages and disadvantages of each are outlined in detail.

The cohort in a prospective study is simply a group of people with a common experience over a defined period of time. Some of the ways to select a cohort include accessibility, history of exposure, and availability of needed records. The cohort is divided according to their exposure to the risk factor (exposed versus not exposed, or gradients of exposure). At the end of the study, the two groups are compared to measure the strength of association between the risk factor(s) and the health outcome. In comparing the two groups, it may be more accurate to use person-years rather than population units in the data. Several methods of analysis are available, each dependent upon the study design and nature of the data. The cohort model can be modified by making it retrospective in nature rather than prospective. In this scenario, called a retrospective cohort study, the cohort assembly, baseline data, and follow-up took place in the past. Both are prospective in nature; the principal difference is the starting point.

In regard to the case-control technique, the researcher must establish criteria so that no ambiguity exists between cases and controls. The cases must clearly have the condition or disease, while the controls must be disease- or condition-free. Generally, it is best to select the case group from newly diagnosed cases that meet the criteria and took place within the specified time frame. Sources can include physician and employer records, hospital records, death certificates, birth certificates, and the like. As in the cohort study, once the data are collected, the cases and controls are compared to determine the strength of the association between exposure and health condition. The odds ratio gives us that answer.

While analytic epidemiologic studies do not show causation, they do approximate it. In addition to the rates (relative risk, odds ratio, attributable risk) that can be developed, the researcher can look at temporal sequence, consistency, specificity of effect, biological gradients, and existing data or theory.

Like any research effort, analytic epidemiology can have problems of error. The major ones are bias, random variation, and random misclassification. Bias refers to any error in design, conduct, or analysis of a study. Subsequently, bias includes selection bias, information bias, and confounding. All sources of error need to be controlled as much as possible. Some ways to do this include matching, homogeneous grouping, stratified sampling, poststratification, and multivariate analysis. Logistic regression is quite common because it avoids the problems of matching.

Confidence intervals can be established for relative risk and odds ratios. A confidence interval sets a range or interval of numbers in which we have a designated degree (usually 95% or 99%) of assurance that the population parameter lies within the upper and lower limits of the confidence interval. The confidence interval is important not

only to ensure confidence that the rate represents the true association, but to test the hypothesis at a preset level of significance.

CRITICAL THINKING QUESTIONS

1. What are the principal differences and similarities between epidemiologic descriptive and analytic investigations?
2. What are the major differences between cohort studies and case-control studies?
3. How would you explain current cohort study contrasted with a retrospective cohort study and a case-control study?
4. Why are case-control studies usually less expensive and generally require a smaller sample size than cohort investigations?
5. What methods are available to control for bias in analytic epidemiologic studies?

SUGGESTED ACTIVITIES

1. You work in an obstetrical clinic and want to investigate the association of small fetal size on ultrasound at 10–19 weeks gestation with maternal tobacco use and subsequent early preterm birth and low birthweight. Write out the hypothesis for this retrospective study, and design how you would conduct it.
2. As a counselor at an inner-city teen center, you have watched the rate of teenage suicide escalate slightly. Your director wishes to examine this phenomenon using the large number of teens who participate in her community program.
 - a. List the variables that should be in the study.
 - b. Write out the hypothesis.
 - c. Design the study (cohort, case-control?).
 - d. Write out the way you would analyze the data with specific emphasis on the epidemiologic component (prevalence, incidence, odd ratio, person-years).
3. One of your responsibilities as a health educator is to encourage men to get a colonoscopy. You wonder if men who personally know of someone who died from colon cancer might be more likely to adhere to their physician's recommendation for a colonoscopy. The data show the findings below:

	Colonoscopy	
	No	Yes
Know of someone	150	50
Not know of someone	30	175

- a. Calculate the odds ratio.
- b. Are the odds of adhering to the recommendation for a colonoscopy greater or less if they knew of someone who died from colon cancer?
- c. To what might you attribute these results?
- d. What do these results imply regarding an educational program?

REFERENCES

- Ary, D. (2006). *Introduction to research in education*. Belmont, CA, Thomson/Wadsworth.
- Chen, X. K., Wen, S. W., Fleming, N., Yang, Q., & Walker, M. C. (2008). Increased risk of neonatal and postnatal mortality associated with teenage pregnancy had different explanations. *Journal of Clinical Epidemiology*, 61, 688–694.
- Chok, B., & Gomez, E. (2000). The reliability and application of the neck disability index in physiotherapy. *Physiotherapy-Singapore*, 3(1), 16–19.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley & Sons.
- Fleiss, J. L., Levin, B. A., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, NJ: Wiley.
- Herbst, A. L., Ulfelder, H., & Poskanzer, D. C. (1971). Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *New England Journal of Medicine*, 284(15), 878–881.
- Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2007). *Designing clinical research* (3rd ed.). Philadelphia: Lippincott, Williams & Wilkins.
- Hurwitz, E. S., Barrett, M. J., Bregman, D., Guinn, W. J., Pinsky, P. A., Schonberger, L. B., et al. (1987). Public Health Service study of Reye's syndrome and medications. Report of the main study. *Journal of American Medical Association*, 257(14), 1905–1911.
- Jamieson, D. J., Costello, C., Trussell, J., Hillis, S. D., Marchbanks, P. A., Peterson, H. B., et al. (2004). The risk of pregnancy after vasectomy. *Obstetrics and Gynecology*, 103(5 Pt 1), 848–850.
- Kauffman, R. P., Baker, T. F., Baker, V. M., DiMarino, P., & Castracane, V. D. (2008). Endocrine and metabolic differences among phenotypic expressions of polycystic ovary syndrome according to the 2003 Rotterdam consensus criteria. *American Journal of Obstetrics and Gynecology*, 198(6), 670–673.
- Landmark, K., & Abdelnoor, M. (2000). Current smokers develop more posterior myocardial infarctions probably due to increased tendency to thrombosis. *Scandinavian Cardiovascular Journal*, 34(1), 73–78.
- Matthews, D. E., & Farewell, V. (2007). *Using and understanding medical statistics* (4th ed.). New York: Karger Basel.
- Newman, T. B., Browner, W. S., Cummings, S. R., & Hulley, S. B. (2007). Designing cross-sectional and case-control studies. In S. B. Hulley, S. R. Cummings, W. S. H. Browner, D. G. Grady, & T. B. Newman, *Designing clinical research*. Philadelphia: Lippincott, Williams & Wilkins.
- Nicolai, L. J., Ogden, L. G., Muehlenbein, C. E., Dziura, J. D., Vazquez, M., & Shapiro, E. D. (2007). Methodological issues in design and analysis of a matched case-control study of a vaccine's effectiveness. *Journal of Clinical Epidemiology*, 60, 1127–1131.
- Peterson, H. B., & Kleinbaum, D. G. (1991). Interpreting the literature in obstetrics and gynecology: I. Key concepts in epidemiology and biostatistics. *Obstetrics & Gynecology*, 78, 710–717.

- Rockett, I. R. H. (1994). Population and health: An introduction to epidemiology: *Population Bulletin*, 49(3).
- Rosner, B. (2006). *Fundamentals of biostatistics* (6th ed.). Belmont, CA: Thomson-Brooks/Cole.
- Schlesselman, J. J. (1982). *Case-control Studies: Design, conduct, analysis*. New York: Oxford University Press.
- Simon, S. R., Peters, A. S., Christiansen, C. L., & Fletcher, R. H. (2000). The effect of medical student teaching on patient satisfaction in a managed care setting. *Journal of General Internal Medicine*, 15(7), 457–461.

Analyzing and Interpreting Data: Descriptive Analysis

KEY TERMS

central tendency	nominal measurement	population
descriptive statistical analysis	nonparametric data	ratio measurement
inferential statistical analysis	ordinal measurement	sample
interval measurement	parameter	statistic
linear correlation	parametric data	

Case Study

After earning his doctorate in health education, Armando decided to apply his educational skills to medical education. Beginning as a faculty member in the Department of Medicine, he eventually found himself being appointed as the Director of Education for the Internal Medicine residency program at a community-based hospital where he worked. The program accepted medical students who had completed their M.D. degree at an accredited medical school. They would enter this residency program; and, when they completed it 3 years later, they would have a specialty in internal medicine. The 3-year program had a total of 15 medical residents (5 at each of the 3 years). The department chair asked Armando to provide the status of the program to include such things as: how well the incoming students did on their board scores; how the current medical residents did on their annual nationwide exam; and how they scored on such things as clinical skills, professionalism, medical knowledge, patient satisfaction, interaction with other health care professionals, and scholarly activity. Armando knew he had to collect and analyze a lot of data to create the requested report. He wondered how he should describe the program using the available data.

The Meaning of Statistics

Statistics is a language that can be employed to express concepts and relationships that cannot be communicated in any other way. To the neophyte health researcher, it is a language to be in awe of, or to fear; in contrast, the seasoned health scientist who understands the logic of statistics and appreciates its expressive power views it as a language to organize, analyze, and interpret numerical data.

Another approach to looking at statistics is in terms of its functional aspects. On the one hand, it can function to describe data: to explain how the data look, what the center point of the data is, how spread out the data may be, and how one aspect of the data may be related to one or more other aspects. For example, Armando wants to describe the average score on the national examination, the range of scores, the relationship between their medical knowledge score and their clinical skill score, and the relationship between their clinical score and patient satisfaction score. His primary concern, then, is to describe this group of medical residents. Subsequently, this branch of statistical analysis is called **descriptive statistical analysis**. No conclusions can be extended beyond this immediate group, and any similarity to those outside the group cannot be assumed.

A different function of statistics is to infer. **Inferential statistical analysis** involves observation of a sample taken from a given population. Conclusions about the population are inferred from the information obtained about the sample. Suppose you were to observe the health habits of a random sample of poor inhabitants from the Appalachian region. You could then make inferences about the health habits of the total population of poor persons in the Appalachia. Unlike descriptive data analysis, generalizations can be made from the sample to the respective population.

Inferential data analysis can be used for estimation and prediction. For example, scores obtained on the Graduate Record Examination may predict how well an incoming candidate may perform in a graduate health science program. Extrapolation is a component of inferential statistics but not of descriptive statistics.

Statistical Analysis and Data

Levels of Measurement

While there are several ways to classify variables, one of the principal techniques is the preciseness of measurement of the variable. We will discuss four levels of measurement.

The first measure is called **nominal measurement**. At this level, variables are simply placed into different categories. For example, “gender” is nominal data since 1 or 0 can be assigned to male and female, respectively.

The next higher level measures both groups and ranks the data through ordering of categories. This is called **ordinal measurement** or data. Examples of ordinal measurements in the health sciences are abundant—dosage levels, degree of education, severity of illness, and even social class. The limitation to some categories may be the amount of information available to do the ordering. For example, how much difference is there between lower middle class and upper middle class? In short, the rankings may be relative rather than absolute.

The next level of measurement categorizes, orders, and provides a meaningful measure of the differences in the ordering. That is, **interval measurement** variables can be separated by how much they differ from one another. These variables are not nebulous, unlike socioeconomic status. Examples of interval measurements include height, weight, blood pressure, and the like; their intervals are equal. Celsius temperatures are another example, with the temperature difference between 5 and 15 degrees Centigrade (10 degrees Celsius) being the same as that between 20 and 30 degrees Centigrade.

When interval data have a true zero point, they may be called **ratio measurements**. Height is an example of a ratio variable, as is measure of temperature in Kelvin. As a point of note, Celsius temperature is an interval measurement because the zero is an arbitrary point.

Parametric and Nonparametric Data

Two types of data are recognized in the application of statistics: parametric and nonparametric. **Parametric data** are either interval or ratio data. Parametrical statistical tests assume that the data are normally or near normally distributed. In comparison, data that are either counted (nominal scale) or ranked (ordinal scale) are called **nonparametric data**. Nonparametric statistical tests, often referred to as distribution-free tests, do not require the more restrictive assumption of a normally distributed population.

Frequently parametric tests are considered the more powerful of the two. However, this is only the case when all the assumptions and restrictions of parametric data have been met. If there is lack of homogeneity of variances, unequal sample sizes (*n*'s), and oppositely skewed distributions, using the *t*-test is not as powerful as converting the data to ordinal scale and applying nonparametric tests. Generally, nonparametric tests have wider applications and are less difficult to compute.

Population and Sample

A **population** can be defined as the set of elements we are planning to study. In the health sciences it usually refers to a group of people—all the patients in the hospital, all those living in the county, all graduate students in the health sciences, and so on. However, the population can be something other than a group of people, such as all daily maximum temperatures or all automobiles produced in a given time frame. A **sample** is a subset of the population.

Parameters and Statistics

A **parameter** is a characteristic of a population, whereas a **statistic** is a characteristic of a sample. Although different texts often employ different symbols, some of the more common ones are:

	Sample Statistic	Population Parameter
Mean	M, \bar{X}	μ
Standard deviation	s, sd	σ
Variance	s^2	σ^2

Descriptive Data Analysis Techniques

There are several statistical techniques available to the health science researcher who wishes to describe the observed research group:

- Measures of central tendency
 - Mean
 - Median
 - Mode
 - Geometric mean
- Measures of spread or variation
 - Range
 - Standard deviation
 - Variance
 - Coefficient of variation
- Standard measures
- Measures of relationship
 - Spearman rank order correlation
 - Pearson product-moment correlation

It is the responsibility of the health science researcher to select the technique that best fits the data and that will explain the data in a manner comprehensible to the intended audience.

Measures of Central Tendency

Armando, in our medical residency example, was asked to describe the group using average board scores for incoming residents, average scores on the national examination, average number of patient encounters, and average number of patients and conditions seen in various clinical settings. This is a typical request because most people wish to find a value about which the observations tend to cluster. The three most common measures used to describe the centering point of a set of data are the mean, median, and mode. Collectively, they are called measures of **central tendency**.

Mean. The mean (M or \bar{X}) of a set of data is commonly referred to as the arithmetic mean or average. It is computed by summing all the observations in the group and dividing by the number of observations. The formula is:

$$M = \frac{\Sigma X}{N}$$

- where M = mean
 X = scores in a distribution
 Σ = sum of
 N = number of scores

Table 11.1 Scores and Arithmetic Mean on National Examination

Resident	X		
1	90		
2	95		
3	87		
4	93		
5	96		
6	89	$M = \frac{1319}{15} = 87.93$	
7	90		
8	75		
9	82		
10	86		
11	85		
12	97		
13	78		
14	84		
15	92		
$N = 15$	$\Sigma X = 1319$		

For example, Armando recorded the scores of the medical residents and computed the average score. His data is displayed in Table 11.1.

The arithmetic mean, which may be considered the fulcrum or balance point of a distribution, is one of the most useful statistical measures because it provides much information; is affected by all the scores; and serves as a basis for the computation of other important measures, such as variability.

Median. The median is a measure of position in that it is the point above and below which one-half of the scores fall. In other words, it is the middle-most position.

If Armando examined the scores from the medical residents on their national examination, he would find that the median or middle score is 89, as shown in Table 11.2. There are seven scores above and seven scores below the score of 89.

If there was an even number of medical residents, the median would be the middle point between the two middle scores. For example, if there were only four residents with the following scores:

Resident	X
3	87
1	90
4	93
2	95

Table 11.2 Scores and Median on the National Examination

Resident	X
8	75
13	78
9	82
14	84
11	85
10	86
3	87
6	89
1	90
7	90
15	92
4	93
2	95
5	96
12	97

Median = 89

The median is the midpoint between the scores 90 and 93, making it 91.5. Using 91.5 as the median, there are two scores above and two scores below.

Unlike the mean, the median is not influenced by extreme scores. Therefore, in some instances it may be a more realistic measure of central tendency than the mean. However, the median is usually reserved for when a quick measure of central tendency is needed or when distributions are markedly skewed.

Mode. The mode is simply the most frequently occurring score. By examining the scores of the 15 residents in Table 11.1, it can be seen that the most frequently occurring score is 90. The mode is the quickest estimate of central value and shows the most typical case.

Geometric Mean. The geometric mean is often used in laboratory data. This is especially true with data in the form of concentrations of substances. An example might be the concentration of penicillin in urine following treatment for *Neisseria gonorrhoeae*. The geometric mean is calculated using the antilogarithm of $\log x$. Rosner (2006) provides an excellent illustration for infectious disease.

Measures of Spread or Variation

The measures of central tendency are useful, but oftentimes more information is needed for an accurate description of the sample or population. This is particularly true in a comparison of two groups whose means are identical. In such situations, it is important to know whether the scores or observations for each group tend to be quite similar (homogeneous) or spread apart (heterogeneous). Measures of variation, to include range, variance, and standard deviation, may be employed to show the degree of spread or variation among scores. As an

Table 11.3 Patient Encounters by Residents in the Intensive Care Unit (ICU) and Ambulatory Care (Amb) Unit

ICU Residents	ICU Encounters	Amb Residents	Amb Encounters
Sam	15	Amy	1
Lizzi	2	Rachael	5
Carol	2	Mark	5
Emily	2	Will	6
Sara	3	Katie	4
Bill	1	Alison	1
Eddie	3	Rick	5
		Penny	5
	$\Sigma X = 28$		$\Sigma X = 32$
	Mean = 4		Mean = 4

example, Armando compared the number of patient encounters for residents on the intensive care (ICU) unit, which had patients with severe illness, to those on the ambulatory care (Amb) unit. Table 11.3 illustrates his findings. A comparison of experience on the two units shows that each group of residents had an identical average number of encounters with four for each group. Obviously this description does not offer much if it was placed in his report. However, if he examines the measures of spread or variation within each group, differences are apparent. This illustrates the need for measures of variation or spread.

Range. The range is the simplest measure of variation. It is the difference between the highest and lowest scores. For example, from Table 11.3, the ICU group has a range of 14 (15–1), while the Amb group has a range of 5 (6–1). The range may be used justifiably as a hasty measure of variability; but, since it takes into account only the extremes and not the bulk of observations, it is not generally employed.

Standard Deviation and Variance. The most useful measures of variation are standard deviation and variance. Deviation is defined as the distance of the measurements away from the mean. In a study dealing with a sample, the variance is the sum of the squared deviations from the mean, divided by $N-1$. Although it will be discussed later, it is important to understand that the variance is obtained from squared deviations from the mean, thereby making the variance a different unit of measurement than the mean. The formula is

$$s^2 = \frac{\Sigma(X - M)^2}{N - 1}$$

or

$$s^2 = \frac{\Sigma x^2}{N - 1}$$

where $x = (X - M)$

Table 11.4 Variance of the ICU Group Patient Encounters

X	x or $(X - M)$	x^2
15	11	121
2	-2	4
2	-2	4
2	-2	4
3	-1	1
1	-3	9
3	-1	1
$\Sigma X = 28$	$\Sigma X = 0$	$\Sigma x^2 = 144$

Using the information in Table 11.3, the variance for the ICU group is calculated as:

$$s^2 = \frac{144}{6} = 24$$

The complete calculations are in Table 11.4. Using the same formula, the variance for the Amb unit group, as shown in Table 11.5, is 3.71.

Armando would find that the number of patient encounters in the ICU group varied much more than those in the Amb group. That is, the data for the ICU group are heterogeneous, while the data for the Amb group are homogeneous.

The standard deviation is computed by obtaining the square root of the variance. Notice that this takes away the squaring of deviations, thereby making the

Table 11.5 Variance of the Amb Group Patient Encounters

X	x or $(X - M)$	x^2
1	-3	9
5	1	1
5	1	1
6	2	4
4	0	0
1	-3	9
5	1	1
5	1	1
$\Sigma X = 32$	$\Sigma x = 0$	$\Sigma x^2 = 26$

standard deviation the same unit of measurement as the mean. By formula, standard deviation is:

$$s = \sqrt{\frac{\sum x^2}{N - 1}}$$

Therefore, the standard deviation for the ICU and Amb groups, respectively are:

$$\text{Group 1 } s = \sqrt{\frac{\sum x^2}{N - 1}} = \sqrt{\frac{144}{6}} = \sqrt{24} = 4.90$$

$$\text{Group 2 } s = \sqrt{\frac{\sum x^2}{N - 1}} = \sqrt{\frac{26}{7}} = \sqrt{3.71} = 1.93$$

As with the variance data, comparison of the standard deviations reveals that the spread of score is much greater in Group 1 (ICU) than in Group 2 (Amb). As stated previously, the variance is the sum of squared deviations, whereas the standard deviation is not squared, thereby leaving it in the same measurement units as the arithmetic mean. This holds true whether the measurement is in centimeters, units of blood, people, or whatever. This is one of the principal reasons why the standard deviation is reported more often than the variance as a measure of spread.

When the health science researcher is working with data from an entire population, there is a slight change in the formula along with different symbols. The population mean, μ (mu), is the sum of all values divided by N (the total number in the entire population). By comparison, the sample mean, M or \bar{x} , is an estimation of the population mean. Similarly, the sample variance (s^2) is an estimate of the population variance (σ^2 , or sigma squared), which is determined by the formula:

$$\sigma^2 = \frac{\sum x^2}{N}$$

The standard deviation of the population is ascertained by:

$$\sigma = \sqrt{\frac{\sum x^2}{N}}$$

The difference between formulas and samples is the use of $N - 1$ rather than N for division in the sample formula. The reason for employing $N - 1$ in the sample formula is to provide an equation that gives an unbiased sample variance. In other words, $N - 1$ allows for a more accurate estimate of the population variance and standard deviation.

Coefficient of Variation. As noted previously, the arithmetic mean and standard deviation are usually reported together to describe data. This is important because, for example, a standard deviation of 10 would mean something different if the arithmetic mean were 20 rather than 500. Also, the standard deviation and the mean are in the same units. However, there are occasions when the health science researcher wishes to

compare the variability of different samples, each having different arithmetic means and perhaps differing units of measurement.

Suppose a researcher wanted to compare birth weights of infants of pregnant teens based upon their site for prenatal care and delivery. However, one site recorded the birth weights in grams and the other in ounces. Because the standard deviation for each group is based on the unit of measurement, this would be an inaccurate comparison. To overcome this problem, the researcher would use the coefficient of variation (CV). It is calculated as:

$$CV = 100\% \times \frac{s}{M}$$

If the birthweights at the site using grams as the measurement unit had a mean and standard deviation of 3121.8 grams and 432.6 grams, respectively, the CV would be:

$$CV = 100\% \times \frac{432.6}{3121.8} = 13.9\%$$

The site using ounces as the unit of measurement had a mean of 89.9 ounces and a standard deviation of 12.2 ounces. The calculation is:

$$CV = 100\% \times \frac{12.2}{89.9} = 13.6\%$$

The comparison shows little difference in the variability of birth weights between the two sites.

Standard Measures

Standard scores are equal units of measurement and are very useful in reporting test scores or doing research involving test results. For example, if a subject scored a 75 on a health science test and 85 on a statistics test, which one is really better? Obviously the raw data are of little use, since the tests may differ in the number of items and scaling. For comparison, the scores must be transformed to a common equal-interval scale in which they are called *standard scores*. The standard score, also called the *z-score*, is calculated by subtracting the mean from the observed score and dividing the result by the standard deviation:

$$z = \frac{X - M}{s}$$

where M = mean

X = observed score of the observed distribution

s = standard deviation of the observed distribution

Suppose the health science test had a mean (M) of 70 with a standard deviation (s) of 3, whereas the statistics test had a mean (M) of 88 with a standard deviation (s) of 6. Using the formula, the z -score for the health science test would be +1.67 and for the statistics test -0.5 . The comparison shows the health science test score is much better than the statistics test score. Note that z -scores can be negative as well as positive.

Since z -scores can be expressed in negatives and decimals, oftentimes they are converted so as not to be cumbersome. For example, the subtests of the SAT scores are simply converted z -scores. The mean is set at 500, and the standard deviation is 100. To transform the score, simply multiply the z -score by the standard deviation and then add the mean. Conversion of the health science score to this system is as follows:

$$\text{Transformed score} = z(100) + 500 = 1.67(100) + 500 = 667$$

Intelligence test scores (IQ) have means of 100 with a standard deviation of 15.

Standard scores rather than raw scores can be used by teachers in summing test results for the semester. These give each measure or test the same weight. If the distribution is normal in shape, the standard score can be converted to a percentile rank from a table in most statistics books.

Measures of Relationship

Some of the most interesting questions posed in health science research revolve around the relationship of one variable to another, such as smoking and lung cancer. The method used most frequently to describe the relationship between two or more variables or between two or more sets of data is **linear correlation**. The degree of relationship is expressed by the coefficient of correlation, symbolized by r .

Some of the unique characteristics of the correlation coefficient are that it is a pure number, it is nondimensional, and it may take on values between -1.00 and $+1.00$. A correlation coefficient of zero indicates no relationship between the variables in question. The closer the r is to 1.00 (negative or positive), the stronger the relationship. A perfect positive correlation of $+1.00$ specifies that for every unit increase in one variable, there is a proportional unit increase in the other variable. Concomitantly, a perfect negative correlation of -1.00 means that for every unit increase in one variable, there is a unit decrease in the other variable. Perfect correlations are highly unlikely in dealing with human health concerns.

The scatterplot is a means of displaying the relationship between variables and is developed by graphically plotting each pair of variables that correspond to the x and y axis, respectively. The line drawn through or near the coordinate points is referred to as the line of best fit or regression line. Figure 11.1 demonstrates several correlations and their regression lines.

The beginning health science researcher must be careful not to fall into the trap of attributing a cause-and-effect relationship to variables that might be related. For instance, Kuzma and Bohnenblust (2005) reports a strong relationship between a child's foot size and handwriting ability. As explained, however, this is likely because both increase with age rather than being a direct cause-and-effect relationship. Spurious relationships must be viewed with caution and interpreted judiciously.

Spearman Rank Order Correlation. The Spearman rank order correlation is used to determine the relationship between two ranked variables (rather than interval or ratio variables). That is, the Spearman rank order correlation, signified by r_s , is designed for nonparametric data. Frequently, a health science investigator may employ it to

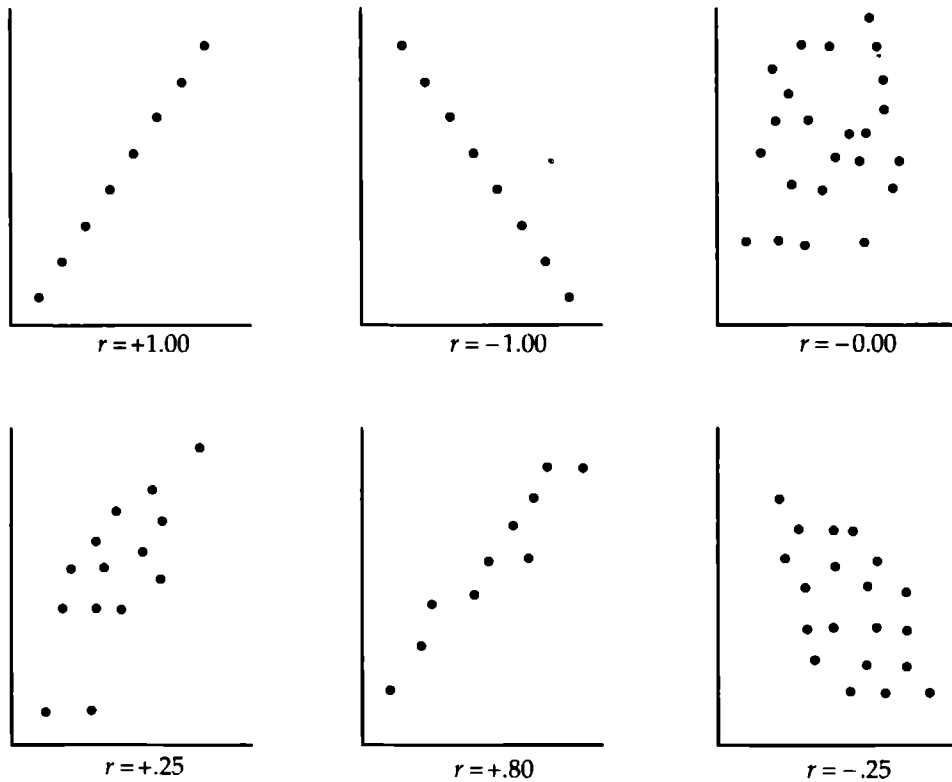


Figure 11.1 Correlation Scatterplots

compare judgments by a group of judges on two objects or the scores of a group of subjects on two measures (for further information on this usage, consult Siegel and Castellan, 1988). A less frequent but valuable usage is to compare judgments by two judges on a group of objects or items. Hays (1994) discusses the Spearman rank order correlation for assessing interjudge equivalence. In cases of multiple judges and multiple objects, the Friedman two-way analysis of variance or Kendall coefficient of concordance would be used.

The equation for Spearman rank order correlation coefficient is:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

where n = number of pairs

d = difference between paired ranks

$\sum d^2$ = sum of the squared differences between ranks

If Armando, in our case study, were to have two independent physicians rank 10 ICU patients according to the severity of their medical complications, he would use the

Table 11.6 Ranking of Severity of Medical Complications by Two Independent Physicians

Patient Number	Physician 1's Rank Order (X)	Physician 2's Rank Order (Y)	d ($x - y$)	d^2 ($x - y$) ²
1	1	2	1	1.00
2	4	1	3	9.00
3	7.5	9.5	-2.5	6.25
4	2	3	-1	1.00
5	7.5	8	-0.5	0.25
6	3	4	-1	1.00
7	5	9.5	-4.5	20.25
8	6	6	0	0.00
9	9	5	4	16.00
10	10	7	3	9.00
				$\Sigma d^2 = 63.75$

Spearman technique. Table 11.6 displays the data. Note that ties in ranks are handled by averaging the ranks. Using the formula, Armando would obtain:

$$r_s = 1 - \frac{6(63.75)}{10(99)} = .62$$

The Spearman rank order correlation procedure is an acceptable method for parametric data when there are fewer than 30 but greater than 9 paired variables. The ease of computation allows it to be used by health teachers for a single classroom of students.

Pearson Product-Moment Correlation. The Pearson product-moment correlation, the most often used and most precise coefficient of correlation, is used with parametric data. The basic formula, symbolized by r , is:

$$r = \frac{N \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N \Sigma X^2 - (\Sigma X)^2 N \Sigma Y^2 - (\Sigma Y)^2}}$$

This is the raw score equation that is convenient for both calculator and computer use. Statistics books can be explored to obtain equations written in a different format (Kuzma & Bohnenblust, 2005).

Armando wished to investigate the relationship between how a good a teacher (of medical students) a resident might be and how satisfied patients are with their care, especially patient interaction. He believed that there would be a strong relationship in that the best teaching residents would get better scores on patient satisfaction surveys. His database contained scores on their teaching and scores on patient satisfaction, with the highest score possible being 20 for each category. Table 11.7 illustrates the data and necessary calculations for the 15 residents and their chief resident (for a total of 16 participants). The correlation of $r = .69$ shows that Armando is correct in his assumption.

Table 11.7 Teaching Score and Patient Satisfaction Score

Teaching Score X	Patient Satisfaction Score Y	X^2	Y^2	XY
11	11	121	121	121
15	12	225	144	180
18	12	324	144	216
12	11	144	121	132
11	11	121	121	121
16	13	256	169	208
14	13	196	169	182
13	12	169	144	156
19	11	361	121	209
14	10	196	100	140
16	15	256	225	240
17	14	289	196	238
15	13	225	169	195
16	16	256	256	256
15	13	225	169	195
15	14	225	196	210
$\Sigma X = 237$	$\Sigma Y = 201$	$\Sigma X^2 = 3589$	$\Sigma Y^2 = 2565$	$\Sigma XY = 2999$
$r = \frac{16(2999) - (237)(201)}{\sqrt{[16(3589) - (237)^2][16(2565) - (201)^2]}}$ $r = \frac{47,984 - 47,367}{\sqrt{[57,424 - 56,169][41,040 - 40,401]}}$ $r = \frac{617}{\sqrt{[1255][639]}}$ $r = \frac{617}{\sqrt{801,945}}$ $r = \frac{617}{895.51} = .69$				

Personal Computers and Information Delivery Systems

As computer technology has increased, the cost of hardware and, in many instances, of software has decreased. Consequently, most entry-level health science researchers are able to analyze their data on desktop or laptop computers. The power of these computers has grown significantly, allowing very sophisticated computations through a myriad of software programs. In addition to the obvious data analysis, several software packages have

an integrated suite to address issues of data access, management, and presentation. Examples of computer programs that can be employed by the health science researcher include: Statistical Analysis System (SAS), Statistical Package for the Social Sciences (SPSS for Windows), Minitab, Statistica, and S-Plus.

Selection of a statistical analysis package can be an arduous process. Some considerations for selection and possible purchasing are:

1. Know what statistical analyses you need to do now or in the near future. The package you purchase is likely to be updated in the future (for ease of use and so forth), so you can probably plan on getting another package as your needs change. Table 11.8 illustrates some of the procedures that can be found in various programs.

Table 11.8 Typical Statistical Procedures Available in Computer Software

Statistical Component	Typical Usage
Frequencies	Counts, percentages, central tendencies, dispersion, distributions
Descriptives	Central tendencies, geometric mean, variance, standard deviation, standard error of the mean, coefficient of variation, N and $N - 1$ denominators
Confidence intervals	User-specified, such as one group mean, paired mean difference, unpaired mean difference, variance, t -test
t -tests	One group, two group paired; two group unpaired, one sample to compare sample mean to reference mean of choice
Cross-tabulation	Two-way contingency tables, summary statistics, Fisher's exact test, continuity correction, phi coefficient for 2×2 tables, Cramer's V , Mantel-Haenszel's chi-square, Goodman and Kruskal's lambda, tau, gamma
ANOVA and ANCOVA	Balanced and unbalanced designs, several covariates, several independent variables, post-hoc tests such as Duncan's multiple range, Scheffe's test, Student-Newman-Keuls, Tukey's procedure
Correlation	Bivariate to include Pearson's r , Kendall's tau-b, Spearman, cross product deviations and covariances; partial correlation
Regression	Several methods: forced entry; forced removal; backward elimination; forward entry; stepwise procedures; polynomial; plots such as scatterplots, outlier, normal probability, partial, histogram
Nonparametric tests	Chi-square; one sample tests such as Kolmogorov-Smirnov, Poisson; two independent samples Mann-Whitney, Kolmogorov-Smirnov Z ; k -independent samples Kruskal-Wallis; two related samples Wilcoxon, sign, McNemar; k -related samples Kendall's W , Friedman
Survival analysis	Actuarial estimates, linear rank tests, Kaplan-Meier; plots like cumulative survival, hazard
Graphing capabilities	Importing and exporting files; chart production directly from statistical results; graphs; automatic templates; page layout; interface with several printers
Data management	Computerized editing for missing, inconsistent, or out-of-range data; red flags for omissions; use of correct forms; consistency among key variables such as age and birth date

2. Know whether you need a full database system.
3. Think about your knowledge of variables, variable labeling, and variable manipulation, and select a program that meets or slightly exceeds your knowledge level.
4. Display of data is important, and you should consider the nature of the graphic displays you would like to have as well as the look of tables. For example, do you need plotting capabilities? Do the tables generated by the program indicate all the necessary variables and labels?
5. Does the package have a tutorial segment? If so, does it allow you to exit and then reenter where you left off, rather than having to start at the beginning again?
6. Is online help readily available in the package, and is it understandable?
7. Is technical support service available locally, through a help line, on the Internet, or by fax or e-mail?
8. If you plan to buy, know how much money you have to spend, and be careful not to be sold something you neither need nor can afford.

Computer software can take the drudgery out of many research tasks. However, as the world becomes more mobile, access to information in a portable and simple manner is becoming more important. Wireless technologies extend information delivery as well as advanced data analysis. Instead of computers, health science researchers can use light-weight devices such as cell phones, pagers, handheld PCs, and calculators for computing. Accessing data from the Center for Disease Control while on the golf course or at the gym will be simple. Crucial to mobile computing, as well as more standard computers, are packages available on the Internet.

Case Discussion

In order to describe the various aspects of the medical residency program in internal medicine, Armando needed to analyze data so that they could be presented in an understandable fashion. He initially concentrated on national examination scores, looking at measures of central tendency. However, when he compared the number of patient encounters for residents on the intensive care unit with those on the ambulatory care unit, he quickly discovered that the mean or average was of little value in describing the two groups. Therefore, he decided to look at patient encounters from a different vantage point—variation or measures of spread. He saw that this approach provided a better description of patient encounters between the two groups than did the average or mean. Armando also wanted to see how physicians might differ or agree on the severity of medical complications. To carry that out, he used the Spearman rank order correlation. In addition, he wished to report on the relationship between teaching ability and patient satisfaction scores. He conducted a Pearson product-moment correlation to find that a positive relationship did exist between the two variables.

SUMMARY

This chapter discussed descriptive analysis. Statistics, it was seen, is simply a language to express concepts and relationships that cannot be communicated in any other way. As we saw, statistics can be divided into descriptive and inferential. Descriptive statistics describe the population or sample. Inferential statistical analysis involves observation of a sample taken from a population, and conclusions about that population are then inferred from the sample.

Statistical analysis takes several things into consideration. One is the level of measurement, which can be nominal, ordinal, interval, or ratio. Another issue is the type of data—parametric and nonparametric. The researcher must know the differences between parameter and statistic as they apply to a population and sample.

The measures of central tendency examined included the mean, median, mode, and geometric mean for laboratory data. Measures of spread or variation discussed were the range, standard deviation, variance, and coefficient of variation. Relationship measures presented were Spearman rank order correlation and Pearson product-moment correlation.

Computer software is readily available in either abbreviated student packages or packages for professionals. Many of the major companies have packages for basic statistics and other packages for advanced statistics. The Internet has several computer resources and support services, some of which are connected directly with universities, while others connect directly to the software company.

CRITICAL THINKING QUESTIONS

1. If a professor told the class the mean or average score of the last examination and also mentioned that the standard deviation was very large how should the students interpret what is implied?
2. What is the relationship between standard deviation and variance?
3. What are the differences between parametric and nonparametric data, and how do those differences impact the choice of statistical tests?
4. What measure of central tendency would you use to describe a group of people with normal blood pressure as compared to a group experiencing severe hypertension? Justify your answer.
5. When would you use the Spearman correlation technique rather than the Pearson product-moment correlation formula?

SUGGESTED ACTIVITIES

1. A clinician looked at the effect of position on blood pressure levels. Ten participants had their blood pressure taken while lying down with their arms next to them

and then standing up with their arms out at shoulder level. The data are shown below:

Lying Down		Standing	
Systolic	Diastolic	Systolic	Diastolic
100	75	110	82
120	80	128	85
118	70	124	78
122	74	118	75
146	93	139	95
108	72	110	74
160	93	162	98
134	76	128	68
102	68	110	70
104	68	98	62

- a. Compute the mean, median, and standard deviation for the difference in systolic and diastolic blood pressure, respectively, between the positions (lying down and standing).
 - b. Interpret your results as if you were going to put them in a report.
2. Make a checklist of points you believe should be taken into consideration if you were to purchase a statistical software package.
 3. Looking at the list below, label each item as to the level of measurement (nominal, ordinal, interval, or ratio).
 - a. Age
 - b. Blood Pressure
 - c. Grades from an exam
 - d. Ethnicity
 4. Use a calculator, a software package, or Microsoft Excel to calculate the mean, median, mode, variance, and standard deviation for each of the following two groups of scores, which represent student performance on a health science test, with the highest possible score being 50. (Microsoft Excel has several statistical

Group A		Group B	
35	21	20	25
49	28	50	44
33	41	16	29
44	37	21	40
45	39	23	26

tests that can be used for descriptive analysis and often comes bundled with other packages.)

- a. Which group has the greatest variability in scores?
- b. Which group has the highest mean score?
- c. How would you describe each group when comparing it to the other?

REFERENCES

- Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth: Harcourt Brace College Publishers.
- Kuzma, J. W., & Bohnenblust, S. E. (2005). *Basic statistics for the health sciences* (5th ed.). Boston: McGraw-Hill.
- Rosner, B. (2006). *Fundamentals of biostatistics* (6th ed.). Belmont, CA: Thomson-Brooks/Cole.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Analyzing and Interpreting Data: Inferential Analysis

KEY TERMS

alternative hypothesis	interval estimation	point estimation
analysis of covariance (ANCOVA)	level of significance	post hoc test
confidence interval	meta-analysis	region of acceptance
correlation hypothesis test	multiple linear regression	region of rejection
critical region	multivariate analysis of variance (MANOVA)	simple linear regression
directional hypothesis	null hypothesis	test statistic
estimation of parameters	one-tailed test of significance	three-way ANOVA
hypothesis testing	one-way ANOVA	two-tailed test of significance
		two-way ANOVA

Case Study A

Bob, as head of infection control at the community hospital, was worried about the slowly rising increase in health care–associated methicillin-resistant *Staphylococcus aureus* (HA-MRSA). He knew HA-MRSA was a problem in most health care facilities and usually attacked those most vulnerable—older adults and people with weakened immune systems, burns, surgical wounds, or serious underlying health problems. Bob was aware of the risk factors and knew that if the health care professionals would wash their hands more (leaving a patient’s room, before entering a patient’s room, and so on) the incidence could be lowered greatly. He convinced the administration to enhance the policy and run a campaign of hand washing. Bob predicted that within 6 months there would be a noticeable difference in the number of HA-MRSA infections.

Case Study B

As a health educator, Mary was well aware of the fact that of all racial and ethnic groups in the United States, human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) have hit the African Americans the most. The reasons, she knew, were due to some of the barriers faced by African Americans rather than race or ethnicity. These barriers included poverty, sexually transmitted diseases, the stigma of having HIV/AIDS, and prejudice toward people who do things that put them at risk. Her review of data from the Centers for Disease Control (CDC) showed that Blacks, including African Americans, experienced more illness, shorter survival times, and more deaths with HIV/AIDS than did comparable groups. She thought that faith-based programs would help reduce the number of HIV/AIDS cases and therefore designed a randomized study with a control group to educate people about HIV/AIDS, including risk-taking behaviors. She strongly felt the program would make a significant difference in knowledge about HIV/AIDS and attitude toward those with HIV/AIDS.

Inferential Analysis

Statistics is both a collection of numbers, as seen in the chapter on descriptive statistics, and a “process” or a way of working with those numbers to address research questions. Examples of such questions are: *Will a new medicine be more effective than the current one to lower blood pressure? Will computer self-paced instruction be more effective in teaching about disease than the traditional instruction?* For Bob, the question is: *Will increased hand washing decrease the incidence of HA-MRSA?* Mary’s question is: *Will her faith-based program change both knowledge and attitudes about HIV/AIDS in the African American population?*

The researcher gathers data (numerical information), organizes it, and then analyzes it using various statistical tests to make inferences to answer the research questions. An inference is simply an educated statistical guess. However, this “guess” is based on a statistical framework in order to make decisions in a systematic, objective manner. Generally, inferential statistics involves making educated guesses or inferences from samples to a population. When inferring from a smaller group to a large one, it is essential that the smaller group represents the larger group, and that is why random sampling is a critical part of the process. The two principal types of statistical inference are **estimation of parameters** and **hypothesis testing**.

Inferential statistics use procedures that allow researchers to make those inferences about the population based on the descriptive statistics obtained from analyzing the data collected on the sample. The descriptive measures from the sample are the statistics, while the descriptive measures from the population are the parameters. Inferences are thus made about the parameters from the statistics.

Estimation of Parameters

We use estimation procedures to determine a single population parameter when there is no preestablished hypothesis about the value of the population characteristic. For example, Mary may use estimation procedures to estimate the average knowledge score of African American participants who completed the faith-based HIV/AIDS program.

Mary can use what is called a **point estimation** or an **interval estimation**. In point estimation, assume that she calculated a mean knowledge score of 55 (out of a possible score of 100) from a sample of 100 participants. She would state that the mean of 55 is the point estimate of the population mean (μ). In other words, the sample mean (\bar{x}) or (m) is the best point estimate of the true value of μ . Of course, there is the possibility of error so that the actual population mean may be smaller or larger. Consequently, Mary may want to use a range of possible values for μ rather than a single point estimate. This range is called a **confidence interval**.

In addition to establishing a range of possible values when using interval estimation, the researcher establishes a degree of confidence. That is, the researcher can say that the estimation is made with a certain degree of confidence (see Figure 12.1). More often than not, the degree of confidence is set at 95% or 99%. Understanding the normal curve, recall that:

$$68\% = \text{Mean} \pm 1.0 (SD)$$

$$95\% = \text{Mean} \pm 1.96 (SD)$$

$$99\% = \text{Mean} \pm 2.58 (SD)$$

where SD = standard deviation

It is good to know that about 68% of the all scores fall within the limit of plus or minus one standard deviation either side of the mean. Similarly, 95% of the scores or cases fall

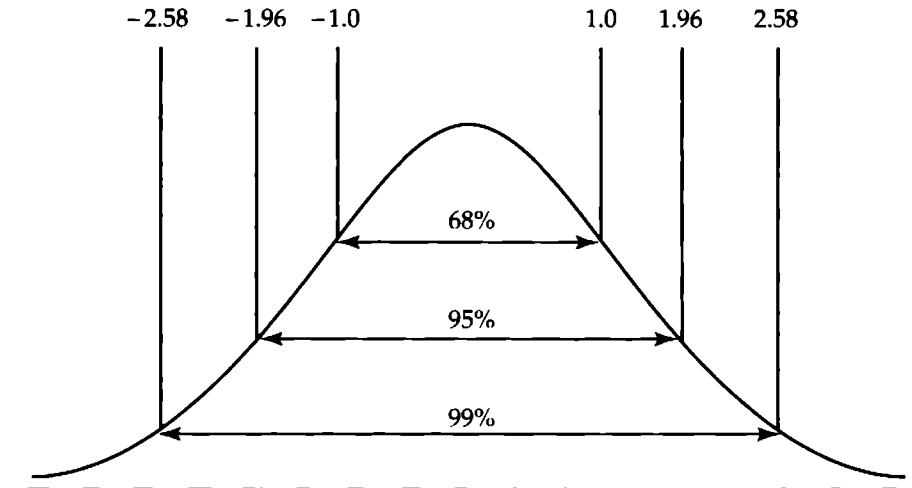


Figure 12.1 Percent of areas under normal curve with 2-scores

within 1.96 standard deviations from the mean, and 99% of all cases or scores fall within 2.58 standard deviations from the mean.

In Case Study B, Mary had a mean knowledge score of 55 for her sample, and let's suppose the standard deviation was 6. If she wished to calculate a confidence interval with 95% confidence, her calculations would be:

$$95\% = 55 \pm 1.96 (6)$$

$$95\% = 55 \pm 11.76$$

$$95\% = 43.24 \leq \mu \leq 66.76$$

She would report with 95% confidence that the true population mean falls between a score of 43.24 and a score of 66.76. Stated differently, if she used 100 samples, 95 out of 100 such samples would contain the population mean. Reversed, Mary would acknowledge that she could be incorrect 5 out of 100 times. The concept of the normal curve and the z-distribution is important in the discussion of hypothesis testing.

Hypotheses Testing

In scientific investigation, the researcher develops a hypothesis or statement about the outcome of the investigation. As pointed out in an earlier chapter, hypotheses can be stated in several ways: **null**, **alternative**, or **directional hypotheses**. Examples using Mary's study would be:

Null: There is no difference in knowledge of HIV/AIDS between those participating in the faith-based program and those not participating.

Alternative: There is a difference in knowledge of HIV/AIDS between those participating in the faith-based program and those not participating.

Directional: Those participating in the faith-based program will have greater knowledge of HIV/AIDS than those not participating.

Hypothesis testing is simply a statistical means to determine if the hypothesis is correct. Usually, the researcher tests the null hypothesis of "no difference." If statistical significance is found, then the investigator rejects the null hypothesis. A directional hypothesis, developed when there is adequate prior information to make such a prediction, can be used and tested if deemed appropriate. It should be realized that obtaining statistical significance is more likely when a directional hypothesis is used. This is discussed in the *Critical Region: Region of Rejection and Region of Acceptance* section of this chapter.

The steps in hypothesis testing are summarized as follows:

Step 1: State the null (H_0) and alternative (H_A) hypotheses

$$H_0: \mu_1 = \mu_2 \text{ or } H_A: \mu_1 \neq \mu_2$$

Step 2: State the level of significance (e.g., .05)

Step 3: Compute the test statistic (such as the z-score)

- Step 4:** Determine the critical region, which is the z distribution with one or two tails
- Step 5:** Reject the null hypothesis if the test statistic z falls in the critical region. Do not reject if it falls in the acceptance region.
- Step 6:** State appropriate conclusion.

Level of Significance

The acceptance or rejection of a hypothesis is based upon a **level of significance** (alpha level, or α), which corresponds to the area in the critical region. Many research efforts in health science establish the level of significance at the 5% (.05) alpha level, although it may be set at the .025, .01, or .001 levels. Rejecting the null hypothesis at the .05 alpha level suggests a 95% probability that the differences between the two variables is real, that is, not the result of chance. In other words, there is a less than 5% probability that the differences are caused by error or chance.

In a two-tailed test of significance, the 5% area of rejection is split between the upper and lower tails of the curve since the null hypothesis is nondirectional. By comparison, in the directional one-tailed test, the 5% area of rejection is at either the upper end or the lower end of the curve. As a general rule of thumb, the following probabilities and interpretations are widely accepted by health science researchers (Kuzma & Bohnenblust, 2001):

Probability Value	Interpretation
>.05	Result is not significant.
<.05	Result is significant.
<.01	Result is highly significant.

Computing the Test Statistic

Now that the null hypothesis has been stated and level of significance established, the researcher should select an appropriate **test statistic**. While this will be covered in more detail later in this chapter, computing the test statistic requires knowing your data well;

Box 12.1 SAYING THE SAME THING . . . BUT DIFFERENTLY

- The finding is significant at the .05 level.
- The alpha level is .05.
- $\alpha = .05$.
- Type I error rate is .05.
- The p value is .05.
- The confidence level is 95%.
- There is 95% certainty this result is not due to error or chance.
- The region of rejection area is .05.

for example, are they parametric or nonparametric? Appendix A lists several test statistics with brief explanations.

Critical Region: Region of Rejection and Region of Acceptance

Following the steps, the researcher looks up the critical z-score, or the **critical region**—the one that corresponds with the level of probability or significance (illustrated in Figure 12.1). Critical z-scores can be found in standard normal tables. Most researchers apply a **two-tailed test of significance**, which means that each “tail” or end of the sample distribution is used (Figure 12.2). A **one-tailed test of significance** should

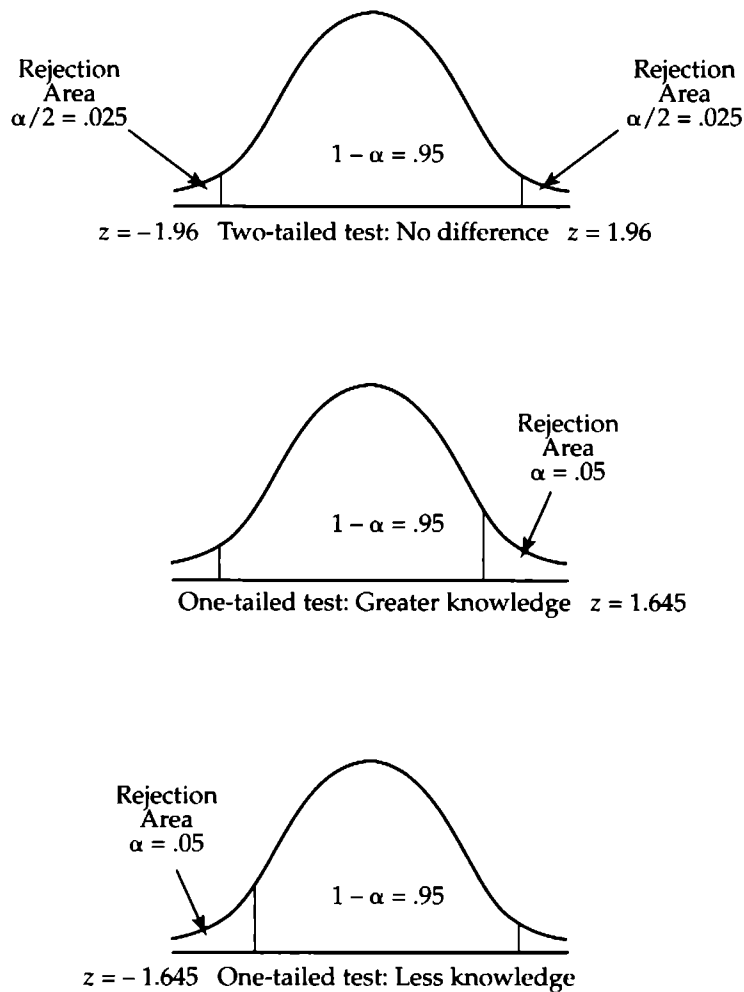


Figure 12.2 Two-Tailed and One-Tailed Regions

be used when there is a directional hypothesis. Examine Figure 12.2 closely with special attention to the differences in critical z-scores for two-tailed versus one-tailed tests even though both have the same probability level of $\alpha = .05$.

If the researcher's test statistic is below the critical z-score (the **region of rejection**), the null hypothesis can be rejected. If it is above the critical value (the **region of acceptance**), the researcher cannot reject the null hypothesis.

Appropriate Conclusion and Type I and Type II Errors

If the researcher rejects the null hypothesis, then it can be concluded that there is evidence to support the alternative hypothesis. However, the researcher could be incorrect in this conclusion. Recall, that if the level of significance is set at 95% there is a 5% chance (1 in 20) that the null hypothesis was rejected when, in fact, it was correct. Type I error is the rejection of a null hypothesis when it in fact is true. The alpha level of significance determines the probability of a Type I error. If the health science researcher rejects a null hypothesis at the .01 level, there is a 1% risk of rejecting it when it is actually true. Similarly, using the .05 alpha level of significance, the researcher is taking a 5% risk of rejecting the null hypothesis even when it is true.

However, if the null hypothesis is accepted when in fact it is false, a Type II error occurs. Therefore, if the health science researcher establishes the alpha level of significance as high as .01, the possibility of a Type I error is reduced, but the chance of a Type II error increases, as shown in Table 12.1.

Table 12.1 Alpha Level and Probability of Type I and II Errors

Alpha	Type I Error	Type II Error
.01	Decreased	Increased
.02	—	—
.05	Increased	Decreased

As an example, suppose that Mary compared two educational techniques to be used—Technique A and Technique B. If she rejected the null hypothesis and claimed that Technique B was the preferred method for teaching, new equipment may be purchased and additional personnel budgeted. Later, in subsequent experimentation or in actual programs, it may be found that Technique B failed to bring about the expected results. Although the ultimate truth about the falsity of the null hypothesis would still be unknown, evidence supporting the null hypothesis would be abundant, leaving Mary humiliated and with a depleted budget. She probably committed a Type I error.

On the other hand, if she accepts the null hypothesis and it is later found that there is a difference between Techniques A and B in that one brings about better results, she again may be embarrassed. Typically, Type I errors lead to

Table 12.2 Type I and Type II Errors

Researcher's Decision	True Conditions	
	True	False
Reject null hypothesis	Type I error	Correct decision
Fail to reject null hypothesis	Correct decision	Type II error

unwarranted changes, whereas Type II errors maintain the status quo when a change should occur. (See Table 12.2.)

Inferential Data Analysis Techniques for Comparing Mean Scores

Prior to choosing a statistical test the health science researcher must determine the criteria for using the test (Berg & Latin, 2008). When comparing two mean scores a *t*-test is used. If three or more mean scores are compared, generally the analysis of variance (ANOVA) is used. The statistical criteria for using these tests to compare means are:

1. Compare the means of two or more groups.
2. The data are drawn from a normally distributed population. This is more important for small samples than for large sample sizes, especially with unequal sample sizes.
3. The variance/standard deviation in each group is identical or at least similar.
4. The data are ratio or interval with continuous data and equal intervals.

The *t*-Test for Two Independent Sample Means

Health science research workers frequently draw two samples from a population and assign them to a control group or to an experimental group. After the experimental group has been exposed to the treatment, the researchers may wish to compare the experimental to the control group. Because the mean is likely the most satisfactory measure for characterizing a group, health scientists find it important to determine if there is a difference between the mean of the experimental group and the mean of the control group. To accomplish this, a *t*-test is used to determine the probability that the difference between the means is a real difference rather than a chance difference.

In such a situation, the null hypothesis is expressed as:

$$H_0: M_1 = M_2 \text{ or } H_0: M_1 - M_2 = 0$$

The alternative hypothesis is:

$$H_A: M_1 > M_2$$

The formula for the t -test is as follows:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

- where M_1 = mean of experimental group
 M_2 = mean of control group
 N_1 = number in experimental group
 N_2 = number in control group
 s_1^2 = variance of experimental group
 s_2^2 = variance of control group

When the samples are greater than 30 subjects, the t critical values are expressed as z scores. Therefore, the obtained t -value from the formula is compared with the z -distribution for acceptance or rejection of the null hypothesis. If the obtained t -value exceeds the z -score of 1.96 (two-tailed test), the researcher may conclude that a significant difference exists between the two means at the .05 level. Concomitantly, an obtained t -value greater than a 2.58 z score (two-tailed test) allows the researcher to reject the null hypothesis at the 0.1 level of significance.

One of the goals that Mary may have about an educational program is a reduction in the fear and panic displayed by many people. If she had two randomly selected groups, with one being experimental and the other being control, she could compare the mean of the two groups on a measure designed to reflect fear. To obtain the necessary data to test the null hypothesis that there is no difference between the experimental and the control, she could compare the mean of the two groups on a measure designed to reflect fear. The necessary data to test the null hypothesis that there is no difference between the experimental and control groups at the .01 level would be:

Experimental Group	Control Group
$N_1 = 34$	$N_2 = 32$
$M_1 = 75.25$	$M_2 = 70.25$
$s_1^2 = 21$	$s_2^2 = 22$
$t = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{75.25 - 70.25}{\sqrt{\frac{21}{34} + \frac{22}{32}}}$	
$t = \frac{5}{\sqrt{.62 + .69}} = \frac{5}{\sqrt{1.31}} = \frac{5}{1.14}$	
$t = 4.39$	

Table 12.3 Critical Values for Large Samples

	.05 Level	.01 Level
Two-tailed test	1.96	2.58
One-tailed test	1.64	2.33

Because the obtained t -value from the formula exceeds the z score of 2.58, Mary may reject the null hypothesis at the .01 level of significance. Table 12.3 shows the t -critical values for the rejection of the null hypothesis in samples with an N greater than 30.

When the samples are fewer than 30 in number (small samples), a t -table is used rather than the normal probability table. The reason for this is that the distribution curves of small samples are different from the normal curve.

The formula for testing the significance of the difference between two small sample means is:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{(N-1)s_1^2 + (N-1)s_2^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

The t -value, which is obtained from the formula, is then compared with the t -value from a t -table with $N_1 + N_2 - 2$ degrees of freedom. If the t -value obtained through the formula exceeds the table t -value at a specific probability level, then the null hypothesis is rejected at that level.

The t -Test for Two Dependent Sample Means

Often the health science researcher is involved with samples in which the composition of one group has bearing on the composition of the other group. For example, the subjects may be matched on one or more characteristics, or the same subjects may be in a pretest-posttest experiment. In such cases, the two groups are no longer independent, so a special t -test for dependent or correlated means is required. The measure to be analyzed is the difference between the paired scores.

The formula used is:

$$t = \frac{M_D}{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{N}}{N(N-1)}}$$

where D = the difference between the paired scores

M_D = the mean of the differences

$\sum D^2$ = the sum of the squared difference scores

N = the number of pairs

The t -value obtained is compared to those in the t -table with $N - 1$ degrees of freedom at the appropriate level of significance.

Box 12.2 ORIGINALITY: WHO IS TO JUDGE (LETTER TO THE EDITOR OF *LANCET*)

Sir—Collier and Vallance (Aug 28, p. 510) claim that it is difficult to assess originality. But there are some reliable tests. One is to submit your ideas to a reputable journal as a hypothesis. If the paper is accepted then it certainly is not original; if it is rejected then it might be. Another approach is to submit your work to peer review or even better to a committee of experts. If the work receives acclaim then it means that it is a part of the conventional wisdom, and is not original. If rejected, it might be original; if dismissed out of hand, it

probably is. The third, and most reliable, method is to present your work as a poster at a meeting. If nobody comes to view it, it means that nobody else is working in the field and your work is original—it might even be important. If a crowd gathers it is because they are all doing the same work as you.

J. Morris, Department of Pathology,
Lancaster Moor Hospital, Lancaster UK.
Lancet (1993, August 28) 342
(8870):510

Analysis of Variance (ANOVA)

When comparing the mean of two groups, on one independent variable, the health science researcher employs the *t*-test. However, if two or more groups are involved, one of the most powerful methods for comparing means is analysis of variance (ANOVA). If there is only one independent variable in the study, the ANOVA is called a **one-way ANOVA**. For example, in Case Study B, Mary has three types of instruction for teaching about HIV/AIDS. She intends to have a common knowledge exam. Random samples of her target population are assigned to each of the three instructional methods. The independent variable is the type of instruction (three types), and the dependent variable is the score on the exam. The null hypothesis is $H_0 = \mu_1 = \mu_2 = \mu_3$. In other words, there is no difference in the mean exam scores among the three instructional groups. It is possible to compute a *t*-test between the means of each pair, but the problems with this approach are (1) the need to ascertain a level of significance to compensate for “over-testing,” (2) the necessity of computing several tests, and (3) the possibility of errors in calculating so many tests. This is particularly so if several groups are involved—five different groups would require 10 separate *t*-tests. ANOVA is able to avoid these problems.

In ANOVA, as in the *t*-test, a ratio of observed differences/error is used to test hypotheses. The ratio, called the *F*-ratio, uses the variance of group means as a measure of observed differences among groups. The within-groups variance (V_w), simply the sum of the variances of each of the groups, is the denominator in the *F*-ratio. The between-groups variance (V_b), which measures the variation among the means of the groups, is the numerator in the *F*-ratio:

$$F = \frac{V_b}{V_w} = \frac{\text{between-groups variance or treatment}}{\text{within-groups variance or error}}$$

The total groups variance (V_{total}) equals the scores for all groups combined into one composite group.

The rationale of the F -ratio is as follows. The between-groups variance shows the influence of the experimental variable or treatment, while the within-groups variance represents the sampling error in the distributions. If the between-groups variance fails to be much greater than the within-groups variance, the health scientist would conclude that the difference between the means is likely caused by sampling error. On the other hand, if the F -ratio is substantially greater than 1, it would appear that the difference is likely the result of the treatment.

To determine whether the F -ratio is great enough to reject the null hypothesis at the pre-determined level of significance, the researcher must consult an F -table. Like the t -table, it contains the critical values necessary for testing. In entering an F -table, the appropriate degrees of freedom must be used. The between-groups variance (V_b) has $k - 1$ degrees of freedom (where k is the number of groups), and the within-groups variance (V_w) has $k(N - 1)$ degrees of freedom (where N is the number of observations in each group).

The analysis of variance is the first step in the analysis of such designs. If a significant F -ratio is obtained, it is only known that somewhere in the data something other than chance is operating. The researcher must employ a special form of the t -test to isolate the presence, nature, and extent of the influencing variable. Examples of such special t -tests are Duncan's multiple range test and tests by Newman-Keuls, Tukey, and Scheffé.

These **post hoc tests** differ somewhat in their ability to produce significant results (Berg & Latin, 2008). The Duncan Multiple Range test is more liberal than the Scheffé test and thereby is more likely to produce significant results when pairs of means are compared. It is important that the researcher be able to justify the post hoc test used.

If two independent variables are in the ANOVA analysis, it is called a **two-way ANOVA**. Using the example of Mary from above, we would now add a second explanatory variable, such as life experience with HIV/AIDS, coded as one of three ordinal categories: know someone with HIV/AIDS, living with someone who has HIV/AIDS, or inflicted with HIV/AIDS. In this scenario there is a null hypothesis for each of the independent variables. The researcher, in conducting the two-way ANOVA, may be able to calculate a statistical test for the interaction between the two independent variables. The interaction is simply the combined effect of the independent variables on the dependent variable (Wiersma, 2000). If this were to be done, a third null hypothesis would be used to note there is no interaction effect.

As you might anticipate, three or more independent variables can be used in a single analysis. In these instances the researcher may have a **three-way ANOVA** or even greater with more independent variables. This may be referred to as **multivariate analysis of variance (MANOVA)**. Factorial designs discussed in the chapter on experimental and quasi-experimental research lend themselves to MANOVA given the number of independent variables and possible interactions.

Analysis of covariance (ANCOVA) is a unique version of ANOVA in that it determines the effect of one or more independent variables while simultaneously controlling for the effect of some other variables (called covariates). Mary may wish to control for the covariate (variable) level of education. If she wishes to study the relationship between instructional method and experience with HIV/AIDS on knowledge of HIV/AIDS, she

can control for level of education. In so doing, another variable, level of education (elementary school, high school, and college) would be added. The analysis would be analysis of covariance. ANCOVA can also be of value when comparison groups can only be matched on the principal variable and not on others.

Measures of Relationship and Predictions

The chapter on descriptive data analysis addressed the issue of the relationship of one variable to another, explaining that the method used to describe such a relationship is linear correlation. For example, one might want to look at the relationship between disease severity and overall functionality. Simple linear correlation is not concerned with causation but simply the degree of the relationship symbolized by r . However, a related statistical procedure called regression is concerned with causation in many studies. In this type of research the investigator wants to determine the contribution of one or more independent (causing) variables on the dependent variable (the one caused). Regression can also be used to predict the value of one variable over that of other variables. The term **simple linear regression** refers to a study in which there is only one independent variable and one dependent variable (both are continuous), and the relationship can be illustrated in a straight line (see *Measures of Relationship* on page 247). It is the simplest form of prediction since only one predictor variable is used (Berg & Latin, 2008). For example, simple linear regression analysis could be used in our Case Study A to examine the relationship between hand washing and HA-MRSA and to predict HA-MRSA from hand washing (using the amount of time spent on hand washing so that it is a continuous variable). Of course, the incidence of HA-MRSA is continuous.

As the term implies, **multiple linear regression** assesses the linear relationship between two or more continuous independent variables and a single continuous dependent variable (Lang & Secic, 1997). In our case study, Bob may want to add a variable: the number of times the health care worker wore gloves out of the operating room rather than dispensing them before leaving. Therefore, he would include both independent variables in predicting HA-MRSA. Of course, additional independent variables could be included in the analysis. In this case, an additional variable could be, on a scale of 1 to 10, the degree of knowledge of CDC infection control guidelines. The results of the analysis would illustrate the strongest predictor to the weakest predictor. Table 12.4 illustrates possible findings for his study.

Table 12.4 Summary of Multiple Correlation of Variables with HA-MRSA

Variable	R	R ² × 100
Hand washing	−0.60	36.0%
Gloves	−0.80	64.0%
Degree of knowledge	−0.85	72.0%

The r between hand washing and HA-MRSA was $-.60$ thereby explaining 36% of the variance in HA-MRSA. The second best predictor, glove disposal, increased the multiple R to $-.80$ which explained the increased variance to 64%. The third variable increased the variance to 72%. Keep in mind when reviewing the table that the improvement in variance is due to the pooled correlation of the variables. For example, hand washing and glove disposal have a combined R of $-.80$, and together they explain the variance of 64%. As stated previously, the value of the independent variables (in so far as contribution to the dependent variable) can be recognized in the results.

Multiple linear regression is used in several ways, and it is recommended that texts dealing with this statistical technique be consulted. Similarly, in simple correlation, a statistics book can explain how a correlation coefficient r can be used in a **correlation hypothesis test** to test the null hypothesis ($H_0: r = 0$).

Nonparametric Tests of Significance

We have discussed tests of statistical significance known as *parametric statistics*. A parameter is a population score, whereas a statistic is a score for a sample that is randomly drawn from the population. There are assumptions made when using parametric statistics:

- Scores in the population are normally distributed about the mean.
- Population variances of the groups are approximately equal.

When deviations from these assumptions are in the data, parametric statistics should not be used, but rather one of the nonparametric statistical tools should be selected. These techniques do not make any assumptions about the population variance or shape of the data (Borg & Gall, 1999).

The advantages of nonparametric tests are (1) they do not have the many restrictions required for parametric tests; (2) they are very suitable for health surveys and experiments in which outcomes are difficult to quantify; and (3) they offer ease of computation. On the other hand, they (1) are less efficient, (2) are less specific, and (3) fail to deal with all the special characteristics of a distribution. Some of the most frequently used nonparametric tests are presented in this section.

The Chi-Square Test The chi-square test (χ^2) is generally employed in causal comparative studies and in comparison of observed and theoretical frequencies. As a test of independence, it is used to estimate the likelihood that some factor other than chance accounts for the apparent relationship—not to measure the degree of relationship. It is used when the research data are in the form of frequency counts.

The Mann-Whitney U-Test The Mann-Whitney U -test is the nonparametric counterpart of the parametric t -test. Simply, it is designed to test the significance of the difference between two randomly drawn samples from the same population. Usually each sample has less than 20 subjects because more than that allows the sampling distribution of U to approach a normal distribution wherein the t -test may be used.

The Sign Test The sign test is a procedure for determining the significance of the differences between two correlated or dependent samples. For example, experimental and control groups may be matched on several variables, or the subjects may be matched with themselves in a pretest-posttest situation wherein they act as a control group in one instance and an experimental group in another. The sign test is particularly useful when the treatment effect cannot be measured but only judged to result in inferior or superior performance.

The Median Test The median test, a nonparametric test, determines the significance of the difference between the medians of two independent groups, whereas the sign test operates with two correlated groups. It is an application of the χ^2 test with a 2×2 table and one degree of freedom.

The Wilcoxon Matched-Pairs Signed Rank Test The Wilcoxon matched-pairs signed rank test is employed by the health researcher to ascertain whether two samples differ from each other to a significant degree when there is a relationship between the samples. Although similar to the sign test, it is more powerful because it tests not only direction but also magnitude of difference between matched groups.

Table 12.5 Statistical Procedures

Procedure Name	Test Statistic	Parametric or Nonparametric	Purpose of Procedure
<i>t</i> -test for independent samples	<i>t</i>	P	Test the difference between the means of two independent samples
<i>t</i> -test for dependent (paired) samples	<i>t</i>	P	Test the difference between the means of two related groups or sets of scores
ANOVA	<i>F</i>	P	Test the difference among the means of a number of groups
Chi-square test	χ^2	NP	Test the difference in proportions in two or more groups
Mann-Whitney <i>U</i> -test	<i>U</i>	NP	Test the difference in the ranks of scores of two independent groups
Median test	χ^2	NP	Test the difference between the medians of two independent groups
Wilcoxon matched-pairs signed rank test	<i>Z</i>	NP	Test the difference in the ranks of scores of two related groups or tests of scores
Kruskal-Wallis test	<i>H</i>	NP	Test the difference in the ranks of scores of three or more independent groups
Kendall's coefficient of concordance	<i>W</i>	NP	Test that a correlation is different from zero

Note. Adapted from *Nursing Research* by D. Polit & B. Hungler, 1995, New York: Lippincott. Copyright 1995 by Lippincott.

The Kruskal-Wallis Test Developed along the same lines as the Mann-Whitney U -test, the Kruskal-Wallis test is the nonparametric correspondent to the parametric one-way ANOVA procedure. This would be used when the researcher wished to determine the significance of difference among three or more groups.

The Kendall Coefficient of Concordance This coefficient of concordance, frequently referred to as Kendall's concordance coefficient W or the concordance coefficient W , is used in research efforts involving rankings made by independent judges. The Kendall coefficient shows the degree to which such judges agree in their assignment of ranks.

Table 12.5 depicts a summary of some of the statistics tests we have discussed.

Meta-Analysis

Health science investigators usually conduct a literature search that addresses a particular research question in an attempt to “eyeball” the results and draw some overall conclusions. This approach, however, is very subjective, and the researcher can easily arrive at incorrect conclusions. This can be overcome by carrying out a systematic review that requires uniformity in identifying studies, displaying the results, and calculating a summary estimate of the overall results if appropriate. The statistical components of such a review are called **meta-analysis**.

Hulley et al. (2007) lists nine steps in conducting a good systematic review including meta-analysis:

1. Develop a very clear and precise research question
 - Research questions are similar for any investigation.
2. Identification of completed studies should be comprehensive and unbiased.
 - Develop a definitive strategy well before the results of the individual studies are known.
 - The strategy should be able to be repeated by others.
 - Searchers should include any appropriate databases.
3. Inclusion and exclusion criteria must be well-defined.
 - These criteria should be established before conducting the search.
 - Examples of criteria are: acceptable population, time period of the study, intervention, control groups, blinded, outcomes, and follow-up.
 - List studies that were considered but excluded and the reason why they were excluded.
4. There must be both uniform and unbiased abstraction of the findings and characteristics of the study.
 - Use predesigned forms that define eligibility and criteria.
 - May want to use two or more researchers for abstraction.
 - If necessary, contact authors to get more information.

5. Data from each study must be presented in a clear and uniform fashion.
 - Include characteristics such as sample size, length of follow-up, study outcomes, population characteristics, and methodology.
 - Show results of individual studies.
 - Use meta-analysis for summary estimates, confidence intervals and, where appropriate, subgroup and sensitivity analyses.
6. When appropriate, the calculation of a summary estimate of effect and confidence interval should be based on the findings of all eligible studies.
 - Select an appropriate method for calculating summary effect and confidence interval because different approaches using the same studies can produce different results (Cooper & Hedges, 1994; Hulley et al., 2007; Petitti, 1994).
7. There should be an assessment of heterogeneity of the findings of individual studies.
 - If the studies differ on important variables (outcomes, blinding) they should not be combined.
 - Similarly, findings should not be combined if they vary widely.
 - The variability in findings is called heterogeneity.
8. Similarly, there should be an assessment of potential publication bias.
 - Publication bias means that, generally speaking, studies with positive results are more frequently published than are those with negative results.
 - Unpublished works should be checked out by speaking with investigators, reviewing meeting presentations, and the like.
 - Depending on the nature of the data, they may not be able to be used in the meta-analysis.
9. Subgroup and sensitivity may be possible to include.
 - In reviewing the study, look for possible subsets of data that could be included in the process (especially when looking at large database studies).
 - Sensitivity analysis shows how sensitive the meta-analysis was to such items as design decisions and use of inclusion/exclusion criteria.

There are several limitations of meta-analysis, including (1) inclusion of poorly conducted studies in the equation, (2) cost of conducting the study, (3) criteria being used for including primary sources are difficult to agree upon, and (4) incomplete data are sometimes used because they are the only data available.

According to Daly, Kellehear, and Gliksmann (1997), the advantages of meta-analysis include (1) being able to identify a finding in a diverse array of similarly designed studies, (2) providing a systematic overview of findings in a particular area of study, (3) determining larger research questions, (4) enabling an alternative method when other methods are inappropriate, and (5) being able to conduct methodological assessments of research designs.

The literature shows a broad range of meta-analysis studies from structural brain change in ADHD (Ellison-Wright & Bullmore, 2008), to the effects of stress management intervention (O'Cleirigh & Safren, 2008), to marijuana and alcohol use among adolescents (Lemstra et al., 2008). There is no doubt that systematic review and meta-analysis will continue to be a valuable aid to scholars.

Case Discussion of Case Study A

Bob was concerned about the relationship of hand washing by the health-care professionals and the incidence of HA-MRSA. He knew there was a direct relationship between the two so therefore started a 6-month campaign to increase hand washing. He predicted that the incidence of HA-MRSA would decrease as hand washing increased. He checked this by using a simple single linear regression model, which would show a negative relationship between the two variables. To examine other variables, Bob used a multiple linear regression model by adding the use of gloves and knowledge about HA-MRSA. As to be expected, as hand washing, gloving, and knowledge increased, the incidence of HA-MRSA decreased, showing a strong negative relationship with all three variables and accounting for 72% of the variance.

Case Discussion of Case Study B

Mary wanted to implement faith-based HIV/AIDS programs for African Americans in hopes of increasing knowledge (including of risk-taking behaviors) and modifying attitudes toward people who had contracted HIV/AIDS. She developed null and alternative hypotheses and set the level of significance at .01. Mary chose to use a *t*-test for independent sample means since she drew random samples from the target population. The result was rejection of the null hypothesis, thereby showing that her program did in fact make a difference on her measured variables. If she had more than two means to compare, Mary would have used ANOVA techniques; if she wanted to control for a particular variable, such as level of education, she would have chosen ANCOVA.

SUMMARY

Chapter 12 began with a general discussion about inferential analysis and testing statistical significance. The differences between estimation of parameters and hypothesis testing were elucidated. Six steps in hypothesis testing were outlined and presented. The null and alternative hypothesis were reviewed. In regard to tests of significance, the critical region, level of significance, and Type I and Type II errors were discussed. Inferential data analysis techniques were explained and included the *t*-tests for unpaired and paired sample means, analysis of variance, and multivariate analysis of variance. Measures of relationship and prediction were addressed, including multiple regression. The nonparametric tests of significance that were discussed included Chi-square, Mann-Whitney, sign, median, Wilcoxon matched-pairs, Kruskal-Wallis, and the Kendall coefficient. The chapter concluded with a discussion of meta-analysis.

CRITICAL THINKING QUESTIONS

1. What is the difference between descriptive and inferential statistics?
2. How would you explain “estimation parameters”?
3. What are the steps in hypothesis testing?
4. Why would you use ANOVA rather than a *t*-test when comparing three sample means?
5. Which are more powerful, parametric or nonparametric tests of significance?

SUGGESTED ACTIVITIES

1. Read three articles on a topic that is of interest to you, and critique the data analysis utilized in each article.
2. Data were collected from a pretest and posttest of students taking a 2-week course in substance abuse. Would you use a *t*-test (and, if so, which one), or would you use ANOVA? Given the data below, complete the exercise to see if there is a significant difference between the two tests.

Pretest	Posttest
84	90
85	75
70	89
65	91
71	85
72	79
84	88
85	85
66	82
67	99

3. Describe some topic areas with which you might use meta-analysis.
4. Conduct a literature search for three studies that used multiple regression. Read each, and be able to explain how it was used and the results obtained.
5. Read Appendix A, and be able to explain how each test would be used.

REFERENCES

- Ary, D. (2006). *Introduction to research in education*. Belmont, CA: Thomson/Wadsworth.
- Berg, K. E., & Latin, R. W. (2008). *Research methods in health physical education, exercise science, and recreation* (3rd. ed.). Philadelphia: Lippincott Williams & Wilkins.
- Borg, W., & Gall, M. (1999). *Educational research: Applying a practical guide*. New York: Longman.
- Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Daly, J., Kellehear, A. & Gliksman, M. (1997). *The public health researcher*. New York: Oxford.
- Ellison-Wright, I., Ellison-Wright, Z., & Bullmore, E. (2008). Structural brain change in attention deficit disorder identified by meta-analysis. *BMC Psychiatry*, 8, 51.
- Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2007). *Designing clinical research* (3rd ed.). Philadelphia: Lippincott, Williams & Wilkins.
- Kuzma, J. W., & Bohnenblust, S. E. (2001). *Basic statistics for the health sciences* (5th ed.). Boston: McGraw-Hill.
- Lang, T. A., & Secic, M. (1997). *How to report statistics in medicine*. Philadelphia: American College of Physicians.
- Lemstra, M., Bennett, N. R., Neudorf, C., Kunst, A., Nannapaneni, U., Warren, L. M., et al. (2008). Meta-analysis of marijuana and alcohol use by socioeconomic status in adolescents aged 10–15. *Canadian Journal of Public Health*, 99(3), 172–177.
- O'Cleirigh, C., & Safren, S. (2008). Optimizing the effects of stress management interventions in HIV. *Health Psychology*, 27(3), 297–301.
- Petitti, D. (1994). *Meta-analysis, decision analysis and cost effectiveness analysis*. New York: Oxford University Press.
- Wiersma, W. (2000). *Research methods in education*. Boston: Allyn & Bacon.

Techniques for Data Presentation

KEY TERMS

cross tabulations
contingency tables
preparing figures

presentation of tables
quantitative material
single- or multiple-
variable form

table placement

Case Study

Caleb, a senior at Newtown University, is enrolled in a health sciences research course. The major semester assignment is to conduct a research study using the components learned in the course. In other words, Caleb is to state the problem, derive hypotheses, conduct a literature search, conceive the methodology, actually carry out the experiment, present and analyze the data, and summarize the project, including conclusions and recommendations. If he has read the text to this point, Caleb will be equipped with the necessary knowledge to complete the assignment. This chapter will enable Caleb to present the data he collects for his study. Usually, the form of data presentation includes tables, figures, and graphics.

Table Presentations

The first part of this chapter will discuss the **presentation of tables** in manuscripts for publications, reports, books, term papers, master's theses, and doctoral dissertations. The purpose for including tables and several types of tables will be described. In addition, we will discuss the relationship between the table and the text, which is a very important consideration. Finally, the format of tables will be detailed so that you can construct a proper table. The section concludes with guidelines for including tables.

Purpose of Tables

Tables usually represent **quantitative material**, and sometimes words that present qualitative comparisons or descriptive information. As an example of a word table, Caleb could use a table to depict some of the questions and their responses in the questionnaire he used. Word tables should not repeat what has been discussed in the text but rather illuminate that discussion. Using tables to depict collected data enables the reader to have a clear understanding and comprehension of the masses of numbers that have been collected during the project. The analysis of original data should be presented in these tables so that readers are not burdened with long lines of numbers that disrupt the smooth flow of the text. Data should be presented so that their significance is easily recognized by the reader.

Tables are used to present information in the form of totals and subtotals, rank-order relationships, and results of statistical analyses (Slade, 2003). Once you have decided what data belongs in the table, there are a few additional considerations:

1. Rounded-off values may display patterns more clearly than precise numbers.
2. It is easier to compare numbers down a column than across a row.
3. Row and column averages can provide a visual focus that allows the reader to inspect the data easily.
4. Ample spacing between rows and columns can improve a table because the white space creates a perceptual order to the data. (American Psychological Association [APA], 2001, p. 173).

It should be noted here that some institutions of higher education suggest or mandate use of a particular style or publication manual. You should consult with the appropriate personnel to determine which, if any, style manual is utilized. Many university health science departments use the *Publication Manual of the American Psychological Association* (APA, 2001), because it is preferred for most journals to which potential social sciences authors might submit their projects for publication.

Relating the Tables to the Text

A good table should supplement the text of the paper. However, you should refer to all tables and their data in the text. A discussion of the highlights of the table is all that is necessary in the paper, and each table should have a brief introduction that explains the manner in which the data are presented and suggests their general meaning. Two rules apply to the inclusion of tables:

1. Each table should be understandable without reference to the text.
2. The text should be complete so that the reader may follow it without referring to the tables.

In the text, tables are referred to by their numbers. They are numbered consecutively, using arabic numerals; for example:

Table 4 shows . . .
 . . . behavioral scores with no pretest (see Table 4)

The actual **table placement** is sometimes a difficult decision. Should it go after the analysis? Before? Here are some rules that may be helpful:

1. Each table should be placed entirely on one page when possible.
2. Text material may be placed on the same page with a table of about one-half page or less.
3. A table should be separated from the text by four spaces above and below (Table 13.1).
4. A page containing both a table and text should begin with the text material.
5. If all the preceding conditions cannot be met, a table should be placed between paragraphs (Slade, 2003).

Table 13.1 Mean Knowledge Scores of Pretested and Unpretested Students

This summarizes the mean knowledge scores of the students who took the health science test. The students were asked about their health knowledge in regard to smoking, nutrition, and exercise. In both the pretested and unpretested groups, girls achieved slightly higher scores on the test. However, scores of both groups were very similar and showed no difference between the pretested and unpretested groups.			
	Group	N	Health Science Test
Girls	Pretested	120	18
	Unpretested	110	20
Boys	Pretested	118	17
	Unpretested	108	19

Note: Maximum score was 30.

Single-Variable Tables

Single variables are usually used in a descriptive or explanatory study. Measurements such as the range, mean, mode, or median may be depicted in such a table. In addition, frequency distributions and grouped data may be presented. Table 13.1 is an example of a single variable table in that one score was reported. Table 13.2 depicts the frequency distribution for the scores on the health science knowledge test.

Percentage Tables

Percentages can provide an efficient way to summarize information. These data can be reported in **single- or multiple-variable form**. The listing of percentages as presented in Table 13.3 should be summed. An indication of sample size is also important so that the percentages will not be misleading. There is an important question you might be asking here: *What happens to the nonrespondents? Should they be included in the computation for the percentages?* There are two methods that you could choose to answer this dilemma. The first is to subtract the number of nonresponses from the total sample size and use this new figure as the base for the percentages. In our example, Caleb may have a sample size of 456, but only 411 students reported their gender. In this case,

Table 13.2 Frequency Distribution of Scores on Health Science Knowledge Test

Score	Frequency	Score	Frequency	Score	Frequency
1	0	11	18	21	38
2	0	12	20	22	20
3	0	13	22	23	17
4	0	14	20	24	14
5	0	15	25	25	13
6	0	16	35	26	7
7	6	17	36	27	4
8	7	18	40	28	0
9	15	19	41	29	0
10	14	20	44	30	0

the 45 nonrespondents would be omitted, and 411 rather than 456 would be the base for computing the respective percentages. A second method of dealing with nonrespondents is to use the total sample size (456) as the base and include the nonrespondents as a percentage. In our case study of 456, assume that 205 were males, 206 were females, and 45 were nonrespondents. If we used the first method (using 411 as a base), we would be able to write that 50% were males and 50% females. By the second method (456 as base), we would have 45% males, 45% females, and 10% nonrespondents. If we continue to use nonrespondents as part of the analysis, the base number remains constant and adds stability to all the analyses.

Table 13.3 Sex Distribution for Students Taking Health Science Knowledge Test

Sex	Percentage
Males	45
Females	55
Total	100
	(N = 456)

Contingency (Bivariate) Tables

Health science studies, as well as other social science investigations, sometimes focus upon the relationship between two variables. Tables are used to display the way the values of the variables are associated. Interrelationships are examined and are thus called **cross-tabulations** or **contingency tables**. In these kinds of tables, all combinations of categories of all the variables are presented. However, the most usual form is the two-variable (bivariate) table, with each variable being dichotomous. Therefore, two dichotomous variables present a table with four cells, sometimes labeled a *four-fold*

table. Through evolution, one variable has been termed the *column variable* and is labeled across the top so its categories form vertical columns down the page and usually represent the independent variable. The other variable, the *row variable*, is labeled on the left margin, forming categories of rows across the page and thus is the dependent variable. The intersections of the categories of these two variables form the interior of the table.

An easy way to construct this contingency table is to list the total frequencies in each category for each variable. The simplest dichotomous variables and those most frequently used are gender and race. If we had 100 respondents to a survey concerning health risk factors, we would have 100 as a total for gender and race (each person has a gender and race). Table 13.4 shows that we have 50 males: 30 blacks and 20 whites. The numbers outside the square are referred to as *marginals*. The row marginals are 60 and 40, and the column marginals are 50 and 50. The row and column marginals provide no information about each other but are found in the interior of the cells. You should, once again, be sure to note the N (100) in the lower right-hand corner, which is attained by adding either the row or column marginals.

Absolute numbers should be used in the cells when a statistical analysis will be conducted. This analysis is always placed at the bottom of the table. If, however, no statistical analysis will be used, then percentages should be used in the cells. Usually, the independent variable is presented in percentages; thus the columns become percentaged.

Multivariable Tables

Multivariable tables contain three or more variables and are similar to those presented in the previous sections. Caleb would use such a table in his paper if he wanted to report on a number of variables and their association or correlation. The format is typically a correlation matrix, which presents the correlations between all the pairs of variables in the analysis. If we had five variables, we would want to show the relationship between each pair of variables. The matrix, as in Table 13.5, would list all the variables except for the last one along the left-hand margin of the table. At the top of the table, the variables are listed beginning with the second variable. The obtained correlations are presented once for each pair of variables. The dependent variable should be placed in the last column, column E in Table 13.5. Now you can easily determine the relationship between the dependent variable and each of the independent variables.

Table 13.4 Race by Gender

Race	Male	Female	
Black	30 (Cell a)	30 (Cell b)	60 $a + b$
White	20 (Cell c)	20 (Cell d)	40 $c + d$
	50 $a + c$	50 $b + d$	$N = 100$ $a + b = c + d$

Table 13.5 Correlation Matrix-Multivariable Table

Variable	B	C	D	E
A	.10	.30	.40	.35
B		.72	.50	.62
C			.21	.41
D				.48

Another type of multivariable table that is frequently used is the analysis of variance table (ANOVA). Table 13.6 is an example of how to set up the ANOVA table.

Word Tables

Some tables may consist mainly of words. These tables present qualitative comparisons or perhaps descriptive information. An example can be found in Chapter 9 (see Table 9.1), where various evaluation models were described. Word tables should explain the discussion in the text but not repeat that discussion. When preparing a word table, you should maintain the same format as you would use in other tables, and be sure to double-space all parts of the word table.

Table Format

As we have mentioned previously, each institution will have a set of guidelines for personnel to use when submitting reports, articles, and the like. Caleb would need to find out what is needed for his class and institution. The following discussion should enable you to ascertain the many different terms associated with constructing a table. Table 13.7 has been devised to depict these many terms.

Table Numbers

All tables should be numbered consecutively with arabic numerals. In a book chapter, use sequential numbers preceded by the chapter number and a decimal point. We have used this method throughout this book. In the text of your report, paper, or thesis, refer to the tables by number, not by the title. If your manuscript includes an appendix with

Table 13.6 Analysis of Variance (ANOVA) Table

Source of Variance	SS	df	MS	F	Level of Significance
Between groups	261.1	2	100.1	21.0	.01*
Within groups	687.8	144	477.7		
Total	888.9	146	577.8		

* $p < .01$

tables, identify the tables of the appendix with capital letters and arabic numerals: Table B.1, for example, would be the first table of Appendix B.

Table Titles

The title of a table should be clear and brief and should explain the table. *Avoid* using information contained in the headings of the table; such as the following for Table 13.7:

Relation Between Attitude and Knowledge Tests

This would be unclear and not tell what data are contained in the table. Another bad title would be:

Mean Health Science Knowledge and Attitude Scores of Girls and Boys in Grades 7, 8, and 9 Who Were Pretested and Not Pretested

This is too detailed and duplicates the information in the table headings. A better example is:

Mean Scores of Students with Pretesting and Without Pretesting

This is a *good* table title in that it explains clearly what the data will tell the reader.

Table 13.7 Mean Scores of Students with Pretesting and Without Pretesting

table number
and table title

Group	N ^a	Grade		
		7	8	9
Knowledge Tests				
Boys				
Pretested	118	17	18	20
Unpretested	108	19	22	21
Girls				
Pretested	120	18	19	20
Unpretested	110	20	21	20
Attitude Tests				
Boys				
Pretested	118	18	20 ^b	22
Unpretested	108	20	22	23
Girls				
Pretested	120	19	20	21
Unpretested	110	22	24	23

Note: Maximum score on each test was 30.

^aNumbers of students out of 125 in each group who completed both tests.

^bTwo boys had identical answers.

Table Body

The body of a table contains the data. Mullins (1983) suggests the following guidelines when constructing the body of a table:

1. Use as few entries as possible without eliminating vital information.
2. Within each table, use the same rules for retaining decimals and for rounding.
3. Arrange entries so that the most important comparisons are between adjacent numbers.
4. To prevent confusion of percentages and numbers, place a percent sign (%) after the first number in a column of percentages that add up to 100%. Also, use “percentage” in the column heading.
5. If a column head does not apply to an item in a row stub (this is called a *cell*), leave the cell blank.
6. If rounding prevents the sum of percentages in an additive column from being 100%, use a footnote to explain why this is so.
7. Do not use intersecting lines to connect items in different columns.

Stub

Stub is the name for the rows in the far left-hand column of the table. The names of these columns should be short, clear, and grammatically consistent. The stubs usually present the independent variables. If you use abbreviations, use them consistently in all tables and use a note to explain the abbreviations in your first table.

Stub Column

The stub column is the column of row stubs and their subcategories. In Table 13.7 these rows are labeled *Boys*, *Girls*, *Pretested*, and *Unpretested*. The subcategories should be indented at least one space from the margin to distinguish them from the row stubs.

Stub Head

The stub head is the title of the stub column. In Table 13.7, *Group* is the stub head.

Column Head

The column head names the column and should be grammatically consistent with others in the table. These heads usually name the dependent variables, dependent upon the discipline and guidelines followed. In Table 13.7, for example, 7, 8, and 9 are the column heads.

Column Spanner

The column spanner identifies two or more columns, each of which has its own column head. In Table 13.7, *Grade* is the column spanner.

Table Spanner

Table spanners cover the entire width of the body of the table, allowing for further divisions within the table. *Knowledge Tests* and *Attitude Tests* in Table 13.7 denote table spanners. Table spanners can also be utilized to combine two tables into one, as they are in Table 13.7.

Notes

Tables have three types of notes: general notes, specific notes, and probability level notes. These are always placed below the table.

General Note

General notes explain, qualify, or provide information relating to the entire table. This may include an explanation of abbreviations, symbols, and so on. These notes are designated by the word *Note* (italicized, followed by a period). In Table 13.7, it reads:

Note. Maximum score on each test was 30.

In addition, general notes indicate that if a table has been reprinted from another source. To do this, you must obtain permission to reproduce or adapt all or part of a table from a copyrighted source. Give credit to the original author and to the copyright holder.

Example of Note from Book

Note. From *Contemporary Human Sexuality* (p. 42) by J. Turner & L. Rubinson, 1993, Upper Saddle River, NJ; Prentice-Hall. Copyright 1993 by Prentice-Hall. Reprinted by permission.

Example of Note from Article

Note. From “Acquaintance rape: The influence of alcohol, fraternity membership, and sport team membership” by M. Fritner and L. Rubinson, *Journal of Sex Education and Therapy*, 19(4), 272–284.

Specific Note A specific note refers to a particular column or to an individual entry. Specific notes are indicated by italicized superscript lowercase letters (superscript *a* in Table 13.7, for example). Within the headings and body of the table, the superscripts are ordered horizontally from left to right across the table by rows, beginning at the top left. Each table is independent of any others; therefore, notes always begin on each table, with superscript *a*.

Probability Level Note A probability level note indicates the results of significance tests. Asterisks indicate the probability levels of tests of significance. When more than one level appears in a table, use one asterisk for the lowest level, two for the next, and so on. These levels and the number of asterisks do not have to be consistent from table to table. Table 13.8 is an example of a probability level note.

Table 13.8 Example of a Probability Level Note

<i>E</i>
1.70*
3.86**

* $p < .05$ ** $p < .01$

Note Format The ordering of the notes in a table is (1) general; (2) specific; and (3) probability level.

Example of Order of Notes

Note. Maximum score on each test was 30.

^aNumbers of students out of 125 in each group who completed both tests.

* $p < .05$

** $p < .01$

Each type of note begins at the margin on a new line below the table, beginning with the general note. The first specific note begins flush left on a new line, and all subsequent specific notes follow one after the other. The first probability level notes follow one another. These parts of a table have been presented so that you can easily construct a good, coherent, and useful table. The following are some guidelines you might use when constructing your tables:

- Is the table necessary?
- Is the entire table—including the title, headings, and notes—double-spaced?
- Are all comparable tables in the manuscript consistent in presentation?
- Is the title brief but explanatory?
- Does every column have a column heading?
- Are all abbreviations, underlines, parentheses, dashes, and special symbols explained?
- Are all probability level values correctly identified, and are asterisks attached to the appropriate table entries? Is a probability level assigned the same number of asterisks if it appears in more than one table?
- Are the notes in the following order: general note, specific note, probability note?
- Are all vertical rules eliminated?
- Will the table fit across the width of a journal column or page?
- If all or part of a copyrighted table is reproduced, do the table notes give full credit to the copyright owner? Have you received written permission from the copyright holder and sent a copy to the APA production office?
- Is the table referred to in text? (APA, 2001, p. 176)

Figure Presentations

This section describes how figures should be utilized in a manuscript. The various types of figures will be discussed, as well as how to cite figures in the text. A general discussion of instructions for preparing figures with an accompanying list of guidelines completes the section. We have determined, for use in this text, that figures encompass any type of illustration other than a table. These may be in the form of charts, graphs, photographs, maps, drawings, and some kinds of computer printouts. The author provides these materials for the publisher to include.

Purpose of Figures

Figures, as mentioned, refer to charts, graphs, drawings, maps, and photographs. They are used to present data very clearly and concisely. The inclusion of a figure should be carefully considered, because figures can be expensive to produce, both for the author and publisher. Therefore, figures should be used only when they actually contribute something to a paper. Some points to consider when including a figure are:

- What idea do you need to convey?
- Is the figure necessary? If it duplicates text, it is not necessary. If it complements text or eliminates lengthy discussion, it may be the most efficient way to present the information.
- What type of figure (e.g., graph, chart, diagram, drawing, map, or photograph) is most suited to your purpose? Will a simple, relatively inexpensive figure convey the point as well as an elaborate, expensive figure? (APA, 2001, p. 176)

Sometimes it becomes confusing to decide if you should use a table or a figure. A good rule of thumb might be: if the data shows trends, it could be better augmented by a figure rather than a table. Remember that a good figure should not duplicate what is contained in the text, should be easy to read and understand, and should be carefully prepared. You can employ a professional artist to do the work, or you can attempt the project yourself. Check the guidelines from your college or university or from the publisher (if you are submitting a manuscript for publication).

Types of Figures

There are many types of figures. Those we will discuss here include graphs, charts, dot maps, drawings, and photographs.

Graphs Graphs usually show how things are compared or distributed. These come in the form of percentages or absolute values. There are several types of graphs: line graphs, bar graphs, circle graphs, and scatter graphs.

Line Graphs. Line graphs are used to show trends or results of a line series experiment. The independent variable is plotted on the x axis (horizontal), and the dependent variable is plotted on the y axis (vertical). See Figure 13.1 for an example. The length of the y axis should be approximately two-thirds the length of the x axis. The grid marks

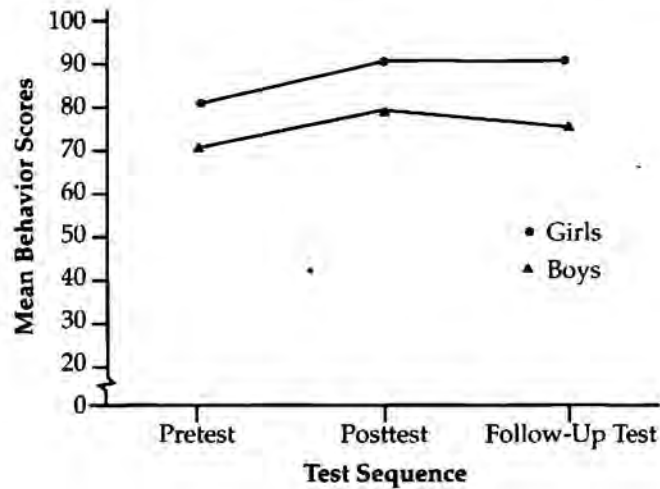


Figure 13.1 Example of Line Graph

(dashes) on the axes denote the units of measurement. If changes on the axes are disproportionate, the differences will be distorted. Thus, the curve or slant of the line must accurately depict the data. Notice the double slash on the axes in Figure 13.1. This indicates that the origin of the coordinates is not zero.

Bar Graphs. Bar graphs are easy to read and construct (see Figure 13.2). Solid vertical or horizontal bars present one kind of data. There are also subdivided bar graphs (each bar

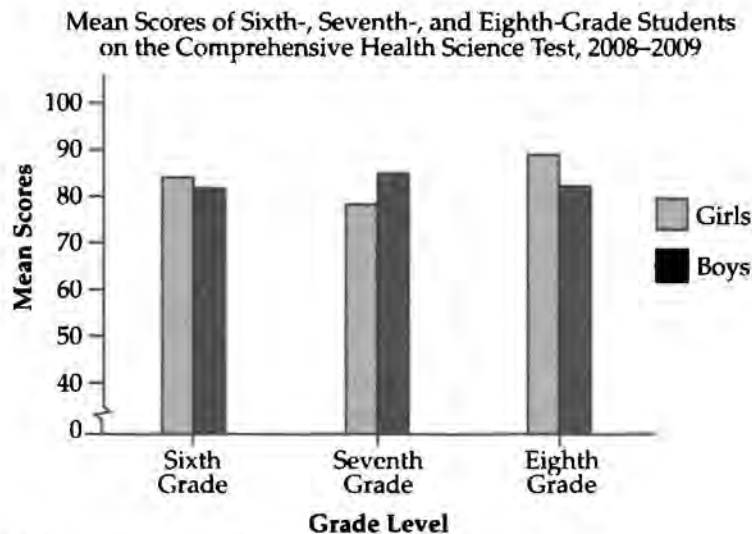


Figure 13.2 Example of Bar Graph

Time Allotted for Health Science Classes
Newbury High School, 2000–2009

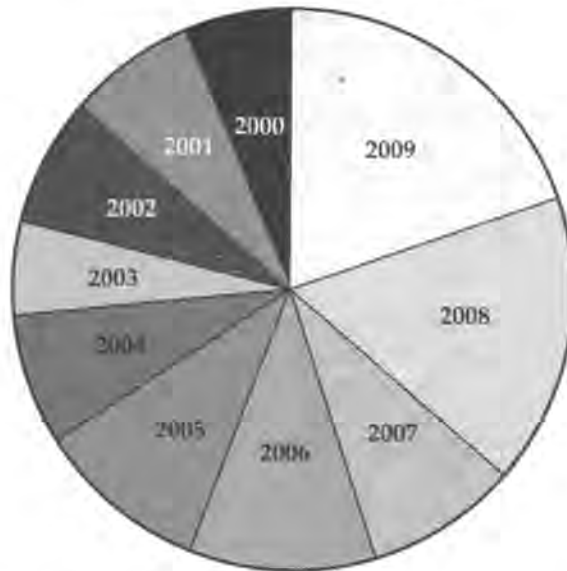


Figure 13.3 Example of Circle Graph

shows two or more kinds of data); multiple bar graphs (whole bars represent different variables in one data set); and sliding bar graphs (bars are split by a vertical line that serves as the reference for each bar).

Circle Graphs. Circle or pie graphs show percentages and proportions (see Figure 13.3). A general rule to follow, for clarity, is to depict only five or fewer items. The segments should be ordered from large to small, with the largest segment beginning at the 12:00 position and moving clockwise to the smallest. The differences in segments should be highlighted from light to dark, with the smallest portion being the darkest. Using different patterns of lines and dots shows the shaded patterns.

Scatter Graphs. Scatter graphs consist of single dots that are plotted as on a line graph, but the dots are not joined together (see Figure 13.4). The dots represent where the variables intersect, and a cluster of dots along a diagonal indicates a correlation.

Charts Charts can describe relationships between group segments in the sequence of operations in a process. These are usually depicted by boxes that are connected by lines. Examples include charts of organizations (see Figure 13.5), flow charts that show a step-by-step process, and schematics that show components in a system (e.g., a circuit board).

Drawings Drawings are usually prepared by a professional artist because they are difficult to accomplish. The drawing should be as simple as possible so that the author's idea can be easily conveyed. A drawing enables the author to augment the

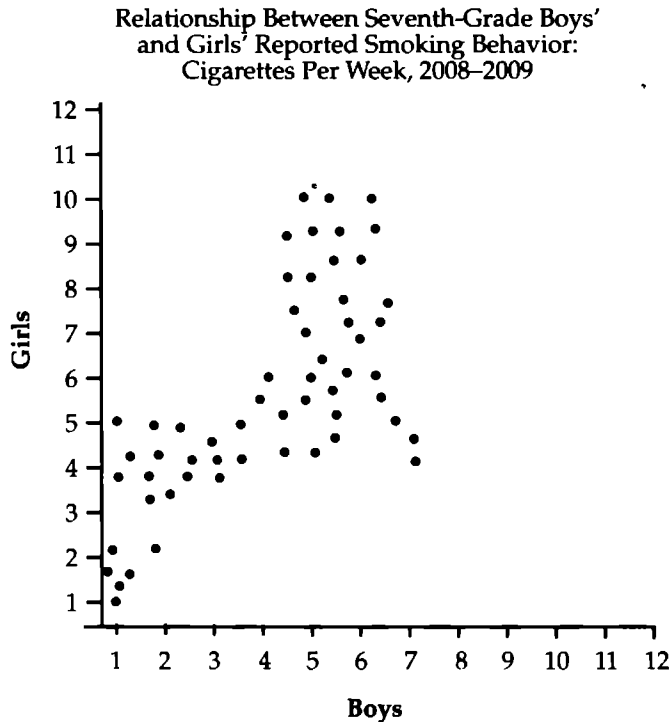


Figure 13.4 Example of Scatter Graph

manuscript by providing ideas and images from different viewpoints: two-dimensional and side views.

Photographs Professional photography is important when photographs are used in a manuscript. They can provide focus and interest points and add something of value to the words in a manuscript. Inform the photographer that there should be a strong contrast between the subject and the background. Check guidelines carefully as the use of black-and-white film is often mandatory. When photographing people, attempt to get a signed consent form from those people. When using a photograph from another source (book, journal, or other), obtain the original picture (because photographs of photographs do not reproduce adequately). You must also obtain written permission to reprint from the copyright holder and must acknowledge the holder in the figure caption (APA, 2001).

Citing of Figures

When using a figure, it should be placed as close as possible to the first reference made to it in the manuscript. The figures should be consecutively numbered with arabic numerals in the order as they are mentioned (e.g., *Figure 1*, *Figure 2*, *Figure 3*). They can be linked to a chapter number, if appropriate. The number should be written lightly in pencil on

Department of Public Health and Health Behavior Organizational Chart

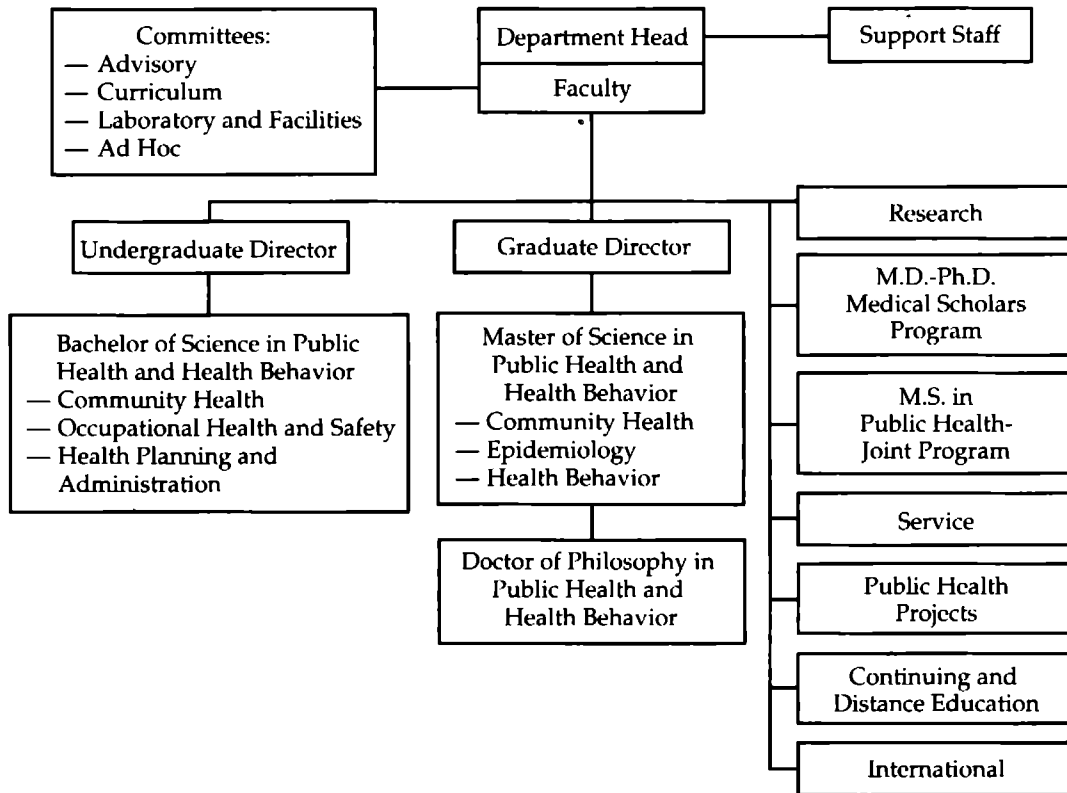


Figure 13.5 Example of Organizational Chart

the back of the figure, near the edge. In addition, (also on the back of the figure) note the top (top) of the figure and write the figure's title (again, lightly). In the manuscript refer to figures by numbers; for example:

Figure 1 shows . . .

The data are related (see Figure 1)

Avoid writing "see the figure above or below" on a specific page. This is because the placement of figures cannot be determined until the manuscript is typeset. The printer should be apprised of the approximate placement of the figure by a break in the text and a note:

Insert Figure 1 about here

The same procedure is used for noting the placement of tables.

Captions and Legends

The *caption* is the explanation of the figure and is placed below it. The caption describes the contents of the figure in a sentence or a phrase.

Example of Figure Caption

Figure 1. Time and set point between attitude and behavior surveys.

Information that is needed to clarify the limits of measurement should be placed in parentheses after the caption. It is important that all terminology used in the text and the figures agree.

The *legend* is a key to the symbols in the figure. There are some standard legend symbols: ■, ▲, and ●. Figure 13.6 is an example of a figure with a legend. The legend is put into the figure and should have the same kind and size of lettering that is utilized in the figure.

All figure captions should be typed (double-spaced) on a separate page as a figure captions list. The notation *Figure Captions* should appear in the top center of the page,

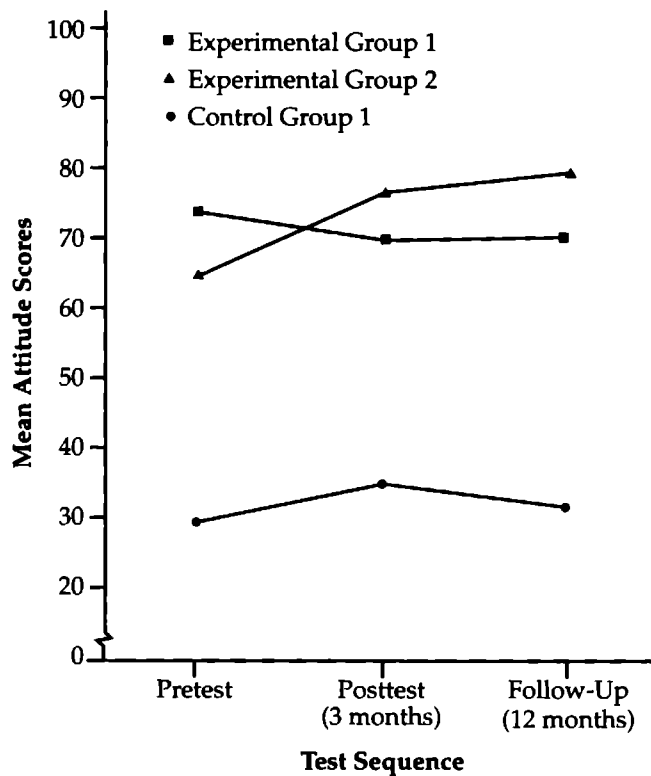


Figure 13.6 Mean Attitude Scores of Eighth-Grade Students Over 12 Months

and each caption should be flush to the left margin. Underline *Figure* as well as the number; follow this with a period and the text of the caption. Capitalize only the first word and proper names. Double-space if there is more than one line. An example is:

Figure 1. The relationship of skills and attitudes to health science knowledge.

Instructions for Preparing Figures

The following list provides some general instructions for **preparing figures** for manuscripts, follow this with reports, term papers, theses, dissertations, and so on, that will be submitted in hardcopy:

1. Place each figure on a separate page measuring $8\frac{1}{2} \times 11$ in. Sometimes you will have to use larger paper. In these cases, check your local format procedures.
2. Check to make sure that the figure is necessary.
3. Do not type letters; employ a professional or use a lettering stencil or appropriate software.
4. Minimize the number of lines, and check to see if the data are plotted properly.
5. Make sure all words are correctly spelled.
6. Check the legend to ensure that it is clear.
7. Use arabic numerals, and be sure that all figures are mentioned in the text.
8. In pencil, lightly write *TOP* on the back of each figure to indicate which side is the top.
9. Identify all figures with their figure numbers by lightly writing on the back of each in pencil.
10. Receive written permissions for all figures for which they are necessary and include them in the package.
11. Write all figure captions as a captions list on a separate page.

In some cases you may submit a project electronically. However, it is still important to prepare a completed hard copy to ensure all tables, figures, graphs, and text are present and correct. Below are some helpful hints to consider when submitting your data:

1. The figures on the screen will not necessarily print at the same size.
2. Print out all figures, tables, and graphs before you submit them.
3. Check to see that all items are numbered correctly.
4. Preview your figures before submission.
5. Prepare the data points so that they are easily visible within the parameters of the figure.
6. Ensure that your figures and graphs can be exported through graph export systems such as PDF by Adobe, EMF file format, etc.
7. Generally, you should utilize black and white figures and graphs.

8. If your figure is part of a proposed journal article, check with the journal to ascertain their accepted formats (EXCEL for figures, Photoshop, PowerPoint, Illustrator, Freehand, Corel Draw, and others for illustrations and other images).

Graphics and the Computer

Computer programs have given researchers distinct advantages in the area of graphics. Graphics programs enable researchers to easily depict data in the form of graphs, figures, and tables without the labor intensive efforts of measuring, cutting, and pasting by hand necessary in the past. While you may still want to sketch graphics by hand before beginning to create them in a software program, it will generally be easier to create your final graphics on a computer. It is recommended that you consult with members of your faculty as to which graphics programs are acceptable for data presentation.

Case Discussion

Caleb, in working through his project, would need to consider the objective(s) of his study, the nature of his data and findings, and his audience (for the report or presentation), and then select the best mode of presentation. If he has access to a computer with a statistics and graphics package, he should be able to generate several different types of tables, graphs, and charts. However, he must be careful not to “overkill” his presentation; instead, he must keep it simple and to the point.

SUMMARY

This chapter gave an overview of how to present tables and figures in a term paper, research report, thesis, or dissertation. Each student should check with his or her college's or university's required format before embarking upon writing. The use of a guide makes writing and working with tables and figures much easier.

Tables are usually representative of quantitative material. However, words may be used in a table to present qualitative comparisons or descriptive information. When deciding to utilize a table, make sure that the table enhances the manuscript and gives the reader a clearer understanding and comprehension of the numbers mentioned in the manuscript. A good table *supplements* the text and must be mentioned or referred to in the body of the paper.

There are several types of tables that you may need to use in a paper: single variable, percentage, contingency (bivariable), and multivariable. Regardless of the type, the format is the same for any kind of table. Formatting is dependent upon the particular style you or your institution adheres to and should be carefully followed. Several guidelines in constructing a table were mentioned.

Figures can be difficult to construct because they may require professional lettering and spacing. Figures include charts, graphs, drawings, maps, and photographs. A cautionary note: Make sure that the figure is necessary and that it augments the text. When

submitting figures along with a manuscript, include a separate page listing the figures because this ensures correct captions. Each figure should be titled in light pencil on its reverse, along with a notation as to which end is the top. Legends denote the symbols used in the figure. Symbols are used to differentiate between groups, sexes, and so on. It is important that these symbols remain consistent within and between the figures in the manuscript. There have been many applications for computer graphics, especially in the fields of engineering, chemical sciences, arts, education, and business. Graphics enhance the text, allow faster productivity, and can include a variety of interactive activities.

CRITICAL THINKING QUESTIONS

1. Why is it necessary to present tables and figures in a consistent manner?
2. Differentiate between single and multivariate tables. Give an example of each.
3. Discuss bivariate tables. Explain some of the categories utilized in each of these tables.
4. What are the purposes of figures?
5. Explain when you would use a scatter graph.
6. Distinguish between captions and legends. Give an example of each.

SUGGESTED ACTIVITIES

1. Prepare the following types of tables, using data from any journal article: single variable, percentage, contingency, and multivariable.
2. Using a different article than the one you used in Activity 1, prepare a figure to illustrate an apparent trend.
3. Use a computer graphics program to prepare one table and one graph.
4. Search the Web for the Centers for Disease Control and Prevention home page and find two topics that interest you. Utilize the appropriate website and find one table for each topic. Attempt to interpret each table, and critique (both positively and negatively) the tables. How would you improve upon the presentation of the material?
5. Determine how a chart differs from a table or graph. Use the Web to find a chart and then rework it into a table or graph. Explain how the content of the material might change once you have accomplished this activity.

REFERENCES

- American Psychological Association. (2001) *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: APA.
- Mullins, C. (1983). *A guide to writing and publishing in the social and behavioral sciences* (2nd ed.). New York: Wiley.
- Slade, C. (2003). *Form and style* (3rd ed.). Geneva, IL: Houghton Mifflin.

Writing a Research Report

KEY TERMS

acknowledgment page
communication of the study

copyright
research questions

standardized measures
title page

The preceding chapters have dealt with the many aspects of conducting research. Once the research study is complete, the project needs to be communicated to other interested professionals. The communication may take the form of a term paper, research report, manuscript for a journal article, master's thesis, or doctoral dissertation. This chapter will provide some guidelines to help you prepare the research report. Keep in mind, however, that style manuals, university and college requirements, and each journal require different types of written preparation. Make sure to check the format that is required *before* you embark upon the writing of your study.

The Report as a Communication Document

Health science is exactly that: a science you have learned through many years of schooling. A good health scientist begins with a theory, devises hypotheses based on that theory, designs and carries out an investigation to test the stated hypotheses, and analyzes the collected data to ascertain if these hypotheses can be accepted (or rejected). The last step, which can be the most exciting, is to communicate this process to other health scientists and interested people. The study has been well planned and thought out, and it was carried through as expected. Usually, several obstacles have been faced and successfully dealt with. These obstacles and other problems undoubtedly alter the search process, therefore causing unexpected results. This is the exciting part! If the research project had progressed as smoothly as planned, the report could have been written before the study was conducted. When you begin to think about writing the research report, you should probably start by thinking about the data you have collected and what these results actually mean. This process leads to finding unexpected results, which may lead to new, exciting, and extremely

relevant information for the field of health science. Some researchers are reluctant to report negative results or results that do not support their theories. All good research should be communicated; we have a responsibility to our colleagues to inform, reaffirm, and present new findings related to human health behavior. It is this communication that encourages others to conduct investigations that will help the growth of the health sciences.

The research report is generally divided into the following sections:

1. **Introduction:** An explanation of the problem and why it is important.
2. **Review of the literature:** A review of relevant studies focusing on the theory and its importance and implication for the study.
3. **Methodology:** A description of the procedures, subjects, and instruments employed in the study.
4. **Presentation and analysis of the data:** A discussion of the method used to analyze the data, a presentation of the findings, and a discussion of what the findings mean.
5. **Conclusions and summary:** Conclusions and a brief summary close the report.
6. **References:** All cited sources are included here.
7. **Appendix:** If appropriate, this section may include instruments, letters of commitment, etc.

We cannot emphasize one point too strongly: Remember, all journals have their own format, as do several style manuals used by different colleges and universities. The following material will enable you to write the research report, regardless of the format, because the information can always be rearranged to suit a particular format.

Preliminaries

We have outlined the many parts to the research report. However, certain *preliminary* pages might be necessary for a manuscript, journal article, thesis, or dissertation.

Title Page

The first page of the report is called the **title page**. The page usually includes the title of the paper, author, relationship of the report to either a course or degree requirement, name of the institution to which the report is submitted, and date of submission. The title of the report should be brief but specific in terms of describing the project. For example, if you were to conduct an experiment to compare the behavioral changes of diabetes patients who completed a new patient education course as compared to the behavioral changes of a matched group who were exposed to the regular patient education instruction, a good title might be: “A Comparison of Experimental and Regular Diabetes Patient Education Instruction on the Behavioral Changes of Diabetes Patients.” A second choice—“Behavioral Changes of Diabetes Patients”—is too short and does not really describe the experiment.

The title should be typed in capital letters, single-spaced, and centered. If there is more than one line in the title, it is divided so that each successive line is shorter, forming an inverse pyramid. Figure 14.1 is an example of a title page.

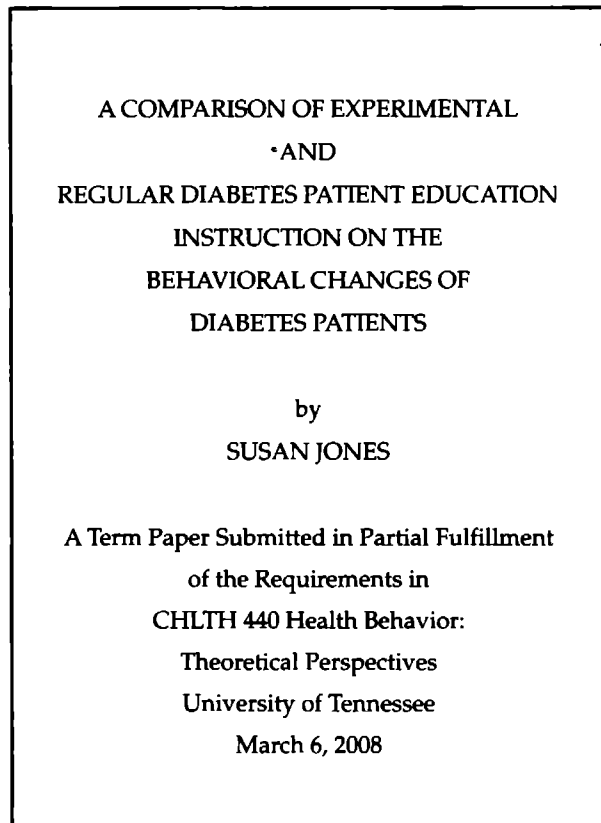


Figure 14.1 Title Page

Copyright

Most doctoral dissertations have a **copyright**. Theses and other reports may also be copyrighted, if you wish. To initiate this, the author must obtain a copyright authorization form (usually available from university or college business offices), complete it, pay the fee, and include a copyright notice at the front of the dissertation or report. The notice appears, centered on a single page, as follows:

© Copyright by
(author's name)
(date)

If the author has received a great deal of assistance from a person or persons, an **acknowledgment page** is included at the beginning of the thesis or dissertation. This page should be kept simple and to the point. Usually, the student's committee members and family members are mentioned. In addition, it is considered correct to give mention to the participants and other site personnel. If you are preparing a report or journal article, and there are groups or individuals who should be acknowledged, you can do so by providing a note to be placed on the first page (at the bottom) of the manuscript. The guidelines for

acknowledgments here are slightly different from those used for theses or dissertations. The author would not acknowledge committee or family members but might mention those who have helped in the data collection process or those who have reviewed the manuscript previous to it being submitted for publication.

Table of Contents

The table of contents provides an outline of the contents of the paper. The major headings and subheadings are included along with the page number of each. Figure 14.2

TABLE OF CONTENTS	
BACKGROUND AND SIGNIFICANCE	1
REVIEW OF THE LITERATURE	2
Prevalence of Dental Disease	2
Importance of Dental Health Education	3
Psychological Role of Parents	4
Efficacy of a Preschool Oral Health Program	7
HYPOTHESES	8
METHODOLOGY	10
Planning Phase	10
Pilot Phase	12
Experimental Phase	12
Data Analysis	16
EVALUATION	17
Students	17
Teachers	18
Parents	18
SUMMARY	19
REFERENCES	21
APPENDIX A—Preschools	22
APPENDIX B—Early Childhood Experts	23
APPENDIX C—Dental Society	24
APPENDIX D—Evaluative Plan	25

Figure 14.2 Table of Contents

shows a table of contents for a research report. The table of contents for theses and dissertations will vary according to the style utilized by individual institutions.

List of Tables and Figures

When writing a report you may utilize figures and or tables to supplement the manuscript. If you do, a separate list for each (tables and figures) should be included after the table of contents. The exact table titles and numbers used in the report are presented along with the pages on which they are located.

Abstract

The abstract is a brief summary of the problems, methods, results, and conclusions of the study; it gives a short rendition of the manuscript. This enables the reader to determine if it is necessary to continue reading the remainder of the report or article. Abstracts for theses and dissertations are usually presented after the table of contents. They are paginated in lower-case roman numerals, as are all preliminary materials mentioned in this section. An example of an abstract is shown in Figure 14.3.

Text or Main Body of the Report

The main portion of the report consists of an introduction, review of the literature, methodology, presentation and analysis of the data, and conclusions and summary. Each of these sections serves as a **communication of the study**.

Introduction

The introduction serves to state the problem and coordinate that problem to the literature that has been published. It also enables the writer to make the argument for the study. Hypotheses or research questions are formed from reviewing the relevant literature. Delimitations, definition of terms, and assumptions are also part of the introduction. Finally, the importance of the research is described.

Statement of the Problem

The problem statement occurs within the first or second paragraph, setting up a rationale for the ensuing background literature. The writer should be sure that the problem is clearly and concisely stated so that readers who may not be informed about the problem get a true conceptual notion about the problem. When stating the problem, the author should cite those sources that have direct bearing on it and follow up with an in-depth review of the literature in a later section. This part of the introduction should begin with broad statements that become more specific until the study is introduced by means of stating the hypotheses or research questions.

THE RELATIONSHIP BETWEEN SELF-EFFICACY THEORY AND
EXERCISE COMPLIANCE IN A CARDIAC POPULATION

Patricia Marie Vidmar, Ph.D.
Department of Health and Safety Studies
University of Illinois at Urbana-Champaign, 1991
L. Rubinson, Advisor

This investigation utilized a cross-sectional study design in order to examine the relationship between self-efficacy theory and exercise compliance in a cardiac population. The study sample was comprised of 20 females and 118 males who had completed a Phase II cardiac rehabilitation program. Of the sample, 43 were enrolled in a Phase III program, while 77 reportedly were exercising on their own at the time of data collection. Only 18 of the respondents reportedly did not exercise at all.

Exercise compliance/behavior was assessed according to guidelines developed by the American College of Sports Medicine (ACSM, 1986). Frequency, intensity, and duration of exercise were assessed and then combined to produce the dependent variable (exercise behavior). Two measures were utilized to assess self-efficacy: the aggregation of six self-efficacy activity scales (total self-efficacy) and the aggregation of 16 perceived barriers to exercise (exercise barriers efficacy measure).

Based on ACSM (1986) guidelines, exercise compliance for this particular population was similar to that delineated in previous studies (44%) (Barnard, Guzy, Rosenberg, & O'Brien, 1983; Bengtsson, 1983; Oldridge & Jones, 1983). A positive correlation was observed between the exercise behavior measure and the exercise barriers efficacy measure ($r = 0.6727, p < 0.001$). The exercise barriers efficacy measure was found to be the most significant predictor of exercise behavior ($R^2 = 0.270, p < 0.001$), although total self-efficacy was also found to be a significant predictor of exercise behavior ($R^2 = 0.180, p < 0.005$).

Because exercise barriers efficacy was found to be the most predictive of exercise behavior, it was suggested that perceived barriers be assessed of all graduates of the formal program (Phase II) and periodically of those enrolled in Phase III. Once these barriers have been identified, the sources of self-efficacy can be employed in an effort to alter and/or negate the barrier(s). Methodologies for implementing the sources of self-efficacy were offered. Future research is recommended to assess compliance with other long-term treatment regimens (dietary changes, smoking cessation, and stress management) and their relationships with self-efficacy following completion of a Phase II program.

Figure 14.3 Abstract

Hypotheses or Research Questions

The next step in writing the research report is to formulate and state the hypotheses or **research questions**. Dependent upon the particular format you are required to use, the review of literature can precede the hypotheses or serve as a rationale and hence be stated after the review of literature. In any case, the hypotheses should be reasonable and simply stated, consistent with known facts or theories, be able to be tested, and express the relationship between two variables.

There are some studies that do not lend themselves to formulation of research hypotheses. These are generally investigations that do not have experimental and control groups but are considered experimental anyway.

Example of Research Questions

These research questions were taken from a study in which self-efficacy was used as a predictor of attrition for clients in a treatment program for poly-drug abuse:

1. Will self-efficacy scores allow prediction of likelihood of attrition for subjects of low socioeconomic status who enter an intensive program for poly-drug abuse treatment?
2. Will self-efficacy scores allow prediction of the likelihood of attrition for client subgroups of different ethnic background and gender?
3. Will subjects who complete the program successfully show significantly different self-efficacy profiles from those who drop out? (Steinhoff-Thornton, 1994, p. 8)

Limitations

Studies are usually conducted within certain boundaries, and the results cannot be extrapolated to other populations. These are the limitations. For example, in a study titled "Contraceptive Self-Efficacy in Adolescents: A Comparative Study of Male and Female Contraceptive Behavior," the authors stated:

This study focused on older adolescents. More specifically, the sample was drawn from college student volunteers. As such, results of this research cannot be generalized to the entire adolescent population. Furthermore, past research on college students' sexual behavior has shown that college students tend to have more liberal attitudes, and tend to be more permissive than others in their age group. (Rubinson & Van den Bossche, 1997, p. 32)

Definition of Terms

Some terms used in the literature and in studies may be ambiguous and thus cause confusion for the reader. To be sure that the reader does not misinterpret any terms, the writer devises a definition wherein the term provides a frame of reference for the reader. In addition, the variables that are being studied should be defined in operational terms. The following definition of terms was taken from a study that utilized self-efficacy theory to examine areas of vulnerability for runaways:

Runaway. A child or youth, ages 10 to 17, by self-report, who left home without parental permission.

Throwaway. A child or youth, ages 10 to 17, by self-report, who has left home against his or her wishes after being told to leave by parents or parental figures. (Kaliski et al., 1990, p. 10)

Assumptions

There are some facts concerning the study that are established, but the writer cannot prove these facts. In a study titled *The Health Belief Model and Contraceptive Behavior Among College Females*, the following were listed as assumptions:

1. The design of the instrument would yield responses that were valid and reliable.
2. That self-report, as in the case of a self-administered questionnaire, is an accurate measure of actual behavior.
3. That individuals are able to project themselves into hypothetical situations and accurately assess their probable feelings and actions. (Robertson, 1983, p. 30)

Significance of the Problem

The major purpose of conducting any study is to provide knowledge and insight into a particular theory or provide relevance for practitioners. In this section, the writer presents the possible implications of what the results of the study will mean to the specific area under investigation. A discussion of how these results will be useful in solving problems and answering questions in the general field is also included (Ary et al., 2005). The significance of the problem should additionally include the applications of the results of the study to health science practitioners, if applicable. That is, if the findings of the investigation will benefit those in the field, then this should be stated. The problem's significance convinces the reader that the experiment is worthy and should be carried out by the investigator.

Review of the Literature

The literature review was discussed in detail in Chapter 3. Here we will explain, briefly, the importance of the review of literature, a suggested process for carrying out the task, and a summary of this particular section.

The review of the literature is intended to give the reader an understanding of *why* you have chosen to conduct your study. The relevant information is put together to give purpose to a question that is important, or that you may believe to be important. The writer uses mainly primary sources to support the previously stated hypotheses or research questions. The general themes to concentrate on in the writing of the review include: *What previous theories are relevant to your problem? What is your knowledge about those previous works?*

The process of conducting the search can at first look unwieldy and forbidding, especially if you have chosen a problem on which there has been much previous research. A good starting point, as was described in Chapter 3, is to conduct the search using note cards, prepare an outline of the major topics you have reviewed, and then provide order to those topics. You may have as many as 20 cards per topic, and this is where the *writing* of the review of literature becomes difficult. Too often, writers abstract each source (article, book, and so on) and just write it down, paragraph after paragraph. This is *not* a good way to write the review: it is boring and provides no insight from the writer. Some studies that are considered the classics pertaining to your topic should be described in detail. Other projects should be mentioned and grouped

together to provide a coherent, well-integrated review. There are ways to handle this in your paper. You might write, "Findings of the above studies have been largely supported by a number of other studies that have employed similar approaches" (Gall, Gall, & Borg, 2005). Then you can reference those other studies. Or you might write, "There are several other studies that support this notion (name, 1996; name, 1997; name, 2001)."

As the literature review takes shape (i.e., when the topics are well researched and referenced properly), you need to develop for the reader a sense of integration and insight into the knowledge that you have amassed. This is difficult because you must be thoroughly familiar with and have a complete understanding of the relevant literature. Interpretation of the findings of all of these studies becomes the most important part of the literature review.

The summary of the literature review should include a brief discussion of the findings and their implications for the study being proposed by the writer. These implications should indicate those areas of agreement and disagreement relevant to the problem, as well as any gaps in the existing literature.

Methodology

The methodology section includes a plan of how the study will be conducted so that the hypotheses or research questions can be ascertained. We have discussed in previous chapters the many research designs that are available for almost any type of study. The writer would choose one that best suits the hypotheses or research questions. The methodology section also includes a description of the subjects, the exact procedures utilized to collect the data, and an explanation of the instruments employed in the conduct of the study.

Subjects

The methodology usually begins with a description of the sample used in the study. Detailed description is necessary so that the reader can determine if the research sample is representative and can be generalized to other, similar populations. If the reader knows this then he or she may be able to apply the results to another study, or even replicate the study. Information such as sex, age, educational level, socioeconomic status, place of residence (urban, rural), and the like are important and should be included. In addition, a discussion of how the subjects were selected should be integrated into this section. Here, the population from which the sample was drawn should be described and methods of selection mentioned, such as randomization or matching. These methods must be very carefully detailed and include the criteria, the number of lost cases, and the effect these lost cases might have on the study. The following is an example of the subjects' section from a master's thesis:

The population defined for the purpose of this study included 412 cases from school unit superintendents and principals of middle, junior high school, and senior high school attendance centers in the state of Illinois in 1975. Only 412 questionnaires were used from the original 871 cases collected by the Illinois Office of Education because the information on some of the questionnaires could not be accurately interpreted. (Tunyavanich, 1975, p. 21)

Procedures

In the procedures section of the paper, a detailed description of the procedures is written. A synopsis of the subjects, the setting, and the variables studied begins the section. One way to organize this part is to present the procedures in chronological order. You might begin by describing the research design with enough detail so that a replication of the work could be attempted. Next, the writer gives a review of how the data were collected, again describing in detail any events that were unusual or that affected the study. A discussion of any steps taken to control or even to reduce errors (e.g., administering measures to all groups at the same time) should be delineated for the reader (Gall et al., 2005). This is necessary so that the reader will be able to reconstruct the study and perhaps avoid pitfalls that you encountered. All is done in the name of progress! If the procedures were complicated and had many variables, groups, and data collection methodologies, you can end this section with a one-paragraph summary, as the following example demonstrates:

Procedure

The instructor of each section of the sex education and family life course administered the questionnaire on the first day of class to the students. The course was held on week-day evenings in the meeting rooms of university housing. The instructors were asked to have the students complete the questionnaire at the very beginning of the first class, prior to any instruction or introduction of materials for the course. Instructors were also asked to follow a guideline given to them when administering the questionnaire. Participation in completing the questionnaire was voluntary and anonymous. Students were asked to place the questionnaire, whether completed or not, in an envelope located in the center of the room. The last student was requested to seal the envelope. The sealed envelope was then returned to the mailbox of the investigator. (Harmata, 1980, p. 56) (See Appendix B [of thesis].)

Instruments

In a research project, independent variables are manipulated and then studied for their relationship with the dependent variables. In order to accomplish this task, instruments or measures are used to assess achievement, behavior change, attitude, or some other construct. In this section of the report, a detailed description of instruments used to collect the data is given. If the instruments are standardized tests, then the description should be brief and include a description of the scores, a review of reliability and validity measures, a mention of each variable that was measured by the scores, and a statement of the relationship of the measure to the hypotheses or research questions (Gall et al., 2005).

Instruments can also be new or adapted from other **standardized measures**. If this is the case with your study, then the instrumentation section of the paper must be more detailed. Here the first construction phase must be explained. Types of items used are shown as examples; reliability and validity scores that have been obtained in pilot testing are revealed; and an explanation of the way the measure was constructed is offered. Finally, you should include, by example, the way in which the measure was scored. Usually, a copy of the instrument along with a key or instructions for scoring are included in the appendix of the paper or report.

Presentation and Analysis of Results

The presentation and analysis of the results of a study may be written together or separately. This depends upon format style and the type of paper, report, or thesis. In articles, again dependent upon the journal's format, the results and analysis (sometimes called discussion) sections are combined. In longer, more complicated studies, the results may be separated from the analysis, although the integration of the two is necessary throughout either section.

Introduction. At the beginning of the results and analysis section, the writer should present evidence that the study procedures actually tested the stated hypotheses or research questions. If, for example, you sent out a survey and obtained a low response rate, this may have influenced the results of the study. This fact must be mentioned in the results and analysis section. Also, if you had experimental and control groups, you should be sure to indicate their homogeneity. If the groups did differ, then explain the procedures you took to deal with these differences. The next paragraph or so should explain the method of data analysis (the coding procedures, how the raw scores were converted, combining responses, patterns of response, and so on). Included here should also be a discussion of the statistical analyses used. If you used ordinary or often-used methods, then the description should be brief, as in the case of analysis of variance. However, if you used more complicated and unusual analyses, then you need to fully explain the analyses and give a rationale for their use. This is accomplished by citing a source for the reader.

At the end of this introduction section, it is recommended that you tell the reader how the results and analysis will be presented. One good method of organizing this section is to use the hypotheses or research questions as organizers. Each hypothesis is restated, one at a time, and the results and subsequent analysis regarding that hypothesis are discussed. This is an easy method to use, but certainly not the only one. You should discuss this matter with an advisor or someone who has had experience in writing reports or theses; their insights will prove helpful to you.

Presenting the Findings. When presenting the findings, you must be careful not to present so many numbers and tables that your results section is ignored. One method is to state the basic finding (related to a hypothesis or research questions) first and then work to the more specific findings. The result should be communicated with words first, then with numbers and statistics. It is also a good idea to provide brief summaries throughout the section because this maintains coherence and clarity for the reader.

Presenting the Analysis (or Discussion). This section on presenting the analysis has three main components: interpretation of the findings, implications of those findings, and application of the findings to practice. If this section is centered on the hypotheses or research questions and is combined with the results section, you do not have to restate the questions or hypotheses. You can begin by accepting or rejecting the hypotheses or giving the answers to the research questions. Here the writer should make some inferences from the findings and interpret the results in relationship to the theory and other research. This section is where the writer compares his or her results to other studies and states the possible flaws in the research. Reasonable explanations are expected; keep them brief and sensible. If, on the other hand, you might

have developed a new theory, or link to a new or even old theory, you should explain that phenomenon as well. Again, even if this discussion is the most exciting event in your life, keep it brief and to the point.

The second part of this section deals with the implications that the results may have for the general field encompassed within your study. For example, you may have found that self-efficacy was a major predictor of smoking behavior among adolescents, but what are the implications for the health science field? Is self-efficacy, as a part of social learning theory, a good mechanism for determining program interventions in antismoking campaigns? These implications should suggest additions to theories and further research that may follow from the present study.

The final section of presenting the analysis should attempt to illustrate how these findings can be used by the practitioners in the field. Will these results provide a new administrative style, teaching method, or community organizer?

Conclusions and Summary

The last section of the paper or the final chapter of a thesis or dissertation is probably the one most often read. It provides a brief review of the investigation and clearly discusses the conclusions reached by the investigator. Many journals no longer require a summary because the material is covered in the abstract. Nonetheless, research papers, reports, theses, and dissertations still require, for the most part, final conclusions and a summary.

Conclusions. In the conclusions, the investigator should indicate whether the findings support or do not support the hypotheses or what the findings mean in relationship to the research questions. The conclusions are actually the major *inferences* of the study based on the results of the experiment. For example, in one of our case studies, Jayne found that the new health science curriculum, “No Smoke” produced more changes in students’ behavior regarding smoking than the old health science curriculum, “Smoking’s Bad.” This is an observed result based upon measurement scores and observations of students. Jayne could then conclude by inference that the “No Smoke” curriculum is more effective in changing smoking behavior than the “Smoking’s Bad” curriculum. This section may also include recommendations for further research and a discussion of new research questions that may have arisen.

Summary. The summary is usually the last section of the report and should briefly restate the problem, describe the procedures, and discuss the principal findings. The writer must be sure not to add anything new here because it is an account of what has already been written in the report.

References

All material that was cited in your report or article must appear in the reference list. The list begins with the word *References* centered at the top of the page. The works are arranged alphabetically by author’s last name. The style manual that your institution has selected will provide exact details as to how to prepare the reference list. In this textbook,

we have used the American Psychological Association's style manual, but there are others you can utilize, including the following:

Slade, C. (2003). *Form and style* (15th ed.). Boston: Houghton Mifflin.
The Chicago manual of style. (2003). 15th rev. ed. Chicago: University of Chicago Press.

Appendixes

The appendixes include materials that were not appropriate to be included in the body of the paper but will be useful to the reader. These materials usually include copies of instruments, keys to those instruments, raw data scores, instructions to subjects, letters of support, and long tables or printouts of secondary data analyses.

The appendix is usually noted by having a piece of paper precede it with the word *APPENDIX*, capitalized and centered. The first page of the first appendix is titled *Appendix A*, centered at the top, and is numbered consecutively from the last pages of the manuscript. Each subsequent piece of material constitutes an appendix and is designated by *B*, *C*, *D*, and so on. Make sure to check your specific style requirements because they may differ from our instructions.

Writing Style

The report, article, thesis, or dissertation need not be pedantic and dull, as so many have been. You should keep your writing to a minimum; it should be clear, concise, simple, and coherent. The main points to strive for when writing are clarity and accuracy (Judd, Smith, & Kidder, 1991). Anything you add to enhance the text, such as humor, is acceptable if it is done in a professional manner. Many new writers want to make their reports flowery, including similes and alliterations. This type of writing style should be avoided in the writing of scientific papers.

Voice and Tense

Because scientific writing is supposed to be objective, writers have traditionally used the impersonal form of expression. For example, they say "the investigator observed" in place of "I observed." The good news is that many journals and style manuals for thesis and dissertations have relaxed this rule to enable the writer to have a more personal style. For example, writers may say "I instructed the data collectors" instead of "The data collectors were instructed." Though personal pronouns are now acceptable, do not overuse them in your report. The term *we* should be used if "we" really collaborated on the study. You may also use *we* occasionally when referring to yourself and the reader, as in "We can see from Table 16. . . ."

The experiment must have been completed if you are now writing the report. Therefore, utilize the past tense throughout the paper. Be especially careful in the review of literature section, where there is a tendency to write "Smith notes" instead of "Smith noted." The studies have been completed, therefore they should be reported in the past tense (Slade, 2003). There are occasions within the report where you can use the present

tense, such as “Table 16 shows the relationship.” In addition, when discussing the implications of your study you might write, “The data from the present investigation suggest.”

Nonsexist Terminology

We are supposed to be an enlightened society and free from stereotypes. To reinforce this idea, writers are not to use sexist language because it might convey attitudes and gender roles that perpetuate stereotypes. Therefore, in 1977 the American Psychological Association published guidelines for nonsexist journal writing. The manual does provide ways to use alternatives, such as the following:

Improper: The health scientist is the best judge of his attitude toward smoking.

Proper: Health scientists are the best judges of their attitudes toward smoking.

In this example the use of the plural was used as an alternative to sexist language. At times, you might have to refer to an individual; this may be accomplished by using “he or she” or “him or her,” if it is not done too often.

When describing your subjects, however, you must report numbers of males and females. This is done by using either male or female pronouns. Attempt to find gender-neutral terms—*flight attendant* rather than *stewardess*, *parenting* rather than *mothering*, and so on—to describe persons in your study. In addition, we should all attempt to avoid gender-role stereotyping when using examples (Judd et al., 1991). Not all physicians are male nor all dental hygienists female.

Rewriting the Report

When you first begin writing research reports, it will seem an arduous and impossible task. But, as with anything else, the harder one tackles the problem, the less insurmountable that task becomes. This is very true with writing. As you write, you will find it helpful to have peers or faculty members review the drafts. Usually their comments will prove to be helpful, and you will learn something about the necessity for rewriting the original draft. Most if not nearly all writers never submit a first draft but rewrite various portions of a manuscript several times. This may seem tiresome—and it is—but it does produce good results.

The first draft is often written as quickly as possible, so that time and energy are saved for subsequent rewrites. You may realize after reviewing the draft (using the input of others as well) that the report needs more literature review or data analysis to support your arguments. Again, this is a difficult process. You may have to almost begin again, but you will be much more satisfied with the results of your extraordinary efforts.

SUMMARY

The research report, term paper, journal article, thesis, or dissertation is really a method of communication. As a health scientist, you have the obligation to report the results of your investigation in a clear and concise manner. To accomplish this we have suggested that you first check with your local authority as to the type of style or manual that you

should utilize. The second step in preparing a manuscript for a journal article is to research that journal's specific style requirements.

A research report consists of several parts: introduction, review of the literature, methodology, presentation and analysis of the data, and summary and conclusions. Each section has several subheadings, which require attention to detail, clear and concise writing, and logical interpretations of the results of your study.

The writing style you grow into and eventually adopt will depend upon what you read and how often you read the literature, journal articles, theses, and dissertations. The style should be free of clichés, personal or sarcastic remarks, asides, and sexist language. Having others review your work before submitting it will be extremely helpful to you in preparing the manuscript. The finished product, with your name affixed to it, is well worth all the effort!

CRITICAL THINKING QUESTIONS

1. Why and how is your report considered a communication document?
2. Identify and describe the sections you will utilize to write your report.
3. Discuss the importance of copyrighting your materials.
4. Differentiate between limitations and assumptions.
5. When preparing your report, what sections will be the most difficult to write? Why?

SUGGESTED ACTIVITIES

1. Determine the style manual your college or university requires you to utilize.
2. Prepare a detailed outline of a research report you are going to submit this semester.
3. Differentiate between hypotheses and research questions. Give examples of each.
4. Utilizing any Web-based library access program, identify five style sources that you would incorporate into writing a research report. Explain how you would use these materials.
5. Access a thesis or dissertation via your computer, and critique the style that the writer utilized. How could you improve upon the report?

REFERENCES

- Ary, D., Jacobs, L., Razevieh, A., & Sorensen, C. (2005). *Introduction to research*. New York: Harcourt.
- Gall, J., Gall, M., & Borg, W. (2005). *Applying educational research* (5th ed.). New York: Longman.

- Harmata, J. (1980). Differentiation between actual and perceived sexual behavior among college students. Unpublished master's thesis, University of Illinois.
- Judd, C., Smith, E., & Kidder, L. (1991). *Research methods in social relations*. New York: Holt, Rinehart, and Winston.
- Kaliski, E., Rubinson, L., Lawrance, I., & Levy, S. (1990). AIDS, runaway behavior and self-efficacy. *Journal of Family and Community Health, 13*(1), 65-72.
- Robertson, N. (1983). The health belief model and contraceptive behavior among college females. Unpublished master's thesis, University of Illinois.
- Rubinson, L., & Van den Bossche, F. (1997). Contraceptive self-efficacy in adolescents: A comparative study of male and female contraceptive behaviors. *Journal of Sex Education and Therapy, 22*(2), 23-29.
- Slade, C. (2003). *Form and style* (15th ed.). Boston: Houghton Mifflin.
- Steinhoff-Thornton, R. (1994). Self-efficacy as a predictor of attrition for African American and white client populations in treatment for polydrug abuse. Unpublished doctoral dissertation, University of Illinois.
- Tunyanavich, N. (1975). The status of health instruction in Illinois public schools. Unpublished master's thesis, University of Illinois.

Common Statistical Procedures

The procedures are listed alphabetically.

Procedure	Common Usage
Analysis of covariance (ANCOVA)	Describes the relationship (difference) between a continuous dependent variable and one or more nominal independent variables (like ANOVA), but also controls for the effect of one or more continuous independent variables.
Analysis of variance (ANOVA)	Used to analyze the means among a number of groups. It is a generalization of the <i>t</i> -test and is used for two or more groups. Generally, the dependent variable is interval level and continuous, and the independent variable(s) are nominal. Assumes homogeneity within group variation and normal distribution. When groups can be classified in a 2×2 contingency table and the response under study is a continuous normal distribution, a two-way analysis of variance is used.
Bartlett's test	This test is for homogeneity of variance over <i>k</i> groups. The <i>F</i> -test is for two groups, whereas this test is for more than two. Useful in deciding whether to use ANOVA.
Chi-square (χ^2) test	Compares the frequency count of what is theoretically expected versus what is actually observed. This test can be used to determine (1) an association between two variables, (2) homogeneity of subgroups, (3) significant difference between proportions, and (4) how well observed data fit a specific model. Yates correction factor can be used when cell numbers are ≤ 5 . Nonparametric test.
<i>F</i> -test	Used to test for the equality of two variances. If they are not significantly different, the general rule of thumb is to use a two-sample <i>t</i> -test with equal variances. If the difference is significant, use a two-sample <i>t</i> -test with unequal variances.
Factor analysis	Used to arrive at one or more composite variables (i.e., factors) from other reduced variables. All variables are classically continuous, but in practice almost all types have been used.
Fisher's exact test	Used for a two-sample problem with binominal distribution and independent samples; all values are expected to be ≤ 5 . Nonparametric test.
Friedman ANOVA	Used to determine if a significant difference exists among more than two dependent groups. Nonparametric test.

Procedure	Common Usage
Kappa statistic	For use with categorical variables to measure reproducibility. An example would be to check the reproducibility of a categorical variable in two surveys. The Kappa statistic quantifies the association.
Kruskal-Wallis test	Used to compare the means among more than two samples, when either the data are ordinal or the distribution is not normal. This is a nonparametric alternative to the one-way ANOVA. When there are only two groups, it is equivalent to the Mann-Whitney <i>U</i> -test.
Linear structural relation analysis (LISREL)	Most often used to test causal relationship, similar to path analysis, but can be used as an approach to factor analysis. This analysis is generally better than path analysis because it does not have to meet so many assumptions.
Mann-Whitney <i>U</i> -test	This is the ordinal equivalent of the <i>t</i> -test (difference between means) and is generally equal to or greater in power than the <i>t</i> -test. This test is equivalent to the Wilcoxon rank-sum test since the same <i>p</i> -value is obtained by applying either test.
Mantel-Haenszel test	In epidemiology, used to assess the association between a dichotomous disease and a dichotomous exposure when confounding is present. Used for estimating the common odds ratio for stratified data.
McNemar test	Used for a two-sample problem wherein the distribution is binominal and the samples are related. In short, this is a two-sample test for binominal proportions for matched-pair data. Nonparametric test.
Multiple logistic regression (logit analysis)	Used to analyze relationships between multiple independent variables and a dependent variable that is nominal (i.e., categorical). The independent variables are classically continuous, but in practice a mixture of variables has been used. Logistic regression can be used to generate odds ratios.
Multiple regression	Examines the strength, direction, and extent of a relationship between a continuous dependent variable and several independent variables. The latter are usually, but not always, continuous. This is an extension of simple linear regression.
Multivariate analysis of variance (MANOVA)	Used to test for prespecified relationships between multiple independent variables and two or more dependent variables. Dependent variables should be at the interval level; interactions are permitted. Tests of significance for MANOVA are usually Wilk's lambda, Roy's greatest root criterion, or Pillai-Bartlett <i>V</i> -tests.
Newman-Keuls multiple comparison	Used to determine which of <i>k</i> means in a one-way ANOVA are significantly different. Newman-Keuls is preferred when pairs of means are contrasted. Multiple comparison procedures may be better than <i>t</i> -tests if there are many groups and not all comparisons between groups have been well thought out in advance.
Path analysis	Regression-based method applied to a specific model based on theory and prior knowledge. For example, it can be used to test relationships based on the Health Belief Model.
Pearson correlation	This correlation is used to show the relationship between two variables (<i>X</i> and <i>Y</i>) that have an underlying normal distribution.

Procedure	Common Usage
Scheffé multiple comparison	Used instead of Newman-Keuls for multiple comparisons in a one-way ANOVA when linear contrasts are more complex than simple contrasts of pairs of means. As with Newman-Keuls, this test helps protect from declaring too many significant differences. This method is preferred for multiple comparisons from a one-way ANOVA when linear contrasts are compared (rather than means).
Simple linear regression	Looks at the linear relationship between two variables and is used to predict the dependent variable (Y) from the independent variable (X).
Spearman rank correlation	This is a correlation coefficient based on ranks. It shows the association between two variables (X and Y), which are not normally distributed.
t -test	With one sample it can be used to test the sample mean against some specified value. With two samples, the t -test can be used to compare means. The samples can be independent or paired; if used for the latter, it is sometimes called a dependent t -test. The formula will vary with the nature of the test. Normal distributions are assumed. The test can be one-tailed or two-tailed. Parametric test.
Wilcoxon matched-pairs test	The most common nonparametric test for the two-sampled repeated measures design. Assumes an ordinal dependent variable, ordinal, continuous distribution, and random sampling. This test is sometimes called the Wilcoxon signed-rank test.
Wilcoxon rank-sum test	Used with two independent samples and is based on ranks. Similar to a t -test, except no assumptions as to normal distribution or equal variances are necessary. This test is equivalent to the Mann-Whitney U -test.
Wilcoxon signed-rank test for ordinal data	Used with two dependent samples, using differences between the individual pairs of observations. Similar to a paired t -test except for nonparametric data.

The following references were used for this table and provide further explanations.

- Cleary, P. D., & Angel, R. (1984). The analysis of relationships involving dichotomous dependent variables. *Journal of Health and Social Behavior*, 25, 334–348.
- Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis*. New York: Wiley.
- Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore: Johns Hopkins University Press.
- Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariate methods* (2d ed.). Boston: PWS-Kent.
- Kuzma, J. (1984). *Basic statistics for the health sciences*. Palo Alto, CA: Mayfield.
- Morrison, D. F. (1976). *Multivariate statistical methods* (2d ed.). New York: McGraw-Hill.
- Rosner, B. (1990). *Fundamentals of biostatistics* (3d ed.). Boston: PWS-Kent.

World Wide Web Research

The World Wide Web (www) offers innumerable sites for research activity. Some government sites provide basic health information about disease and wellness, while others offer databases that can be used in research efforts. There are statistical sites that can analyze your data and assist in determining sample size for your study. More recently, sites have been added that address evidenced-based health care for use by researchers, educators, and policymakers. From a recreational standpoint, some sites provide humor made especially for researchers. Of course there are many more sites than those listed below, and you are encouraged to use your browser to seek other sites. If you find an interesting link in any of the sites in this appendix, you are encouraged to follow it.

Statistical Sites

Title	Address
Fed Stats—The gateway to statistics from over 100 U.S. federal agencies.	http://www.fedstats.gov/
SAS Institute Main Entry Point—Offers information on the newer wireless techniques.	http://www.sas.com/
SPSS Entry Point	http://www.spss.com/
Mathematica—Provides statistical packages for students and teachers, with special editions for elementary and secondary school teachers.	http://www.wri.com/
Minitab Statistical Packages	http://www.minitab.com/
K-12 Statistics—Provides projects and problems to teach statistical concepts.	http://www.mste.uiuc.edu/stat/
Electronic Statistics—The Electronic Textbook—Begins with an overview of the relevant elementary concepts and progresses to a more in-depth exploration of specific areas of statistics (ANCOVA/MANCOVA, discriminant analysis, nonparametric statistics, distribution fitting, factor analysis, multidimensional scaling, reliability/item analysis, cluster analysis, log-linear analysis, nonlinear estimation, canonical analysis, survival analysis, time series analysis, structural equation modeling, etc.). Provides a glossary of statistical terms and a list of references for further study.	http://www.statsoft.com/textbook/stathome.htm

Evidence-Based Health Care Sites

Title	Address
Centre for Evidence-Based Medicine—National Health Service in Great Britain—Has links to several sites and a PowerPoint presentation about EBM.	www.cebm.net
National Institute of Child Health and Development Cochrane Collection on Neonatal Health	http://www.nichd.nih.gov/cochraneneonatal/
The Cochrane Collaboration—Provides systematic reviews.	www.cochrane.org/reviews
The Health Information Research Unit (HIRU) at McMaster University—Conducts research in health information science.	http://hiru.mcmaster.ca
Centres for Health Evidence-Canada—Focuses on presenting and disseminating health knowledge in ways that facilitate its optimum use.	http://www.cche.net/
Centre for Evidence-based Mental Health	http://www.psychiatry.ox.ac.uk/cebmh/index.html
Herbal Medicine—Provides hyperlinked access to the scientific data underlying the use of herbs for health. It is somewhat of an evidence-based information resource.	http://www.herbmed.org

Government Sites

Title	Address
Centres for Disease Control—Division of Adolescent and School Health—Offers YRBSS data and much more.	http://www.cdc.gov/nccdphp/dash/
Centres for Disease Control—Offers a variety of information for research; check the data and statistics section for scientific, surveillance, laboratory, and health statistics.	http://www.cdc.gov/
National Centre for Health Statistics—Offers data on surveys and collection systems (NHIS, NHANES, NIS, NSFG, NHCS, SLAITS, trend data, state data).	http://www.cdc.gov/nchswww/
Statistical Export and Tabulation System—Provides tools to handle large data sets on your PC; check the NCHS, as well as other CDC sites for available data sets.	http://www.cdc.gov/nchs/sets.htm
NCHS Data Warehouse—Provides both tabulated and microdata at national and state levels.	http://www.cdc.gov/nchs/datawh.htm
CDC National Prevention and Information Network—Provides data on HIV, AIDS, and TB.	http://www.cdcnpin.org/
searchgov.com—Provides links to executive, independent, and state and local agencies.	http://www.searchgov.com/

(Continued)

Government Sites (*Cont.*)

Title	Address
U.S. Department of Education—Prescribes funding opportunities and provides other educational information.	http://www.ed.gov/
National Cancer Institute	http://www.nci.nih.gov/
Department of Health and Human Sources Grants Net—Offers an electronic road map to grants.	http://www.hhs.gov/grantsnet/
National Institutes of Health—Leads to information about NIH grant and fellowship programs as well as research contracts.	http://grants.nih.gov/grants/index.cfm
Substance Abuse & Mental Health Services Administration—Provides grant information.	http://www.samhsa.gov/
Agency for Healthcare Research and Quality—Provides information on funding, research findings, and fact sheets.	http://www.ahrq.gov/fund/grantix.htm

Name Index

A

Abdelnoor, M., 219
Aday, L. A., 100, 101, 144–146
American Psychological
Association, 277,
285–286, 308
Anderson, L., 171
Anderson, R. M., 125
Anderson, S., 198, 201, 205
Annemans, L., 125
Ary, D., 3, 4, 223, 302

B

Bailey, K., 112, 169
Bailey, K. D., 14, 64
Baker, F., 199–200
Ball, S., 198, 201, 205
Baranowski, T., 183
Barriball, K. L., 106
Beattie, B. L., 57
Beck, C. T., 142
Becker, H., 173
Becker, M., 5
Beecher, H. K., 54, 65
Bennett, J. H., 37
Berg, B. L., 107–108, 161, 163
Berg, K. E., 263, 268
Best, J., 6
Bettcher, D. W., 125
Bickman, L., 159–160
Biklen, S., 153, 158, 171,
173–174
Bilodeau, B. A., 120
Birch, S., 90
Bloom, B., 188
Bodgan, R., 153, 158, 171,
173–174, 176
Bogdewic, S., 167

Bohnenblust, S. E., 247, 249, 260
Boissel, J. P., 90
Borg, W., 162, 269, 303
Bowling, A., 135
Box, G., 92
Boyd, R., 142, 183
Bracht, G., 82–83
Brandt, A., 51
Branowski(Windsor), 183
Brewer, M., 182
Britan, G., 210
Browner, W. S., 142, 218
Buchler, J., 2
Bullmore, E., 272
Burns, N., 107

C

Campbell, D., 79, 82, 88, 92, 187
Capron, A., 58
Castellan, N. J., 248
Champion, V., 15
Chave, E., 119–120
Chelimsky, E., 183
Chen, X. K., 217
Chok, B., 217
Christiansen, C. L., 217
Clark, H., 82
Clark, N., 142, 183
Cloutier, P., 100
Cohen, L., 2
Cole, D., 57
Committee on Research
Participation, 66
Conant, J., 3
Conklin, K. A., 108
Cook, D. J., 38
Cook, S., 105
Cook, T., 92, 187

Cooper, H., 272
Cooper, R., 207–208
Copernicus, 7
Cornelius, L. J., 144–145
Cottrell, R. R., 56, 135, 142–143
Cristol, D. S., 106
Cummings, S. R., 142, 218

D

Dalcy, F. M., 110
Dalkey, F. M., 125
Daly, J., 272
Davison, B. J., 120
Degner, L. F., 120
Denzin, N., 164
Department of Health and
Human Services, 52, 57
Deutch, M., 105
Dolan, M. S., 39
Donald, A., 37–38
DuBois, J. M., 68
Duffy, M., 88–90

E

Ebbutt, D., 124
Eisner, E., 193–194
Eleazar, K., 38
Ellison-Wright, I., 272
Ellison-Wright, Z., 272

F

Fagerline, A., 100
Fan, X., 112
Fanning, E., 122
Farewell, V., 224
Fendrich, M., 114
Ferguson, L., 120
Fernandez, C. V., 64

Fleiss, J. L., 228, 230
 Fletcher, R. H., 217
 Forthofer, R. N., 145
 Fowler, F. J., 108
 Freeman, H. E., 183
 Frey, J. H., 110
 Fricker, S., 109
 Furlong, D. K., 110

G

Galesic, M., 109
 Gall, J., 162, 303
 Gall, M., 162, 269, 303
 Gallagher, P. M., 108
 Garcia, R., 125
 Garfinkel, Harold, 158, 167–168
 Giacomini, M. K., 38
 Ginsberg, D., 89
 Glaccone, G., 125
 Glaser, B., 156
 Glass, G., 82–83
 Glesne, 164
 Gliksmann, M., 272
 Goddard, C. E., 125
 Goldzicher, 59–60
 Gomez, E., 217
 Goodman, R., 143, 183
 Goon, E. H., 125
 Grady, C., 64
 Grady, D. G., 142
 Grant, P., 106
 Graves, R. M., 108
 Graziano, A. M., 75, 78, 80–81,
 92, 142
 Green, L., 15
 Green, S., 100
 Greene, J., 109–110
 Gribben, S., 52
 Grove, S. K., 107
 Guba, E., 204
 Guyatt, G. H., 38
 Guzman, J. R., 125

H

Hack, S., 82
 Hahn, E. J., 125
 Hahn, S., 90
 Haight, B. K., 87–88
 Halfens, R., 119
 Hallam, S., 82

Harless, J., 198
 Harmata, J., 304
 Hartman, D., 163
 Hasman, A., 119
 Hays, W. L., 248
 Hedges, L. V., 272
 Heflich, D., 110
 Heise, D., 121
 Helmer, O., 124–125
 Hendrix, S., 87–88
 Herbst, A. L., 218
 Hornig, S., 64
 House, E., 189–191, 195–196
 Huberman, A., 177–178
 Hughes, J., 110
 Hulley, S. B., 143–144, 217,
 218, 271–272
 Hungler, B., 270
 Hurlburt, R. T., 143
 Hurwitz, E. S., 220
 Huss, S. N., 87–88
 Huss, W., 207–208

I

Imbos, T., 119
 Ingelfinger, F. J., 65
 Ireson, J., 82
 Isaac, S., 106

J

Jacelon, C. S., 57
 Jadad, A. R., 90
 Johada, M., 105
 James, S., 15
 Jamieson, D. J., 217
 Jansson, M., 100
 Jenkins, G., 92
 Jenkins, J., 120–121
 Jennings, E., 57
 Johnson, G., 100, 114
 Judd, C., 307–308
 Junker, B., 164

K

Kahn, J. V., 6
 Kahn, R. I., 108
 Kaliski, F., 301
 Kalton, G., 145–146
 Kauffman, R. P., 217
 Kellehear, A., 272

Kerlinger, F., 122
 Kidder, L., 307–308
 Kittleson, M. J., 110
 Kitzinger, J., 124
 Klecka, W. R., 108
 Kleinbaum, D. G., 232
 Kodish, E., 64
 Kolodny, R., 63
 Kreuter, M., 15
 Kuzma, J. W., 247,
 249, 260

L

Landmark, K., 219
 Lang, T. A., 263, 268
 Latin, R. W., 268
 Lee, E. S., 145
 Leedy, P., 111
 Leenan, Frans, 107–108
 Lemeshow, S., 145
 Lemstra, M., 272
 Lesley, M. W., 183
 Levin, B. A., 228
 Levine, R. J., 65
 Levy, P. S., 145
 Likert, Rensis, 117–120
 Lincoln, Y., 204
 Lofland, J., 171
 Lofland, L., 171
 Lohr, K., 38
 Luong, A., 106

M

Mager, R., 188
 Manion, L., 2
 Mann, H., 60
 Matthews, D. E., 224
 Mauskopf, J., 38
 Mavis, B. E., 110
 McCartney, J. J., 55
 McCormack-Brown, K. R., 110
 McCoy, C. A., 125
 McDermott, R. J., 110
 McKenzie, J. F., 56, 135,
 142–143
 McMahan, L. F., 100
 Mead, George Herbert, 158
 Meha, R., 110
 Michael, W., 106
 Michel, Y., 87

Miles, M., 177-179
 Mortimer, R., 80
 Mortimore, P., 82
 Mueller, B., 120
 Mullins, C., 283
 Murphy, G., 120
 Murphy, R. & Associates, 201

N

National Center for Education
 Statistics, 109
 National Health Interview
 Survey, 112
 Nederend, S., 108
 Nesbary, D. K., 110
 Neutens, J., 120
 Newman, T. B., 142,
 218, 220
 Newman-Kuels, 267
 Niccolai, L. J., 222
 Nixon, M. K., 100, 108
 Norvell, S., 52
 Norwood, S. L., 142
 Nunnally, J., 122

O

O'Cleirigh, C., 272
 O'Connor, K. G., 105
 Olson, L., 105
 Osgood, C., 120-122
 Oxman, A. D., 38

P

Paik, M. C., 228
 Pakieser, R., 110
 Palmer, R., 209
 Pappworth, M. H., 58
 Paskanzer (Herbst), 218
 Patrick, J. H., 112
 Patton, M., 159, 161,
 169, 172, 192
 Pealer, L., 110
 Peters, A. S., 217
 Peterson, F., 87-88
 Peterson, H. B., 232
 Petitti, D., 272
 Pittman, K., 124
 Polit, D., 142, 270
 Popham, W., 188
 Poskanzer, D. C., 218

Powelson, R., 53
 Pruchno, R. A., 112
 Puffer, S., 90

R

Raulin, M. L., 75, 78, 80-81,
 92, 143
 Reimer, M. A., 57
 Rice, M., 110
 Rich, K., 88-89
 Richardson, W., 183-184
 Riis, J., 173
 Ritchie, M., 87-88
 Robertson, N., 302
 Rockett, I. R. H., 215, 220, 227
 Rogelberg, S. G., 106
 Rose, M. S., 112
 Rosenberg, W., 37-38
 Rosner, B., 225, 242
 Rossi, P. H., 183, 190
 Rothrock, J., 209
 Rubin, H. J., 108
 Rubinson, L., 18, 80, 87-88,
 284, 300-301
 Russell, C. L., 100
 Russell (Jenkins), 120
 Rutman, L., 184

S

Sackett, D. L., 38
 Sacks, H., 168
 Safren, S., 272
 Sapirie, S., 125
 Scheffé (Newman-Keuls), 267
 Schlesselman, J. J., 220,
 222-223, 231
 Schonlan, M., 109
 Scott, C., 122
 Scriven, M., 192-193, 199
 Secic, M., 268
 Sederburg, M. E., 106
 Sellitz, C., 105
 Sharp, S., 105
 Shaughnessy, A., 37
 Shortell, S., 183
 Siegel, S., 248
 Silva, R. R., 90
 Sim, J., 124
 Simon, S. R., 217
 Singer, E., 60

Sivadas, E., 110
 Sjoberg, G., 64
 Slade, C., 277-278, 307
 Slaughter, S., 57
 Slawson, D. S., 37
 Sloan, J. A., 120
 Smith, A., 173
 Smith, E., 307
 Snider, J., 122
 Snow, D., 171
 Snow, R., 84
 Sobel, Dava, 62-63
 Sorokin, W., 121
 Sowell, R. L., 125
 Speizer, H., 109-110
 Stake, R., 195, 202-203
 Stanley, J., 79, 82, 88, 92
 Steinhoff-Thornton, L., 18, 301
 Stone, D., 80
 Stott, W., 173
 Strain, R., 209
 Strauss, A., 156
 Strecher, C., 15
 Streuning, E., 182
 Stufflebeam, D., 191
 Suchman, E., 188-189
 Suci, G., 120-121
 Sully, P., 106

T

Tannenbaum, P., 120-121
 Tayens, M. K., 125
 Taylor, 188
 Taylor, A. W., 108
 Taylor, S., 176
 Thomas, W., 163, 172
 Thompson, E., 209
 Thomson, J., 173
 Thorne, B., 60
 Thurstone, 119-120
 Tordella, M., 120
 Torgerson, D. J., 90
 Tourangeau, R., 109
 Tourney, C. P., 125
 Trow, M., 100
 Tuchfarber, A. J., 108
 Tukey (Newman-K), 267
 Tunyavanich, N., 303
 Turner, J., 284
 Tyler, R., 188-189

U

Ulfelder, H., 218

V

van Alphen, A., 119
Van den Bossche, F., 301
Van Gelder, F., 100
Van Vorst, B., 52
Vazquez, F., 125
Veatch, 59
Vergnenegre, A., 125
Vidmar, P., 18

W

Wagner, J., 173
Wakefield, M., 108

Walizer, M., 166
Wang, L., 112
Watson, J., 90
Watts, M., 124
Weber, Max, 157
Weijer, C., 64
Weiler, R. M., 110
Weitzman, E., 179
White, A. E., 106
Wieder, D., 167
Wiener, P., 166
Wiersma, W., 10, 122, 267
Wiitala, W., 109–110
Williams, J., 100
Wilson, A. H., 124
Wilson, D., 108

Windsor, R., 142, 183
Wold, J. L., 124
Wood, D., 163

Y

Yan, T., 109

Z

Znaniecki, F., 173

Subject Index

A

abstracts, research reports, 299–300
accreditation model, evaluation research, 195–197
acknowledgment page, 297–298
acquired immunodeficiency syndrome (AIDS),
inferential analysis case study, 256–257
additional costs, ethics in human research and,
58–60
alpha errors, sample size, 142
alternative hypotheses, testing of, 259–263
ambiguous terminology, statistical records,
175–176
analysis of covariance (ANCOVA)
analytical epidemiology, 223
experimental research, 77, 311
inferential analysis, 267–268
nonequivalent control group design, 94
analysis of variance (ANOVA)
experimental research, 93, 311
inferential data analysis, 266–268
table numbers, 281–282
analytical studies
epidemiological studies
basic principles, 215
case-control studies, 218–220, 229–230
case studies, 214–215, 232–233
causation, 220
cohort investigations, 217–218, 226–229
confidence intervals, 230–232
controls, 222–224
error problems, 220–222
incidence, 225–226
methodologies, 215–217
prevalence, 224
results analysis, 224–232
report guidelines, 305–306
sample size, 142
survey research, 103–104

anecdotes, qualitative research, 160
antecedent conditions, evaluation research, 187
appendixes, in research reports, 307
applied research, defined, 5–6
art criticism, evaluation research, 194
assumptions
problem statement, 19
in research reports, 302
attitude scale, construction, 117–122
attributable risk rate, cohort studies, 227–228
attribution, evaluation research, 187
audio recordings, privacy and confidentiality and,
61–62
autoregressive integrated moving average (ARIMA)
model, interrupted time series design,
92–93

B

background information, literature review and,
31–32
Bandolier (journal), 34
bar graphs, figure presentations, 287–288
basic research, defined, 5–6
behavioral objective model, evaluation research,
187–189
Belmont Report, 56–57
beta errors, sample size, 142
bias
analytical epidemiology, 221
statistical records, 175–176
bias-free approach, evaluation research, 193
bibliography index, literature review and, 32
BIOETHICSLINE database, 33
Biological Abstracts database, 33
bivariate tables, 279–280
blind study, internal validity, 82
branching questions, defined, 116
budget, research proposal and, 15, 23

C

- captions, figure presentations, 291–292
- case-control studies
 - advantages and disadvantages, 219
 - analytical epidemiology, 214–215, 218–220, 229–230
 - defined, 215–217
 - techniques, 229–230
- case study model
 - case control studies, 214–215, 218–220, 229–230, 232–233
 - cohort studies, 214, 217–218, 226–229
 - evaluation research, 194–195
- CATCH program, 5
- causation, analytical epidemiology, 220
- census surveys, qualitative analysis, 175
- central tendency, measures of, descriptive data analysis, 240–242
- charts, figure presentations, 288
- children, ethics in human research on, 55–57
- chi-square test, 269, 311
- CINAHL database, 33
- CIPP model, evaluation research, 191–192
- circle graphs, figure presentations, 288
- citation guidelines, figure presentations, 289–290
- client identification, evaluation research, 185
- clinical trials, phases of, 88–90
- closed questionnaire, design and construction, 113–114
- cluster/area sampling, 139
- Cochrane Library*, 34
- coding patterns, qualitative data analysis, 176–178
- coefficient of variation, descriptive data analysis, 245–246
- cohort studies
 - analytic techniques, 226–229
 - case study, 214
 - defined, 215–217
 - investigative techniques, 217–218
- column heads, table presentations, 283
- column spanner, table presentations, 283
- committee on research participation (CRP), exempt/nonexempt status and, 68–69
- communication document, 295–296
- communication of the study, research reports, 299–300
- complete participants, qualitative research, 164–167
- computer-assisted telephone interviewing (CATI), survey research, 109
- computer-based graphics, 293
- computer searches
 - literature review and, 35–37
 - qualitative data analysis, 178–179
- conclusions, report guidelines, 306
- concordance, Kendall coefficient of, 271
- confidence intervals
 - analytical epidemiology, 230–232
 - parameter estimation, 258–259
- confidence levels, sample size, 144, 146
- confidentiality, ethics in research and, 60–62
- confounding variables
 - analytical epidemiology, 221
 - experimental research, 76
- connoisseurship model, evaluation research, 193–194
- construct validity
 - defined, 79
 - evaluation research, 187
- contamination, internal validity, 82
- content validity, evaluation research, 187
- contextual evaluation, evaluation research, 210
- contingency (bivariate) tables, 279–280
- continuous recording, qualitative research, 160–161
- contract, evaluation research, 186
- control group
 - analytical epidemiology, 222–224
 - experimental research, 76
- convenience sampling, 140
- Copernicus, 7
- copyright, report writing and, 297–298
- correlation
 - evaluation research, 205
 - Pearson product-moment correlation, 249–250
 - Spearman rank order correlation, 247–249
- correlation matrix-multivariable table, 280–281
- cost analysis, evaluation research, 209
- cost computation, sampling designs, 147
- Council for Education in Public Health (CEPH), evaluation research, 195–197
- counterbalance design, experimental research, 94
- critical region, hypothesis testing and, 261–263
- criticism of relevant studies, 31
- crossover trial, experimental research, 90
- cross-tabulations, 279–280
- culture, qualitative research and, 158
- current cohort study, defined, 217

D

- data analysis
 - case study, 237
 - computers and information delivery systems, 250–252
 - descriptive techniques, 240–250
 - evaluation research, 188
 - inferential analysis
 - basic principles, 257
 - case studies, 256–257, 273
 - hypotheses testing, 259–263
 - mean score comparisons, 263–268
 - meta-analysis, 271–272
 - nonparametric significance testing, 269–271
 - parameter estimation, 258–259
 - relationship and prediction, 268–269
 - measures of central tendency, 240–242
 - measures of spread or variation, 242–246
 - Pearson product-moment correlation, 249–250
 - qualitative data, 176–178
 - research problems and, 9
 - Spearman rank order correlation, 247–249
 - standard measures, 246–247
 - relationships, 247–250
 - statistics, 238–239
- Database of Abstracts of Reviews of Effectiveness (DARE) database, 34
- databases, literature review and, 33–34
- data coding, qualitative data analysis, 176–178
- data collection
 - attitude scale construction, 117–122
 - case study, 99, 126–127
 - cover letter, 122–123
 - Delphi techniques, 124–126
 - flow plan, 100–102
 - focus groups, 124
 - interval-ratio scaling, 119–120
 - Likert scaling, 117–119
 - literature review and, 32
 - mail surveys, 104–106
 - personal interviews, 106–108
 - pretesting and questionnaire revision, 122–123
 - questionnaire design and construction, 112–117
 - response rates, 110–111
 - semantic differential scaling, 120–122
 - statistical records, 175–176
 - survey design, 103–104
 - survey research, 99–100
 - survey sampling, 111–112
 - telephone interviews, 108–109
 - web-based surveys and electronic mail, 109–110
- data organization, qualitative data analysis, 178
- data presentation techniques
 - case study, 276–293
 - computer graphics, 293
 - figure presentations, 286–293
 - captions and legends, 291–292
 - charts, 288
 - citations, 289–290
 - drawings, 288–289
 - graphs, 286–288
 - photographs, 289
 - preparation guidelines, 292–293
- table format, 281–285
- table presentations, 276–281
 - contingency tables, 279–280
 - multivariable tables, 280–281
 - percentage tables, 278–279
 - purpose of, 277
 - single-variable tables, 278
 - text relevance, 277–278
 - word tables, 281
- decision makers, evaluation research, 188
- decision-making model, evaluation research, 191–192
- definition of terms, problem statement, 19
- delimitations, problem statement, 19
- Delphi technique, survey research and, 124–126
- dependent variables, defined, 75
- descriptive studies
 - data analysis
 - case study, 237–238
 - measures of central tendency, 240–242
 - measures of spread/variation, 242–246
 - Pearson product-moment correlation, 249–250
 - Spearman rank order correlation, 247–249
 - standard measures, 246–247
 - statistics, 238–239
 - qualitative research, 156
 - sample size, 143–149
 - survey research, 103–104
- dichotomous questions, defined, 115
- differential selection, internal validity, 81
- differential statistical analysis, 238
- diffusion of treatment, internal validity, 81
- dimensional sampling, 140
- directional hypotheses, testing of, 259–263
- directories, qualitative analysis, 175

- document study, qualitative research, 169–170
- double-blind studies
 - informed consent and, 59–60
 - internal validity, 82
- drawings, figure presentations, 288–289
- duration recording, qualitative research, 160
- dynamic research, defined, 3

- E**
- ecological validity, defined, 83–84
- effect size, sampling errors, 143
- elderly, ethics in human research on, 57
- electronic mail (e-mail), survey research, 109–110
- epidemiology
 - analytical studies
 - basic principles, 215
 - case-control studies, 218–220, 229–230
 - case studies, 214–215, 232–233
 - causation, 220
 - cohort investigations, 217–218, 226–229
 - confidence intervals, 230–232
 - controls, 222–224
 - error problems, 220–222
 - incidence, 225–226
 - methodologies, 215–217
 - prevalence, 224
 - results analysis, 224–232
 - defined, 215
- equivalent time sample design, experimental research, 93
- ERIC database, 33
- error problems, analytical epidemiology, 220–222
- estimated sample design effect, 146
- estimation of parameters, inferential analysis, 257–259
- ethics in research
 - case study, 50–51, 70
 - children as target groups, 55–57
 - elderly as target groups, 57
 - harmful consequences, responsibility for, 62–63
 - human research dilemmas, 51–53
 - informed consent issues
 - double-blind studies, 59–60
 - truth-telling and deception, 58–59
 - Institutional Review Board, 66–69
 - exempt/nonexempt review status, 68–69
 - review form checklist, 69
 - review process, 67–68
 - justification for human experiments, 54–55
 - level of research, 65–66
 - privacy and confidentiality rights, 60–62
 - audio and video recordings, 61–62
 - sponsored research, 64
 - successful research, duty to continue, 63
 - therapeutic and nontherapeutic research, 64
 - unethical research, publication of, 65
- ethnography, qualitative research and, 158
- ethnomethodology
 - evaluation research, 194–195
 - qualitative research and, 158, 167–169
- evaluation process, 184–185
- evaluation research
 - case study, 182
 - categories of, 197–204
 - formative evaluation, 199–200
 - needs assessment, 198
 - responsive evaluation, 202–204
 - summative evaluation, 200–202
 - clients for, 185
 - contracts for, 186
 - data analysis, 188
 - defined, 10
 - design criteria, 187
 - measurement techniques, 186–187
 - methodologies, 186, 204–210
 - case study, 206–207
 - contextual evaluation, 210
 - correlation, 205
 - cost analysis, 209
 - experimental and quasi-experimental designs, 205
 - expert judgment, 206
 - personal assessment, 206
 - simulation modeling, 207–208
 - surveys, 205
 - models, 188–197
 - accreditation model, 195–197
 - behavioral objectives model, 188–190
 - case study model, 194–195
 - connoisseurship model, 193–194
 - decision-making model, 191–192
 - goal-free model, 192–193
 - systems analysis model, 190–191
 - overview, 182–183
 - procedures, 184–185
 - purposes of, 183–186
- evaluation team, measurement by, 186–187
- evidence-based full-test and abstracting services, 34, 45

- evidence-based health care (EBHC)
 defined, 38
 web sites for, 315
- evidence-based medicine (EBM), defined, 38
- expected estimates
 proportion of eligibles, 146
 sampling design, 145
- experimental group, defined, 76
- experimental method, defined, 75
- experimental research, 10
 advantages and disadvantages, 78
 case studies, 74–75, 94–95
 characteristics of, 75
 clinical trials, 88–90
 control mechanisms, 76–77
 design criteria, 84–88
 counterbalanced design, 94
 equivalent time sample, 93
 factorial/multivariable designs, 91
 interrupted time series, 92–93
 nonequivalent control group, 93–94
 one-group pretest-posttest, 85
 one-shot case study, 84–85
 posttest-only control group, 86–87
 pretest-posttest control group, 87
 quasi-experimental designs, 91–92
 Solomon four-group design, 87–88
 static-group comparison, 85–86
 evaluation research design, 205
 external validity, 82–84
 Hawthorne effect, 77
 internal validity, 78–82
 randomized controlled trial, 88–94
 validity, 78–84
- experimental validity, defined, 81
- expert judgment, evaluation research, 206
- external communication, qualitative research, 169–170
- external validity, experimental research, 82–84
- F**
- face validity, evaluation research, 187
- factorial design, experimental research, 91
- Fernald State School ethical case study, 52–53
- field notes, qualitative research, 170–172
- field tests, evaluation research, 200
- figure presentations, 286–293
 captions and legends, 291–292
 charts, 288
 citations, 289–290
 drawings, 288–289
 graphs, 286–288
 photographs, 289
 preparation guidelines, 292–293
- flow plan, survey research, 100–102
- focus groups, survey research and, 124
- formative evaluation techniques, 199–200
- found photographs, qualitative analysis, 173–174
- F-ratio, analysis of variance, 267
- frequency-count recording, qualitative research, 161
- Friedman two-way analysis of variance, 248, 311
- G**
- generalizability, evaluation research, 187
- geometric mean, descriptive data analysis, 242
- goal-based approach, evaluation research, 188–189
- goal-free models, evaluation research, 192–193
- government documents, literature review and, 34–35
- government web sites, 315–316
- graphs, figure presentations, 286–288
- H**
- HAPI database, 33
- harmful consequences, ethics in research and responsibility for, 62–63
- Hawthorne effect, experimental research, 77–78
- Health Belief Model, science theory and, 4–5, 5f
- health care-associated methicillin-resistant *Staphylococcus aureus* (HA-MRSA), inferential analysis case study, 256
- health risk appraisal, evaluation research, 207–208
- health sciences, research methodologies in, 10
- Healthy Cities program, 5
- Helsinki Declaration, 54–55
- historical research, 10
- history effect, internal validity, 80
- homogeneous grouping, analytical epidemiology, 222–223
- human research
 ethical dilemmas in, 51–53
 justification for, 54–55
- hypotheses
 inferential analysis and testing of, 257, 259–263
 qualitative data analysis, 176–178
 research and formulation of, 9
 in research proposal, 20–21
 research reports, 300–301
 sample size and, 142

I

incidence, analytical epidemiology, 225–226
 independent sample means, t-Test of, 263–265
 independent variables, defined, 75
 indexical expressions, qualitative research, 168–169
 indexing and abstracting services, literature review and, 33–34
 inductive reasoning, qualitative research and, 156
 inferential analysis
 analysis of variance, 266–268
 basic principles, 257
 case studies, 256–257, 273
 data analysis, 256–273
 hypotheses testing, 259–263
 mean score comparisons, 263–268
 meta-analysis, 271–272
 nonparametric tests of significance, 269–271
 parameter estimation, 258–259
 relationship and prediction measurements, 268–269
 statistical analysis, 238
 information bias, analytical epidemiology, 221
 information delivery systems, descriptive data analysis, 250–251
 information sources. *See also* literature review
 computer searches, 35–37
 level of evidence, 37–44
 procedures for, 30–35
 purposes of, 29–31
 relevance and validity worksheets, 40–44
 researchable problems, 16–17
 research and, 9
 research proposal and, 22
 writing guidelines for, 45–47
 informed consent
 ethics in human research and, 58–60
 IRB guidelines on, 67
 input evaluation, evaluation research, 192
 Institutional Review Board (IRB)
 ethics of research on children and, 55–56
 exempt/nonexempt status and, 68–69
 review form checklist, 69
 review process of, 67
 role in ethical research of, 66–69
 instruments
 internal validity of, 80
 in research reports, 304
 interaction effect, factorial/multivariable design, 91
 interest, research proposal, 14

internal documents, qualitative research, 169–170
 internal validity, experimental research, 78–82
 International Code of Medical Ethics, 55
 Internet surveys, survey research, 109–110
 interrupted time series design, experimental research, 92–93
 interval measurements
 descriptive data analysis, 239
 parameter estimation, 258–259
 interval-ratio scaling, defined, 119–120
 interval recording, qualitative research, 161
 intervening conditions, evaluation research, 187
 interviews
 questionnaire design and construction, 113
 structured and unstructured, 107–108
 survey research, 106–109
 telephone interviews, 108–109
 introduction, research reports, 299, 305

K

Kendall coefficient, 248
 of concordance, 271
 key word listings, literature review and, 32
 knowledge, research and, 2–3
 Kruskal-Wallis test, 271, 312

L

legends, figure presentations, 291–292
 level of evidence
 formula for, 37
 relevance, 37
 shortcuts for determining, 45
 validity, 38–39
 worksheets for relevance and validity, 40–44
 level of precision, 141–142
 level of statistical significance
 critical regions, 261–263
 hypothesis testing, 260–263
 nonparametric tests of, 269–271
 sample size, 142
 Likert scaling, basic principles, 117–119
 limitations
 problem statement, 19
 in research reports, 301
 linear correlation, descriptive data analysis, 247–249
 linear regression, inferential analysis, 268–269
 linear structural relation analysis (LISREL), analytical epidemiology, 224, 312
 line graphs, figure presentations, 286–287

literature review. *See also* information sources
 computer searches, 35–37
 level of evidence, 37–rt
 procedures for, 30–35
 purposes of, 29–31
 relevance and validity worksheets, 40–44
 researchable problems, 16–17
 research and, 9
 research proposal and, 22
 in research reports, 302–303
 writing guidelines for, 45–47
 logistic regression, analytical epidemiology, 224

M

mail surveys, data collection, 104–106
 main effect, factorial/multivariable design, 91
 management-by-objectives, evaluation research, 189
 Mann-Whitney U-test, 269, 312
 matching
 analytical epidemiology, 222
 case-control studies, 229–230
 maturation effect, internal validity, 79
 mean scores
 descriptive data analysis, 240–241
 inferential data analysis, 263–268
 dependent sample means, 265–266
 independent sample means, 263–265
 measurement techniques
 descriptive data analysis, 238–239
 evaluation research, 186–187
 research and identification of, 9
 median
 descriptive data analysis, 241–242
 tests of, 270
 MEDLINE/MEDLINEplus databases, 33
 meta-analysis, inferential analysis, 271–272
 methodology, in research reports, 303
 mixed sampling, 141
 mode, descriptive data analysis, 242
 multiple-choice questions, defined, 115
 multiple linear regression, 268–269
 multiple regression analysis
 analytical epidemiology, 223
 defined, 312
 multiple-variable tables, 278–281
 multistage cluster sampling, 139
 multivariable design, experimental research, 91
 multivariate analysis of variance (MANOVA)
 analytical epidemiology, 223–224
 defined, 312

experimental research, 93
 inferential analysis, 267–268

N

National Center for Health Statistics (NCHS),
 34–35
 National Commission for the Protection of Human
 Subjects and Behavioral Research, 55–56
 National Guidelines Clearing House (NGC), 34
 National Health and Nutrition Examination
 Survey (NHANES), 35
 National Health Care Survey (NHCS), 35
 National Health Interview Survey (NHIS), 35
 National Immunization Survey (NIS), 35
 National Library of Medicine database, 33
 National Survey of Family Growth (NSFG), 35
 National Vital Statistics System (NVSS), 35
 natural setting, qualitative research, 153
 needs assessment, evaluation research, 198
 new ideas, literature review as tool for, 30
 nominal measurements, descriptive data analysis,
 238–239
 noncontact rate, survey research, 111
 nonequivalent control group design, experimental
 research, 93–94
 nonparametric data
 descriptive data analysis, 239
 tests of significance, 269–271
 nonpersonal documents, qualitative research,
 169–170
 nonprobability sampling, defined, 135, 140–141
 nonsexist terminology, for research reports, 308
 nontherapeutic research, ethics and, 64
 nonverbal communication, qualitative research, 162
 note format, table presentations, 285
 notes, table presentations, 284–285
 null hypotheses
 independent sample means, t-Test of, 263–265
 in research proposal, 21
 testing of, 259–263
 Nuremberg Code, 51, 54–55

O

objectives and effects, evaluation research, 187
 observation, qualitative research, 158–164
 forms, 162–163
 methods, 160–162
 observer training, 163–164
 value and purposes, 159–160
 observed effect, defined, 75

- observer's comments, qualitative research, 171–172
- odds ratio, case-control studies, 229–230
- official documents, qualitative analysis, 174–175
- official statistics, qualitative analysis, 174–176
- one-group pretest-posttest design, experimental research, 85
- one-shot case study design, experimental research, 84–85
- one-tailed test of significance, 261–263
- one-way analysis of variance, 266–268
- open-ended questions, defined, 116
- open questionnaire, design and construction, 114
- operability, research proposal, 14
- ordinal measurements, descriptive data analysis, 238–239
- originality, critical tests of, 266
- outcomes
 - formative evaluation data, 199–200
 - vs.* process, in qualitative research, 156–157
- P**
- parallel group trial, experimental research, 90
- parameter, defined, 239
- parametric data, descriptive data analysis, 239
- participant observation, qualitative research, 157, 164–167
- Pearson product-moment correlation, 249–250, 312
- PedsCCM Evidence-based Journal Club, 34
- percentage tables, 278–279
- personal computers, descriptive data analysis, 250–251
- personal documents, qualitative analysis, 172–173
- personal interviews, survey research, 106–107
- personnel records
 - evaluation research, 206
 - qualitative research, 170
- person-time concept, analytical epidemiology, 225–229
- Phase III clinical trials, characteristics of, 88–89
- phenomenological approach, qualitative research, 157
- photography
 - figure presentations, 289
 - qualitative analysis, 173–174
- photo libraries, qualitative analysis, 173–174
- placebo, experimental research, 77
- point estimation, inferential analysis, 258–259
- population
 - descriptive data analysis, 239
 - sampling, 134, 145
 - validity, 82–83
- post hoc tests, analysis of variance, 267
- poststratification, analytical epidemiology, 223
- posttest-only control group design, experimental research, 86–87
- power factor, sample size, 142–143
- practical applications
 - evaluation research, 209
 - research proposal, 15
- PRECEDE model, science theory and, 4–5
- predictions, measurements of, 268–269
- predictive validity, evaluation research, 187
- preexperimental designs, experimental research, 84–86
- preliminary sources, literature review and, 33
- preordinate evaluation models, 203–204
- pretesting, questionnaire revision and, 122–123
- pretest-posttest control group design, experimental research, 87
- prevalence, analytical epidemiology, 224
- primary sources, literature review and, 32
- privacy rights
 - ethics in research and, 60–62
 - IRB guidelines on, 68
- probability level note, table presentations, 284–285
- probability sampling, defined, 135–139
- problem selection and components
 - participant observation, qualitative research, 166–167
 - researchable problems, 15–17
 - research and, 8–10
 - signification and justification of problem, 21–22
- problem statement
 - literature review and, 30
 - in research proposal, 17–18
 - research reports, 299
- procedures, in research reports, 304
- process
 - evaluation, 192
 - vs.* outcomes, in qualitative research, 156
- product evaluation, 192
- program information, evaluation research, 186–187
- program setting, qualitative research, 161

proportional hazards regression, analytical epidemiology, 224
 proportional stratified sampling, 138–139
 prospective studies. *See* cohort studies
 PsycINFO database, 33
 PubMed database, 34
 purposive sampling, 140
 pyramid of evidence, 38

Q

qualitative research
 case study, 153, 179
 characteristics of, 153, 156–157
 computerized data analysis, 178–179
 data analysis, 176–178
 data collection, 170–176
 field notes, 170–172
 official statistics, 174–176
 personal documents, 172–173
 photography, 173–174
 document analysis, 169–170
 ethnomethodology, 167–169
 observation methods, 159–164
 participant observation, 164–167
 quantitative research *vs.*, 154–156
 theoretical foundations, 157–158
 quantitative research
 qualitative research *vs.*, 154–156
 table presentation, 277
 quasi-experimental design
 basic principles, 91–92
 evaluation research, 205
 questionnaires
 attitude scales, 117–122
 cover letter, 122–123
 design and construction, 112–117
 mail surveys, 104–106
 pretesting and revision, 122–123
 sponsorship issues, 105–106
 types of questions, 114–117
 questions
 research proposal, 17–18
 research reports, 300–301
 quota sampling, 140

R

random assignment, experimental research, 76
 random digit dialing (RDD), telephone interviews, 108–109

randomized controlled trial (RCT)
 crossover design, 90
 experimental research, 88–94
 phases of, 88–89
 randomized matching, experimental research, 76
 random misclassification, 221–222
 random sampling
 simple random sampling, 136–137
 stratified random sampling, 137–139
 random variation, analytical epidemiology, 221
 range, descriptive data analysis, 243
 ranking questions, defined, 115
 rating questions, defined, 115
 ratio measurements, descriptive data analysis, 239
 reactivity, research proposal, 15
 references, in research reports, 306–307
 refusal rates, survey research, 111
 region of acceptance, hypothesis testing, 261–263
 region of rejection, hypothesis testing, 261–263
 relationships
 measures of, 247–249, 268–269
 participant observation, qualitative research, 166–167
 relative risk
 case-control studies, 229–230
 cohort studies, 226–229
 relevance, of evidence, 37, 40–44
 relevant literature
 criticism of, 31
 literature review of, 29
 relevant theory, defined, 30
 reliability, evaluation research, 187
 report writing, research problems and, 9
 representative design, experimental research, 84
 researchable problems, 15–17
 researcher characteristics, questionnaire design and construction, 113
 researcher-produced photographs, qualitative analysis, 174
 research methodology
 research proposal, 15
 world-wide web research, 314–316
 research overview
 basic and applied research, 5–6
 cyclical stages of, 8f
 defined, 1
 design criteria, 9
 in health science, 1–2
 knowledge and, 2–3

- research overview (*Continued*)
 - problem selection and methodology, 8–10
 - scientific approach to, 6–8
 - scientific theory and, 4–5
 - static and dynamic aspects of, 3
 - research procedures, in research proposal, 23
 - research proposal
 - budget considerations, 23
 - checklist for, 24–26
 - hypothesis formulation, 20–21
 - literature resources, 22
 - null hypotheses, 21
 - problem components, 19
 - problem selection, 14–15
 - problem significance and justification, 21–22
 - problem sources, 15–17
 - problem statement and research
 - questions, 17–18
 - research procedures, 23
 - subproblems, 18
 - time schedule for, 23
 - research report writing
 - abstract, 299
 - appendixes, 307
 - assumptions, 302
 - basic components, 295–296
 - conclusions and summary, 306
 - copyright, 297–298
 - definition of terms, 301
 - hypotheses or research questions, 300–301
 - instruments, 304
 - introduction, 299
 - limitations, 301
 - literature review, 302–303
 - methodology, 303
 - nonsexist terminology, 308
 - problem statement, 299–300
 - procedures, 304
 - references, 306–307
 - results analysis and presentation, 305
 - rewriting, 308
 - significance of problem, 302
 - subjects, 303
 - table of contents, 298–299
 - tables and figures, 299
 - title page, 296–297
 - voice and tense, 307–308
 - writing style, 307
 - respondent characteristics, questionnaire design
 - and construction, 113
 - response rates
 - adjustments, 146
 - survey research, 110–111
 - responsive evaluation technique, 202–203
 - responsive evaluation techniques, preordinate
 - evaluation models, 203–204
 - restricted questionnaire, design and construction, 113–114
 - results analysis and presentation, in research
 - reports, 305–306
 - retrospective cohort study, 217–218
 - rewriting guidelines, research reports, 308
 - risks *vs.* benefits, ethics in research and, 67
- S**
- sample
 - defined, 134
 - descriptive data analysis, 239
 - sampling errors, 141–149
 - sampling frame, defined, 134–135
 - sampling techniques
 - analytical epidemiology, 223
 - analytical studies, 142–143
 - case study, 133, 149–150
 - cluster/area sampling, 139
 - nonprobability sampling, 140–141
 - probability sampling, 136–139
 - purpose of sampling, 134
 - random sampling, 136–137
 - sampling frame, 134–135
 - size determination, 141–142, 147–149
 - stratified random sampling, 137–139
 - survey research, 111–112
 - surveys and descriptive studies, 143–149
 - systematic sampling, 137
 - sampling unit, defined, 135–136
 - scatter graphs, figure presentations, 288
 - scatterplots, Spearman rank order correlation, 247–249
 - schedule, research proposal time frame, 23
 - scientific approach, principles of, 6–8
 - scope, research proposal, 14
 - secondary sources, literature review and, 31–32
 - selection-maturation interaction, internal
 - validity, 81
 - semantic differential (SD) scaling, 120–122
 - semistructured interview, survey research, 107–108
 - sentence-completion question, defined, 116
 - sequence effects, internal validity, 81

- significance of the problem, in research reports, 302
 sign test, 270
 simple linear regression, 268–269, 313
 simple random sampling, 136–137
 simulation modeling, evaluation research, 207–208
 single-variable tables, 278
 snowball sampling, 141
 Social Work Abstracts, 34
 Solomon four-group design, experimental research, 87–88
 Spearman rank order correlation, 247–249, 313
 specific note, table presentations, 284
 specimen records, qualitative research, 160
 sponsored research, ethics of, 64
 spread of variation, descriptive data analysis, 242–246
 standard deviation/variance, descriptive data analysis, 243–245
 standard error formulae, sampling design, 14
 standard error of the mean, sampling design, 145
 standardized measures, in research reports, 304
 standard measures, descriptive data analysis, 246–247
 standard score, descriptive data analysis, 246–247
 statement of the problem, research proposal, 17–18
 static-group comparison design, experimental research, 85–86
 static research, defined, 3
 statistical regression, internal validity, 80–81
 statistical validity, defined, 78–79
 statistics
 common statistical procedures, 311–313
 defined, 239
 descriptive data analysis, 238–239
 hypothesis testing, 260–263
 nonparametric tests of significance, 269–271
 qualitative analysis, 174–176
 web sites for, 314
 Stenomask, field note techniques, 172
 stratified random sampling, 137–139
 analytical epidemiology, 223
 sample size and, 147–148
 structured interview, survey research, 107–108
 structured observation, qualitative research, 160
 stub columns, heads and rows, table presentations, 283
 subgroup selection, sampling design, 145
 subject characteristics
 research and identification of, 9
 in research reports, 303
 subjective approach, evaluation research, 195
 subject selection, IRB guidelines on, 67
 subproblems, research proposal, 18
 summary, report guidelines, 306
 summated rating, Likert scaling, 117–118
 summative evaluation, evaluation research, 200–202
 survey research, 10
 design criteria, 103–104
 evaluation research, 205
 flow plan, 100–102
 overview of, 99–100
 sample size, 143–149
 Sylencer, field note techniques, 172
 symbolic interaction, qualitative research, 158
 systematic sampling, 137
 systems analysis model, evaluation research, 190–191
- T**
 table numbers, 281–282
 table of contents, research reports, 298–299
 table place, 278
 table presentations, 276–281
 contingency tables, 279–280
 multivariable tables, 280–281
 percentage tables, 278–279
 purpose of, 277
 single-variable tables, 278
 text relevance, 277–278
 word tables, 281
 tables and figures, research reports, 299
 table spanner, table presentations, 284
 target groups, ethics in human research and children, 55–57
 elderly, 57
 telephone interviews, 108–109
 tense, in scientific writing, 307–308
 terminology, for research reports, 301–302, 308
 testing effect, internal validity, 80
 test statistic, computation of, 260–261
 textual relevance, table presentations, 277–278
 themes, qualitative data analysis, 176–178
 theoretical orientation, literature review writing, 46
 theoretical value, research proposal, 15
 theory, science and, 4–5
 therapeutic research, ethics and, 64
 three-way analysis of variance, 267
 Thurstone scaled value technique, interval-ratio scaling, 119–120

time frame, research proposal, 15, 23
 Title 45, Code of Federal Regulations, Part 46,
 Protection of Human Subjects, 56–57
 title page guidelines, 296–297
 tolerable error range, sampling design, 145–146
 TOXNET database, 34
 training, of observers, qualitative research,
 163–164
 true experimental designs, experimental research,
 86–88
 t-Test
 defined, 313
 inferential mean score comparisons
 dependent sample means, 265–266
 two independent sample means, 263–265
 Tuskegee Study of Untreated Syphilis in the Negro
 Male ethical case study, 51–53
 two-by-two cohort table, 226–230
 two-tailed test of significance, 261–263
 two-way analysis of variance, 267
 Type I error
 hypothesis testing and, 262–263
 sample size, 142–143
 Type II error
 hypothesis testing and, 262–263
 sample size, 142–143
 U
 U. S. Public Health Service (USPHS), human
 research ethics and, 51–53
 unethical research, publication of, 65
 unforeseeable risks, ethics in human research and,
 58–60
 unit of analysis, research proposal, 15

University of Minnesota smoking prevention
 program, 5–6
 unrestricted questionnaire, design and
 construction, 114
 unstructured interviews, survey research, 107–108
 unstructured observation, qualitative research, 160

V

validity
 evaluation research, 187
 of evidence, 38–44
 experimental research, 78–84
 values of researcher, research proposal, 15
 variables
 identification and estimates of, 144–145
 in survey research, 75–76, 91
 video recordings, privacy and confidentiality and,
 61–62
 voice, in scientific writing, 307–308
 voluntary consent, ethics in human research
 and, 54

W

web-based surveys, survey research, 109–110
 weighted distribution, sample size and, 148–149
 Wilcoxon matched-pairs signed rank test, 270, 313
 word tables, 281
 writing guidelines
 literature reviews, 45–47
 report writing, 295–308
 writing style, 307–308

Z

z-scores, descriptive data analysis, 246–247

HEALTH RESEARCH TECHNIQUES MADE EASY!

Research Techniques for the Health Sciences, Fourth Edition focuses on pragmatic aspects of health science research. Underlying concepts and theory are explained and illustrated through a common-sense approach utilizing case studies.

All of the text's case studies are health-oriented in school, community, and hospital settings involving students, teachers, physicians, nurses, and health science researchers. This edition features an emphasis on the currency in the research methods processes.

FEATURES

- **Unique case studies** used to illustrate chapter concepts provide immediate applications for you to follow and solve throughout the text, keeping you actively engaged and stressing context.
- **Ethical research** is strongly emphasized and promoted with discussions of informed consent, right to privacy, responsibility for harmful consequences, and the role of Institutional Review Boards in project approval.
- In addition to data collection and analysis, **techniques of data presentation** are addressed, offering you a variety of ways to present your research clearly, effectively, and with impact.
- New **critical thinking questions** and **suggested activities** at the end of each chapter help emphasize important concepts and encourage understanding through doing.
- **Increased discussion of technology and thoroughly updated references** keep coverage in the book current.

Please visit us at www.pearsonhighered.com for more information.
To order any of our products, contact our customer service department at
(800) 824-7799, (201) 767-5021 outside of the U.S., or visit your campus bookstore.

Benjamin Cummings
is an imprint of

PEARSON

www.pearsonhighered.com

ISBN-13: 978-0-321-59641-3
ISBN-10: 0-321-59641-2

