Ann L. Cupp Curley, PhD, RN, recently retired from her position as the Nurse Research Specialist at Capital Health in Trenton, New Jersey, where she was responsible for promoting and guiding the development of nursing research and evidence-based practice. She has an extensive background in nursing education at the undergraduate, graduate, and doctoral levels, and more than 10 years’ experience in community and public health nursing. Dr. Curley has been principal or co-principal investigator of many research projects and continues to serve as an advisor on DNP project committees and a research consultant. She received a BSN from Boston College, an MSN in community health/clinical nurse specialist track from the University of Pennsylvania, and a PhD in urban planning and policy development from Rutgers, The State University of New Jersey. Dr. Curley has received many honors, including the Nurse.com Nursing Spectrum Excellence Award for Education and Mentorship.
In Memory of Patty Vitale

Good friends are hard to find . . . and impossible to forget.

—Anonymous
CONTENTS

Contributors xi
Foreword Patricia A. Polansky, RN, MS xiii
Preface xv
Acknowledgments xxi

1. Introduction to Population-Based Nursing  1
   Ann L. Cupp Curley
   Introduction  1
   Background  2
   Defining Populations  6
   Using Data to Target Populations and Aggregates at Risk  7
   Summary  17
   Exercises and Discussion Questions  18
   References  19
   Internet Resources  22

2. Identifying Outcomes in Population-Based Nursing  25
   Alyssa Erikson and Sonda M. Oppewal
   Introduction  25
   Identifying and Defining Population Outcomes  26
   National Healthcare Objectives  37
   Summary  45
   Exercises and Discussion Questions  46
   References  47
   Internet Resources  50

3. Epidemiological Methods and Measurements in Population-Based Nursing Practice: Part I  53
   Patty A. Vitale and Ann L. Cupp Curley
   Introduction  53
   The Natural History of Disease  54
   Prevention  55
   Causation  57
   Methods of Analysis  58
   Descriptive Studies  72
4. Epidemiological Methods and Measurements in Population-Based Nursing Practice: Part II 87
Patty A. Vitale and Ann L. Cupp Curley

Introduction 87
Errors in Measurement 87
Confounding 92
Interaction 93
Randomization 94
Data Collection 95
Causality 96
Scientific Misconduct (Fraud) 97
Study Designs 98
Databases 102
Summary 104
Exercises and Discussion Questions 104
References 106
Internet Resources 107

5. Applying Evidence at the Population Level 109
Vera Kunte and Ann L. Cupp Curley

Introduction 109
Asking the Clinical Question 110
The Literature Review 114
Assessing the Evidence 117
Integration of Evidence Into Practice 128
Summary 133
Exercises and Discussion Questions 134
References 134
Internet Resources 137

6. Using Information Technology to Improve Population Outcomes 139
Laura P. Rossi and Ann L. Cupp Curley

Introduction 139
Use of the Internet to Obtain Health Information 140
Using Technology to Improve Population Health 146
E-Resources That Support Population-Based Nursing 150
7. Concepts in Program Design and Development 159
Laura P. Rossi and Ann L. Cupp Curley

Introduction 159
Sources of Data 159
Innovative Care Delivery Models 162
Program Development: Where to Start 163
Implementation 172
Overcoming Barriers and Challenges 178
Summary 180
Exercises and Discussion Questions 180
References 181
Internet Resources 182

8. Evaluation of Practice at the Population Level 183
Barbara A. Niedz

Introduction 183
Monitoring Healthcare Quality 184
Summary 210
Exercises and Discussion Questions 211
References 212
Internet Resources 216

9. The Role of Accreditation in Validating Population-Based Practice/Programs 219
Eileen M. Horton

Introduction 219
Governmental Programs 222
Non-Governmental Programs 227
Planning for Accreditation 231
Summary 233
Exercises and Discussion Questions 233
References 234
Internet Resources 235

10. Building Relationships and Engaging Communities Through Collaboration 237
Sonda M. Oppewal and Barbara A. Benjamin

Introduction 237
Foundation for Population-Focused Practice 237
11. Challenges in Program Implementation 263
Janna L. Dieckmann

Introduction 263
Lewin's Stages of Change 264
Community Engagement 265
Summary 285
Exercises and Discussion Questions 285
References 286
Internet Resources 288

12. Implications of Global Health in Population-Based Nursing 289
Lucille A. Joel and Irina McKeehan Campbell

Core Competencies in Global Health 289
Summary 314
Exercises and Discussion Questions 314
References 315
Internet Resources 317

Index 319
CONTRIBUTORS

Barbara A. Benjamin, EdD, RN, Adjunct Faculty, University of North Carolina, Chapel Hill

Irina McKeehan Campbell, PhD, MPH, Professor, Department of Health Sciences, School of Health and Human Services, National University, La Jolla, California

Ann L. Cupp Curley, PhD, RN, Research Consultant, Falmouth, Massachusetts

Janna L. Dieckmann, PhD, RN, Clinical Associate Professor, School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Alyssa Erikson, PhD, MSN, RN, Associate Professor and Chair, Department of Nursing, California State University Monterey Bay, Seaside, California

Eileen M. Horton, MSN, MSM, RN, NEA-BC, Senior Vice President, Hospital Administration and Chief Quality Officer (retired), Capital Health, Trenton, New Jersey

Lucille A. Joel, APN, EdD, FAAN, Distinguished Professor, Rutgers University School of Nursing, Newark, New Jersey

Vera Kunte, DNP, APN, Nurse Research Specialist, Capital Health, Trenton, New Jersey

Barbara A. Niedz, PhD, RN, CPHQ, Assistant Professor, Rutgers University, School of Nursing, Newark, New Jersey

Sonda M. Oppewal, PhD, RN, Professor, School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Laura P. Rossi, RN, PhD, Assistant Professor, Simmons University, College of Natural, Health and Behavioral Sciences, School of Nursing; Program Manager, Quality and Patient Safety, Department of Orthopaedic Surgery, Massachusetts General Hospital, Boston, Massachusetts

Patty A. Vitale†, MD, MPH, FAAP

†Deceased
FOREWORD

“What’s past is prologue,” wrote William Shakespeare in *The Tempest*. Why read *Population-Based Nursing: Concepts and Competencies for Advanced Practice*? In this book, the third edition, there is a lot more to know and learn. If the first two decades of the 21st century have taught us anything it is that what we know as *healthcare* is changing at warp speed. And, as the anniversary of Florence Nightingale’s 200th birthday looms large, Shakespeare’s words echo in our ears and remind us of needed changes in the practice of the profession of nursing. We inherently know that what has happened in the past sets the scene for the really important morphs yet to be identified and implemented. Those changes will determine nursing’s viability in the ever-changing national and global health marketplace.

While as in the first and second editions successful strategies that nurses have used to improve population outcomes are paramount, the readers will discover new information in the third edition on how to identify healthcare needs at the population level and how to improve overall population outcomes. *Voila.* Not only that, but introduction of the most common study designs and successful program implementation strategies will lead you to correct design selection, successful implementation and most importantly overall success. Problem solved!

The third edition charts a path toward understanding how to successfully integrate new knowledge into practice. This, as experience teaches us, is no small task. Chapter 6 actually describes how technology can be used to truly enhance population-based nursing and describes the role and importance of APRNs in using data to make decisions that lead to new levels of program development and evaluation. While Clara Barton and Florence Nightingale did not have AI they understood that statistics are needed to measure outcomes. Chapter 8 identifies ways and means to evaluate population outcomes and systems changes. These concepts and roles are explored within the competencies of the APRN. The healthcare marketplace is extremely competitive, and executives and managers are like radar screens looking to identify opportunities to distinguish and validate their organizations. This book helps the new APRN identify the ways and means to achieve such validation.

Nurses need to be part of the highest level of care management and policy decision making in partnership with healthcare policy brokers and healthcare policy makers. In Chapter 10, the emphasis is placed on identifying community needs and assessment of resources. Chapter 11 rounds up by providing specific strategies for program implementation coupled with methods to empower the community to advocate for themselves. In the final chapter global health and cultural issue for population-based
nursing theory and practice open one's eyes to recent patterns in international interdisciplinary collaborations including the latest global health competencies. A primer for all practitioners whatever the setting.

This edition targets all of the important aspects in population-based care for the most trusted and recognized of all healthcare professions. Nursing remains, and should remain, a practice centered and caring profession, but current times mandate that nurses discover new and effective strategies for promoting health and providing care. This book gives nurses everything from A to Z describing the role of the APRN in the accreditation process to zeroing in how to eliminate health disparities. By looking back to the lessons and wisdom of the past and opening our minds to the new vistas and parameters of populations and the potential impact of a population-based approach to care it charts the ways and means toward the future. . . , which is now. The past is prologue.

These are nursing's new **Tools of the Trade**.

Patricia A. Polansky, RN, MS
Director, Program Development and Implementation
Center to Champion Nursing in America
RWJF/AARP
Washington, DC
PREFACE

My good friend, colleague, and co-editor Patty Vitale died shortly after we completed the planning for this, the third edition of our book. Patty had enormous energy and zest. Both figuratively and literally she danced her way through life. The joy that she exuded on the dance floor was a reflection of the joy that she had for living. She was dedicated to making the world a better place for children through her work as a pediatrician and an educator. This book is part of the enduring legacy that she left behind.

The original inspiration for this book grew out of our experience while co-teaching an epidemiology course for students enrolled in a doctorate in nursing practice (DNP) program. We found it difficult to find a textbook that addressed the course objectives and was relevant to nursing practice. We decided a population-based nursing textbook, targeted for use as a primary course textbook in a DNP program or as a supplement to other course materials in a graduate community health nursing program, would be of great benefit and value to students enrolled in these programs. This book is the result of that vision. The chapters address the essential areas of content for a DNP program as recommended by the American Association of Colleges of Nursing (AACN), with a focus on the AACN core competencies for population-based nursing. The primary audience for this text is nursing students enrolled in either a DNP program or a graduate community health nursing program. Each chapter includes discussion questions to help students use and apply their newly acquired skills from each chapter.

In this book, the third edition, our goals were to not only update the content of the existing chapters, but also add a chapter on accreditation of population-based programs. We were fortunate that a nurse with extensive experience in accreditation (Eileen Horton) agreed to write the chapter for us. In order to make it easier for readers to enhance their knowledge of the information that is covered in the book, we also decided to add a relevant list of Internet resources to each chapter.

Several events covering a wide range of issues in the healthcare field have occurred over the past few years. These include the attempts by the Trump administration to dismantle the Affordable Care Act and efforts by the 116th Congress to expand the role of public programs in healthcare. Bills being introduced in the House range in scope from broad proposals to create a new national health insurance program for all residents (often referred to as “Medicare for All”) to more incremental approaches that would offer a public plan option in addition to current sources of coverage. It appears that the 2020 elections may well turn into a referendum on healthcare and how it should be paid for. States and local governments are increasingly turning to
legislation in an attempt to quell outbreaks of measles. As of April 2019, officials at the Centers for Disease Control and Prevention (CDC) had confirmed 695 measles cases across 22 states for the current year, a record high since the disease was thought to have been eliminated in the United States in 2000. According to a UNICEF report, among high-income nations, the United States had the most children who went unvaccinated between 2010 and 2017 (CDC, 2019). There is increasing concern over the usage of electronic cigarettes and hookahs by children. There is also increasing interest in the use of social media to address population health. This edition addresses these as well as other current issues in population-based nursing.

As in the first and second editions, this textbook includes successful strategies that nurses have used to improve population outcomes and reinforces high-level application of activities that require the synthesis and integration of information learned. The goal is to provide readers with information that will help them to identify healthcare needs at the population level and improve population outcomes. In particular, Chapter 1, Introduction to Population-Based Nursing, introduces the concept of population-based nursing and discusses examples of successful approaches and interventions to improve population health. In this edition we use the title “advanced practice registered nurse” (APRN). APRN is the title used in The Consensus Model for APRN Regulation, Licensure, Accreditation, Certification and Education (APRN Consensus Workgroup & National Council of State Boards of Nursing APRN Advisory Committee, 2008). This document is the product of the APRN Consensus Work Group and the National Council of State Boards of Nursing (NCSBN).

In order to design, implement, and evaluate interventions that improve the health of populations and aggregates, APRNs need to be able to identify and target outcome measures. Chapter 2, Identifying Outcomes in Population-Based Nursing, explains how to define, categorize, and identify population outcomes using specific examples from practice settings. The identification of outcomes or key health indicators is an essential first step in planning effective interventions and is a requirement for evaluation. The chapter includes a discussion of nurse-sensitive indicators, Healthy People 2020, Healthy People 2030, national health objectives, and health disparities. Emphasis is on the identification of healthcare disparities and approaches that can be used to eliminate or mitigate them. APRNs can advocate needed change at local, regional, state, or national levels by identifying areas for improvement in practice, by comparing evidence needed for effective practice, and by better understanding health disparities. APRNs have an important collaborative role with professionals from other disciplines and community members to work toward eliminating health disparities.

Epidemiology is the basic science of prevention (Gordis, 2014). Evidence-based practice, as it relates to population-based nursing, combines clinical practice and public health through the use of population health sciences in clinical practice (Heller & Page, 2002). Programs or interventions that are designed by APRNs should be evaluated and assessed for their effectiveness and ability to change or improve outcomes. This is true at an individual or population level. Data from these programs should be collected systematically and in such a manner that can be replicated in future programs. Data collection
must be organized and analyzed using clearly defined outcomes developed early in the planning process. Best practice requires that data are not just collected; data must also be analyzed, interpreted correctly, and, if significant, put into practice. Understanding how to interpret and report data accurately is critical as it sets up the foundation for evidence-based practice. With that said, it is important to understand the basics of how to measure disease or outcomes, how to present these measures, and to know what types of measures are needed to analyze a project or intervention.

Chapter 3, Epidemiological Methods and Measurements in Population-Based Nursing Practice: Part I, describes the natural history of disease and concepts that are integral to the prevention and recognition (e.g., screening) of disease. Basic concepts that are necessary to understand how to measure disease, and design studies that are used in population-based research, are discussed. Disease measures, such as incidence, prevalence, and mortality rates, are covered, and their relevance to practice is discussed. This chapter also includes information on primary, secondary, and tertiary prevention, and the concept of causality is introduced. A section on survival and prognosis is included. This material broadens the knowledge of readers with information necessary for advanced practice and interpretation of survival data. The basics of data analysis, including the calculation of relative risk, attributable risk, and odds ratio, are presented with examples of how to use these measures. Study design selection is an important part of the planning process for implementing a program. A portion of Chapter 3, Epidemiological Methods and Measurements in Population-Based Nursing Practice: Part I, is dedicated to introducing the most common study designs, because correct design selection is an essential part of sound methodology, successful program implementation, and overall success.

In order for APRNs to lead the field of evidence-based practice, it is critical that they possess skills in analytic methods to identify population trends and evaluate outcomes and systems of care (American Association of Colleges of Nursing [AACN], 2006). They need to carry out studies with strong methodology and be cognizant of factors that can affect study results. Identification and early recognition of factors that can affect the results or outcomes of a study, such as systematic errors (e.g., bias), should be acknowledged because they cannot always be prevented. In Chapter 4, Epidemiological Methods and Measurements in Population-Based Nursing Practice: Part II, the APRN is introduced to the elements of bias with a comprehensive discussion of the complexities of data collection and the fundamentals of developing a database. More in-depth discussion of study designs is covered, as well as a comprehensive review of ways to report on randomized and nonrandomized studies. Critical components of data analysis are discussed, including causality, confounding, and interaction.

In order to provide care at an advanced level, nurses must incorporate the concepts and competencies of advanced practice into their daily practice. This requires that APRNs acquire the knowledge, tools, and resources to know when and how to integrate them into practice. In Chapter 5, Applying Evidence at the Population Level, the APRN learns how to integrate and synthesize information in order to design interventions that are based on evidence to improve population outcomes. Nurses require several skills to become practitioners of evidence-based care. In this chapter, they learn
how to identify clinical problems, recognize patient safety issues, compose clinical questions that provide a clear direction for study, conduct a search of the literature, appraise and synthesize the available evidence, and successfully integrate new knowledge into practice.

Information technologies are transforming the way that information is learned and shared. Online communities provide a place for people to support each other and share information. Online databases contain knowledge that can be assessed for information on populations and aggregates, and many websites provide up-to-date information on health and healthcare. Chapter 6, Using Information Technology to Improve Population Outcomes, describes how technology can be used to enhance population-based nursing. It identifies websites that are available and how to evaluate them for quality. It also describes potential ways that technology can be used to improve population outcomes and how to incorporate technology into the development of new and creative interventions. APRNs use data to make decisions that lead to program development, implementation, and evaluation. In Chapter 7, Concepts in Program Design and Development, the APRN learns how to design new programs using organizational theory. Nursing care delivery models that address organizational structure, process, and outcomes are described.

Oversight responsibilities for clinical outcomes at the population level are a critical part of advanced practice nursing. The purpose of Chapter 8, Evaluation of Practice at the Population Level, is to identify ways and means to evaluate population outcomes and systems changes, as well as to address issues of effectiveness and efficiency and trends in care delivery across the continuum. Strategies to monitor healthcare quality are addressed, as are factors that lead to success. These concepts are explored within the role and competencies of the APRN.

The healthcare marketplace is extremely competitive. Administrators are constantly on the look out to identify opportunities to differentiate and validate their organization. Achieving accreditation helps to validate programs and organizations in the context of national and professional standards. Developing programs and working toward program accreditation requires competence in each of the DNP essentials. Chapter 9, The Role of Accreditation in Validating Population-Based Practice/Programs, describes the role of the APRN in the accreditation process.

In order for APRNs to make decisions at the community level, APRNs who work in the community need to be part of the higher level of care management and policy-making and decision-making, in partnership with the community-based consortium of healthcare policy makers. Chapter 10, Building Relationships and Engaging Communities Through Collaboration, describes the tools for successful community collaboration and project development. Emphasis is placed on identifying community needs and assessment of their resources. Specific examples are given to guide APRNs in developing their own community projects.

Chapter 11, Challenges in Program Implementation, identifies barriers to change within communities and the importance of developing and sustaining community partnerships. Specific strategies for program implementation are discussed, as well as the
methods to empower the community to advocate for themselves. Specific examples are given in order to guide APRNs in executing a project that has community acceptance and sustainability.

Finally, Chapter 12, Implications of Global Health in Population-Based Nursing, explores the implications of global health for the APRN. Theories of global health, population health, and public/community health are differentiated and compared, to further the understanding of how environmental conditions (e.g., poverty, housing, access to care) affect the health status of individuals and groups. Recent patterns in international interdisciplinary collaborations are reviewed, including the global health competencies developed by the Association of Schools of Public Health (ASPH) and the AACN.

Qualified instructors may obtain access to an instructor’s manual for this title by contacting textbook@springerpub.com.

Ann L. Cupp Curley

REFERENCES


ACKNOWLEDGMENTS

I wish to express my thanks to those professional colleagues who provided direction, guidance, and assistance in writing this book. Thank you also to my family and friends for their support throughout this process. I want to give a very special thank you to my husband, Ed, for his unending patience and excellent advice. Thank you to my publisher, Adrianne Brigido, for her valuable advice, assistance, and infinite patience, and to Cindy Yoo and Kris Parrish at Springer Publishing Company for their assistance and expertise throughout the revision of this book. Finally, I could not have completed the text without the help of the library staff of the Health Services Library of Capital Health especially Erika Moncrief, MS, Director of Library Services.
INTRODUCTION

Evidence-based practice as it relates to population-based nursing combines clinical practice and public health through the use of population health sciences in clinical practice (Heller & Page, 2002). Epidemiology is the science of public health. It is concerned with the study of the factors determining and influencing the frequency and distribution of disease, injury, and other health-related events and their causes (Gordis, 2014). In addition to epidemiology, an understanding of other scientific disciplines, such as biology and biostatistics, is also important for identifying associations and determining causation when looking at exposures and outcomes as they relate to population health.

Population-based care focuses on populations at risk, analysis of aggregate data, evaluation of demographic factors, and recognition of health disparities. It is concerned with the patterns of delivery of care and outcome measurements at the population or subpopulation level. The purpose of this chapter is to provide readers with an understanding of the natural history of disease and the approaches that are integral for the prevention of disease. It addresses the Doctor of Nursing Practice competencies specified in Essential III: Clinical Scholarship and Analytical Methods for Evidence-Based Practice (American Association of Colleges of Nursing [AACN], 2006). We introduce basic concepts that are necessary to understand how to measure disease outcomes and select study designs that are best suited for population-based research. Emphasis is placed on measuring disease occurrence with a fundamental discussion of how to calculate incidence, prevalence, and mortality rates. Successful advanced
practice nursing in population health depends upon the ability to recognize the difference between the individual and population approaches to the collection and use of data and the ability to assess needs and evaluate outcomes at the population level. Concepts surrounding survival data are also discussed along with strategies to guide advanced practice registered nurses (APRNs) on how to calculate and interpret survival data.

THE NATURAL HISTORY OF DISEASE

The natural history of disease refers to the progression of a disease from its *preclinical state* (prior to symptoms) to its *clinical state* (from onset of symptoms to cure, control, disability, or death). Disease is not something that occurs suddenly, but rather it is a multifactorial process that is dynamic and occurs over time. It evolves and changes and is sometimes initiated by events that take place years, even decades, before symptoms first appear. Many diseases have a natural life history that can extend over a very long period of time. The natural history of disease is described in stages. Understanding the different stages allows for a better understanding of the approach to the prevention and control of disease.

**Stage of Susceptibility**

The stage of susceptibility refers to the time prior to disease development. In the presence of certain risk factors, genetics, or environment, disease may develop and the severity can vary among individuals. Risk factors are those factors that are associated with an increased likelihood of disease developing over time. The idea that individuals could modify “risk factors” tied to heart disease, stroke, and other diseases is one of the key findings of the Framingham Heart Study (National Heart, Lung, and Blood Institute and Boston University, 2018). Started in 1948 and still in operation, this study is one of the most important population studies ever carried out in the United States. Before Framingham, for example, most healthcare providers believed that atherosclerosis was an inevitable part of the aging process. Although not all risk factors are amenable to change (e.g., genetic factors), the identification of risk factors is important and fundamental to disease prevention.

**Preclinical Stage of Disease**

During the preclinical phase, the disease process has begun but there are no obvious symptoms. Although there is no clear manifestation of disease, because of the interaction of biological factors, changes have started to occur. During this stage, however, the changes are not always detectable. Screening technologies have been developed to detect the presence of some diseases before clinical symptoms appear. The Papanicolaou (Pap) smear is an example of an effective screening method for detecting cancer in a premalignant state to improve mortality related to cervical cancer. The use of the Pap smear as a screening tool facilitates early detection and treatment of premalignant changes of the cervix prior to development of malignancy.
Clinical Stage of Disease

In the clinical stage of disease, sufficient physiologic and/or functional changes occur, leading to the development of recognizable symptoms of disease. It might also be accurately referred to as the treatment stage. For some people, the disease may completely resolve (either spontaneously or with medical intervention), whereas for some it will lead to disability and/or death. It is for this reason that the clinical stage of disease is sometimes subdivided for better medical management. Staging systems used in malignancies to better define the extent of disease involvement are an example of a system that can help guide the type of treatment modality selected based on the stage. In many cases, staging can provide an estimate of prognosis. Another example is the identification of disability as a specific subcategory of the treatment stage. Disability occurs when a clinical disease leaves a person either temporarily or permanently disabled. When people become disabled, the goal of the treatment is to mitigate the effects of disease and to help these individuals to function to their optimal abilities. This is very different from the goal for someone who can be treated and restored to the level of functioning that he or she enjoyed prior to the illness.

The Nonclinical Disease Stage

This nonclinical or unapparent disease stage can be broken into four subparts. The first subpart is the preclinical stage, which, as mentioned earlier, is the acquisition of disease prior to development of symptoms and is destined to become disease. The second subpart is the subclinical stage that occurs when someone has the disease but it is not destined to develop clinically. The third subpart is the chronic or persistent stage of disease, which is disease that persists over time. And finally, there is the fourth subpart or latent stage in which one has disease with no active multiplication of the biologic agent (Gordis, 2014).

The Iceberg Phenomenon

For most health problems, the number of identified cases is exceeded by the number of unidentified cases. This occurrence, referred to as the “iceberg phenomenon,” makes it difficult to assess the true burden of disease. Many diseases do not have obvious symptoms, as stated earlier, and may go unrecognized for many years. Unrecognized diseases, such as diabetes, hypertension, and mental illness, create a significant problem with identifying populations at risk and estimating service needs. Complications also arise when patients are not recognized or treated during an early stage of a disease when interventions are most effective. Additionally, patients who do not have symptoms or do not recognize their symptoms do not seek medical care and, in many cases, even if they do have a diagnosis, do not take their medications as they perceive that they are healthy when they are asymptomatic.

PREVENTION

Understanding the natural history of disease is as important as understanding the causal factors of disease because it provides the APRN with the knowledge that is required to
design programs or interventions that target populations at risk. Understanding how disease develops is fundamental to the concept of prevention and provides a framework for disease prevention and control. The primary goal of prevention is to prevent disease before it occurs. The concept of prevention has evolved to include measures taken to interrupt or slow the progression of disease or to lessen its impact. There are three levels of prevention.

Primary Prevention

*Primary prevention* refers to the process of altering susceptibility or reducing exposure to susceptible individuals and includes general health promotion and specific measures designed to prevent disease prior to a person getting the disease. Interventions designed for primary prevention are carried out during the stage of susceptibility and can include things such as providing immunizations to change a person's susceptibility. Actions taken to prevent tobacco usage are another example of primary prevention. Tobacco use is one of the 12 leading health indicators used by Healthy People 2020 to measure health. Cigarette smoking is the leading cause of preventable mortality in the United States (Centers for Disease Control and Prevention [CDC], 2018a), and prevention or cessation of smoking can reduce the development of many smoking-related diseases. Taxes on cigarettes, education programs, and support groups to help people stop smoking and the creation of smoke-free zones are all examples of primary prevention measures. The CDC linked a series of tobacco control efforts by Minnesota to a decrease in adult smoking prevalence rates. From 1999 to 2010, Minnesota implemented a series of antismoking initiatives, including a statewide smoke-free law, cigarette tax increases, media campaigns, and statewide cessation efforts. Adult smoking prevalence decreased from 22.1% in 1999 to 16.1% in 2010 (CDC, 2011). In 2013, Minnesota increased the tax on a pack of cigarettes an additional $1.60. Following this increase, smoking decreased by 33% among Minnesota’s 11th graders and by 10% among adult residents. Smokers reported that the tax increase did influence their smoking behaviors (Minnesota Department of Health, 2018). This is an excellent example of a successful statewide primary prevention effort to reduce smoking prevalence through a variety of initiatives.

Secondary Prevention

The early detection and prompt treatment of a disease at the earliest possible stage are referred to as *secondary prevention*. The goals of secondary prevention are to either identify and cure a disease at a very early stage or slow its progression to prevent complications and limit disability. Secondary prevention measures are carried out during the preclinical or presymptomatic stage of disease. Screening programs are designed to detect specific diseases in their early stages while they are curable and to prevent or reduce morbidity and mortality related to a later diagnosis of disease. Examples of secondary prevention include the Pap smear, mentioned earlier, as well as annual testing of cholesterol levels, mammography, and rapid HIV testing of asymptomatic individuals.
Tertiary Prevention

Tertiary prevention strategies are implemented during the middle or late stages of clinical disease and refer to measures taken to alleviate disability and restore effective functioning. Attempts are made to slow the progression or to cure the disease. In cases in which permanent changes have taken place, interventions are planned and designed to help people lead a productive and satisfying life by maximizing the use of remaining capabilities (rehabilitation). Cardiac rehabilitation programs that provide physical and occupational therapies to postoperative cardiac patients are an example of tertiary prevention.

CAUSATION

The Epidemiological Triangle

The relationship between risk factors and disease is complex. Research studies may describe a relationship between a risk factor and disease, but how do we know that this relationship is causal? An understanding of causation is important if APRNs want to effectively impact the health of populations. The epidemiological triangle is a model that has historically been used to explain causation. The model consists of three interactive factors: the causative agent (those factors for which presence or absence cause disease—biologic, chemical, physical, nutritional), a susceptible host (things such as age, gender, race, immune status, genetics), and the environment (including diverse elements such as water, food, neighborhood, pollution). A change in the agent, host, and environmental balance can lead to disease (Harkness, 1995). The underlying assumptions of this model are that causative factors can be both intrinsic and extrinsic to the host and that the cause of disease is related to interaction among these three factors. This model was developed initially to explain the transmission of infectious diseases and was particularly useful when the focus of epidemiology was on acute diseases. It is less helpful for understanding and explaining the more complicated processes associated with chronic disease. With the rise of chronic diseases as the primary cause of morbidity and mortality, a model that recognizes multiple causative factors was needed to better understand this complex interaction.

The Web of Causation

The dynamic nature of chronic diseases calls for a more sophisticated model for explaining causation than the epidemiological triangle. Introduction of the web of causation concept first appeared in the 1960s when chronic diseases overtook infectious diseases as the leading cause of morbidity and mortality in the United States. The foundation of this concept is that disease develops as the result of many antecedent factors and not as a result of a single, isolated cause. Each factor is itself the result of a complex pattern of events that can be best perceived as interrelated in the complex configuration of a web. The use of a web is helpful for visualizing how difficult it is to untangle the many events that can precede the onset of a chronic illness.
Critics have argued that this model places too much emphasis on epidemiological methods and too little on theories of disease causation. As theories evolved about the relationship between smoking and cancer, the U.S. Surgeon General appointed a committee to review the evidence. This committee developed a set of guidelines for judging whether an observed association is causal. These guidelines include temporal relationship, strength of the association, dose–response relationship, replication of the findings, biologic plausibility, consideration of alternative explanation, cessation of exposure, consistency with other knowledge, and specificity of the association (Gordis, 2014). For a more detailed discussion on this model, see Chapter 4, Epidemiological Methods and Measurements in Population-Based Nursing Practice.

**METHODS OF ANALYSIS**

Successful population-based approaches depend on the ability to recognize the difference between the collection and use of data from individuals and populations and the ability to assess needs and evaluate outcomes at the population level. Several of the more recent theories of causation can be helpful in determining whether an exposure is causally related to the development of disease. In particular, calculating the strength of association using statistics is one of several criteria that can be used to determine causality. However, statistics must be used with caution. Health is a multidimensional variable: Factors that affect health, and that interact to affect health, are numerous. Many relationships are possible. There are problems inherent in the use of statistics to explain differences among groups. Although statistics can describe disparities, they cannot explain them. It is left to the researchers to explain the differences. In addition to statistics, one must also be aware of the validity and reliability of the data. There are problems associated with the categorizing and gathering of statistics that can have an effect on how the data should be interpreted. In order to be successful in research, one must do more than just collect data: One must look at the theoretical issues associated with explaining the relationship among the variables. Additionally, even if a relationship is found to be statistically significant, that does not ensure that it is clinically significant. Recognizing limitations in research and in practice are the most important steps prior to making conclusions in any setting. Therefore, it is important that APRNs have a commitment to higher standards with an emphasis placed on adherence to careful and thorough procedural and ethical practice.

Methods derived from epidemiology can be useful in identifying the etiology or the cause of a disease. Among the important steps in this process are the identification of risk factors and their impact in a population, determining the extent of a disease and/or adverse events found in a population, and evaluating both existing and new preventive and therapeutic measures and modes of healthcare delivery. Applying strong epidemiologic methods with a sound application and interpretation of statistics are the foundation for evidence-based practice. The integration of evidence can lead to the creation of good public policy and regulatory decisions.
Descriptive Epidemiology

Rates

Knowledge of how illness and injury are distributed within a population can provide valuable information on disease etiology and can lay the foundation for the introduction of new prevention programs. It is important to know how to measure disease in populations, and rates are a useful method for measuring attributes over time such as disease and injury in any population. Rates can also be used to identify trends and evaluate outcomes and can allow for comparisons within and between groups. The Morbidity and Mortality Weekly Report (MMWR; located at www.cdc.gov/mmwr) is a publication of the CDC and contains updated information on incidence and prevalence of many diseases and conditions. These rates provide healthcare providers with up-to-date information on the risks and burdens of various diseases and conditions (CDC, 2018b). The information obtained from the MMWR can be used to identify trends and provide policy makers with information for designating resources. The following is an example of such information:

Drinking sugar sweetened beverages (SSBs) is associated with several adverse health consequences including obesity, type 2 diabetes, and cardiovascular disease. In 2013 the Behavioral Risk Factor Surveillance System (a telephone survey) investigated self-reported SSB intake in the United States. In this survey of adults aged 18 years of age and older an SSB was identified as regular soda, fruit drink, sweet tea, and sports or energy drink intake. The results revealed that the overall age-adjusted prevalence of SSB intake once or more per day is 30% and ranges from 18% in Vermont to 47.5% in Mississippi. It is most prevalent among adults aged 18 to 24 years (43.3%), men (34.1%), non-Hispanic Blacks (39.9%), unemployed adults (34.4%) and persons with less than a high school education (42.4%). (This excerpt is adapted from an issue of MMWR published on February 26, 2016 [Park, Xu, Town, & Blanck, 2016].)

By publishing rates in percentages and comparing those rates among groups, it highlights the disparity between different demographic profiles related to SSB intake. Information such as this can be useful to both clinicians and policy makers who make decisions about interventions and services.

When calculating rates, the numerator is the number of events that occur during a specified period of time and is divided by the denominator, which is the average population at risk during that specified time period. This number is multiplied by a constant—either 100, 1,000, 10,000, or 100,000—and is expressed per that number. The purpose of expressing rates per 100,000, for example, is to have a constant denominator, and it allows investigators to compare rates among groups with different population sizes. To put it simply, the rate is calculated as follows:

\[
\text{Rate} = \frac{\text{Numerator}}{\text{denominator}} \times \text{Constant multiplier}
\]

In order to calculate rates, the APRN must first have a clear and explicit definition of the patient population and of the event. An important consideration when calculating
rates is that anyone represented in the denominator must have the potential to enter the group in the numerator, and all persons represented in the numerator must come from the denominator.

Rates can be either crude or specific. Crude rates apply to an entire population without any reference to any characteristics of the individuals within it. For example, to calculate the crude mortality rate, the numerator is the total number of deaths during a specific period of time divided by the denominator, which is the average number of people in the population during that specified period of time (including those who have died). Typically, the population value for a 1-year period is determined using the mid-year population.

Specific rates can also be calculated for a population that has been categorized into groups. Suppose that an APRN wants to calculate the number of new mothers who initiate breastfeeding in a specific hospital in 2018. The formula would be:

\[
\frac{\text{Total number of breastfeeding infants in community hospital in 2018}}{\text{Total number of live births in the same community hospital in 2018}} \times \text{Constant multiplier}
\]

In order to compare rates in two or more groups, the events in the numerator must be defined in the same way, the time intervals must be the same, and the constant multiplier must be the same. Rates can be used to compare two different groups, or one group during two different time periods. Returning to the example about breastfeeding, the breastfeeding rates could be compared in the same hospital, but at two different times, before and after implementation of a planned intervention to increase breastfeeding rates.

Formulae for the rates discussed in this chapter can be found in Exhibit 3.1.

**EXHIBIT 3.1**

**LIST OF USEFUL FORMULAE**

**Calculating Rates**

*Incidence rate* describes the occurrence of new disease cases in a community over a period of time relative to the size of the population at risk.

\[
\text{Incidence rate} = \frac{\text{Number of new cases during a specified period}}{\text{Population at risk during the same specified period}} \times \text{Constant multiplier}
\]  

(continued)
Prevalence rate is the number of all existing cases of a specific disease in a population at a given point in time relative to the population at risk.

\[
\text{Prevalence rate} = \frac{\text{Number of existing cases at a specified period}}{\text{Population at risk at the same specified period}} \times \text{Constant multiplier}
\]

Crude rates summarize the occurrence of births (crude birth rate) or deaths (crude death rate). The numerator is the number of events and the denominator is the average population size (usually estimated as a midyear population).

\[
\text{Crude death rate} = \frac{\text{Number of deaths in a population during a specified period}}{\text{Population estimate during same specified period}} \times \text{Constant multiplier}
\]

Specific rates are used to overcome some of the biases seen with crude rates. They are used to control for variables such as age, race, gender, and disease.

\[
\text{Age-specific death rate} = \frac{\text{Number of deaths for a specified age group during a specified time}}{\text{Population estimate for the specified age group during same specified time}} \times \text{Constant multiplier}
\]

Case fatality rate is used to measure the percentage of people who die from a certain disease. This rate tells you how fatal or severe a disease is compared to other diseases.

\[
\text{Case fatality rate} = \frac{\text{Number of individuals dying after disease onset or diagnosis}}{\text{Number of individuals with the specified disease}} \times 100
\]

Proportionate mortality ratio is useful for determining the leading causes of death.

\[
\text{Proportionate mortality ratio} = \frac{\text{Number of deaths from a specified cause during specified time period}}{\text{Total deaths during the same period}} \times 100
\]

(continued)
Calculations Used in Health Impact Assessment

**Number needed to treat (NNT)** is the number of patients needed to receive a treatment to prevent one bad outcome. The NNT calculated should be rounded up to the next highest number. Before the NNT can be calculated, the absolute risk reduction (ARR) must be identified.

\[
\text{ARR} = \text{Incidence in exposed} - \text{Incidence in nonexposed}
\]

\[
\text{NNT} = \frac{1}{\text{ARR}}
\]

The NNT can also be calculated in randomized trials using mortality rates:

\[
\text{NNT} = \frac{1}{(\text{Mortality rate in untreated group} - \text{Mortality rate in treated group})}
\]

**Disease impact number (DIN)** is the number of those with the disease in question among whom one event will be prevented by the intervention.

\[
\frac{1}{\text{ARR} \times \frac{\text{Proportion of people with the disease}}{\text{who are exposed to the intervention}}}
\]

**Population impact number (PIN)** is the number of those in the whole population among whom one event will be prevented by the intervention.

\[
\frac{1}{\text{ARR} \times \frac{\text{Proportion of people with the disease who are exposed}}{\text{to the intervention}} \times \frac{\text{Proportion of the total population with the disease of interest}}{\text{Proportion of the total population with the disease of interest}}}
\]

**Years of potential life lost (YPLL)** is used for setting health priorities. Predetermined standard age at death in the United States is 75 years.

\[
\text{YPLL (75)} = 75 - \text{Age at death from a specific cause}
\]

Add the years of life lost for each individual for specific cause of death = YPLL

**Calculations Used in Screening Programs**

**Sensitivity** is the ability of a screening test to identify accurately those persons with the disease.

\[
\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}
\]

**Specificity** reflects the extent to which it excludes the persons who do not have the disease.

\[
\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}
\]

(continued)
## EXHIBIT 3.1

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>NO DISEASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Test</td>
<td>False positive (FP)</td>
</tr>
<tr>
<td>True positive (TP)</td>
<td></td>
</tr>
<tr>
<td>– Test</td>
<td>True negative (TN)</td>
</tr>
<tr>
<td>False negative (FN)</td>
<td></td>
</tr>
</tbody>
</table>


### Incidence and Prevalence

*Incidence* rates describe the occurrence of new events in a population over a period of time relative to the size of the population at risk. *Prevalence* rates describe the number of all cases of a specific disease or attribute in a population at a given point in time relative to the size of the population at risk. Incidence provides information about the rate at which new cases occur and is a measure of risk. For example, the formula for the incidence rate for HIV is:

\[
\frac{\text{Total number of people who are diagnosed with HIV in a community during 2018}}{\text{Population in that community at midyear of 2018}} \times 1,000 = \text{Rate per 1,000}
\]

Incidence rates provide us with a direct measure of how often new cases occur within a particular population and provide some basis on which to assess risk. By comparing incidence rates among population groups that vary in one or more risk factors, the APRN can begin to get some idea of the association between risk factors and disease. If, in the earlier example of breastfeeding, the APRN discovers breastfeeding rates are significantly different among different ethnic groups, the characteristics of the groups can be compared and the causes for this disparity can be hypothesized and tested.

*Period prevalence* measures the number of cases of disease during a specific period of time and is a measure of burden. The formula for the period prevalence rate for HIV in 2018 is:

\[
\frac{\text{Total number of people who are HIV positive in a community during 2018}}{\text{Population in that community at midyear of 2018}} \times 1,000 = \text{Rate per 1,000}
\]

In the formula given here, all newly diagnosed cases for the year plus existing cases are included. *Point prevalence* is defined as the number of cases of disease at a specific point in time divided by the number of people at risk at that specific point in time multiplied by a constant multiplier. An example of the use of point prevalence would be the information
gathered from a survey in which an investigator asks questions such as who has diabetes, hypertension, epilepsy, or any other disease or event at that specific point in time. Prevalence, whether point or period, cannot give us an estimate of the risk of disease; it can only tell us about the burden of disease for a specified period of time. Prevalence is useful when comparing rates between populations but should be interpreted with caution. Diseases that are chronic will have a higher prevalence because at any given time, those with chronic disease will always have that disease. This can make it challenging to interpret prevalence rates as they do not tell us the risk of developing disease but they can be helpful when trying to determine resource needs for chronic diseases. With diseases that are short in duration, prevalence may not capture the true burden of disease for that population. Additionally, it is important to note that unidentified cases are not captured in either prevalence rates or incidence rates. Rates can only estimate the burden of disease, but they are the best way to draw comparisons using a common denominator.

An example of how prevalence rates are used in the literature is as follows:

The CDC (2018c) reported that the prevalence of obesity was 39.8% for U.S. adults in 2015–2016. Hispanics (47.0%) and non-Hispanic Blacks (46.8%) had the highest age-adjusted prevalence of obesity, followed by non-Hispanic Whites (37.9%) and non-Hispanic Asians (12.7%). The prevalence of obesity was 35.7% among young adults aged 20 to 39 years, 42.8% among middle-aged adults aged 40 to 59 years, and 41.0% among older adults aged 60 and over.

Information on the prevalence of adult obesity in the United States has led to increased attention to factors that cause obesity (especially in children). This has led to the development of new programs aimed at primary and secondary prevention.

**Mortality Rates**

*Mortality rates*, also known as death rates, can be useful when evaluating and comparing populations. As stated earlier, there are many factors that can affect the natural history of disease, and measuring mortality allows investigators to compare death rates among and within populations. The formula for mortality rate is:

\[
\text{Mortality rate} = \frac{\text{Number of deaths in a population during a specified time}}{\text{Average population estimate during the specified time}} \times \text{Constant multiplier}
\]

Mortality rates can be specific or broad in definition and can include any qualifiers for time, age, or disease type. It is important to include those specifics in your denominator to ensure that the population value used is the best estimate of the population at risk. For example, to look at the number of deaths in 2018 due to breast cancer in women aged 18 to 40, the denominator should only include the midyear population of women aged 18 to 40 in 2018. It is also important to include those women who died during that year in the denominator. Again, it is impossible to know exactly how many women in that
age group are at risk using a midyear population, but the key is to use similar sources of measurement so that comparisons can be made assuming similar sources are used to estimate the denominator.

Standardization of crude rates is an important consideration when comparing mortality rates among populations. Standardization is used to control the effects of age and other characteristics in order to make valid comparisons between groups. Age adjustment is an example of rate standardization and perhaps the most important one. No other factor has a larger effect on mortality than age. Consider the problem of comparing two communities with very different age distributions. One community has a much higher mortality rate for colon cancer than the other, leading investigators to consider a possible environmental hazard in that community, when in fact, that community's population is older, which could account for the higher mortality. Direct age adjustment or standardization allows a researcher to eliminate the age disparities between two populations by using a standardized population. This allows the researcher to compare mortality or death rates between groups by eliminating age differences between populations and comparing actual age-adjusted mortality rates to determine whether age truly plays a role in the crude unadjusted mortality rates.

There are two methods of age adjustment: direct, as mentioned earlier, and indirect. The direct method applies observed age-specific mortality or death rates to a standardized population. The indirect method applies the age-specific rates of a standardized population to the age distribution of an observed population and is used to determine whether one population has a greater mortality because of an occupational hazard or risk compared to the general population. (To learn how to perform age adjustment, refer to an advanced epidemiology text.)

The case fatality rate (CFR) is a measure of the severity of disease (such as infectious diseases) and can be helpful when designing programs to reduce the rate or disparity in the population. It should be noted that CFR is not a true rate as it has no explicit time implication but rather is a proportion of persons with disease who died from that disease after diagnosis. It is a measure of the probability of death among diagnosed cases. Its usefulness for chronic diseases is limited because the length of time from diagnosis to death can be long. CFR is also useful in determining when to use a screening test. Screening tests identify disease early so that an intervention or treatment can be initiated in the hopes of lessening the morbidity or mortality of that disease. Those diseases that are rapidly fatal may not necessarily be beneficial to screen unless the screening will allow for a cure or treatment to change the overall outcome or to prevent unnecessary spread of the disease. Screening is useful in identifying disease in asymptomatic individuals in whom further transmission of disease can be prevented or reduced, such as in HIV. CFRs, therefore, can be helpful for comparisons between study populations and can provide useful information that could help determine whether an intervention or treatment is working. The formula for CFR is as follows:

\[
\text{Case fatality rate} \% = \frac{\text{Number of individuals with the specified disease after disease onset or diagnosis}}{\text{Number of cases of that specific disease}} \times 100
\]
CFR is usually expressed as a percentage; so in this case, one would multiply this rate by a constant multiplier of 100 to obtain the percentage of disease that is fatal. It is important in all of these rates to include those who have died from the disease in the denominator. Removing those who have died from the denominator falsely increases the CFR, making the disease appear more fatal or severe (Gordis, 2014).

The proportionate mortality ratio is useful for determining the leading causes of death. The formula for proportionate mortality ratio is as follows:

\[
\frac{\text{Number of deaths from a specified cause during specified time period}}{\text{Total deaths from all causes during the same specified time period}} \times 100
\]

Again, this measure is usually reported as a percentage and reflects the burden of death due to a particular disease. This information is useful for policy makers who make decisions about the allocation of resources. (See Exhibit 3.1 for a list of these formulae.)

**Survival and Prognosis**

Mortality rates are very helpful when comparing groups and looking at disparities among populations. One cannot discuss mortality without having an understanding of survival and prognosis. Many diseases, particularly cancer, are studied over time, with attention placed on survival. Ideally, survival should be measured from the onset of disease until death, but the true onset of disease is generally unknown. Survival rates are usually calculated at various intervals from diagnosis or initiation of treatment. Prognosis is calculated using collected data to estimate the risk of dying or surviving after diagnosis or treatment begins. As mentioned earlier, CFRs give a good estimate of prognosis or severity of disease. However, they are best suited for acute diseases in which death occurs relatively soon after diagnosis. Survival analysis is better suited for chronic diseases or those diseases that take time to progress.

Survival time is generally calculated from the time of diagnosis or from the start of treatment. This can vary from patient to patient, as some patients may seek care immediately after symptoms present or may wait months to seek care. Some patients are diagnosed prior to symptom presentation after they screened positive on a screening test. Some may obtain a diagnosis immediately, whereas others may have poor access to care, and diagnosis is delayed by weeks to months or even years. Once a diagnosis is made, treatment may or may not occur immediately. Additionally, some patients may die before diagnosis or treatment. Because these individuals are not represented in survival analysis, this can lead to a falsely increased survival time. With that said, one can see how difficult it is to establish a true survival time after diagnosis. However, we can estimate survival if we use a common denominator and consistent criteria for measurement.

Before we discuss how to calculate and interpret survival data, we must touch on two important concepts, *lead time bias* and *overdiagnosis bias*. Lead time bias is a phenomenon whereby a patient is diagnosed earlier by screening and appears to have increased
survival due to screening but rather dies at the same time he or she would regardless of screening. In other words, the time from which a patient is diagnosed earlier from screening is the lead time, and the bias is the error that occurs as a result of concluding that screening leads to a longer survival after diagnosis. As can be seen in the following timeline (Figure 3.1), the survival time is longer when screening is implemented, but the ultimate time of death is unchanged. Although this is not true for all screening tests, it is important to recognize the phenomenon of lead time bias as it can affect the conclusions that are made regarding survival, which ultimately can affect a patient’s perceived prognosis.

Overdiagnosis bias occurs as a result of making a diagnosis from screening for a disease or cancer that would not have manifested clinically or has a slow progression, such that the person dies from another etiology. This type of bias has the potential to increase undue stress in individuals and can also falsely increase survival times, especially for diseases with slow progression. In both these types of biases, there is no difference in overall mortality in those screened versus those who were not screened. With that said, considerations must also be made for those screening tests in which a false-negative test reassures a patient who may not seek care and ultimately develops cancer and potentially has decreased survival due to delay in diagnosis. All of these biases need to be taken into consideration when interpreting survival data.

Prognosis is calculated using survival rates. There are two methods of conducting survival analysis and estimating prognosis that are discussed. The first is the actuarial

![FIGURE 3.1 Timeline illustrating lead time bias.](image-url)
method, which measures the likelihood of surviving after each year of treatment (or a predetermined interval). This is calculated as follows: the probability of surviving 2 years if one survived 1 year, or the probability of surviving 3 years if one survived 2 years, and so on. Prognosis is most commonly described in the literature as the probability of surviving 1, 2, 3, or more years. Generally, survival is calculated as a probability $P_1$, $P_2$, $P_3$, and so on. The survival after 1 year is designated as $P_1$; if patients survived 1 year after treatment, those who survived to 2 years = $P_2$; if patients survived 2 years after treatment, those who survived to 3 years = $P_3$; and so on. To calculate $P_1$, divide the number of survivors over the number of patients with the disease at the start of the study or treatment. It is important to note that those who are lost to follow-up (also known as withdrawals) or who are no longer studied must be removed from the denominator. When a study ends or is terminated, those patients are no longer followed and must be taken into consideration in your analysis, and this is called censorship. For simplicity, the following examples will assume no losses to follow-up, but it is important to recognize that those who are lost to follow-up or those who are censored must be taken into account in your calculations. Of note, in a more advanced epidemiology textbook, you will find that those who are lost to follow-up will be subtracted out of the denominator and multiplied by 1/2 to account for the chance they were at risk for half the interval. Again, for the purposes of this text, we will use a hypothetical example in which no patients are lost to follow-up. By definition, here is how to calculate $P_1$, $P_2$, and so on:

- $P_1 = \frac{\text{Number alive after 1 year of treatment}}{\text{Number who started treatment}}$
- $P_2 = \frac{\text{Number alive after 2 years of treatment}}{\text{Number who survived first year of treatment} - \text{Those who dropped out or were lost to follow-up}}$
- $P_3 = \frac{\text{Number alive after 3 years of treatment}}{\text{Number who survived second year of treatment} - \text{Those who dropped out or were lost to follow-up}}$

To calculate the probability of surviving 1, 2, 3, or more years, the calculation is as follows:

- $P_1 =$ probability of surviving 1 year
- $P_1 \times P_2 =$ probability of surviving 2 years
- $P_1 \times P_2 \times P_3 =$ probability of surviving 3 years
- $P_1 \times P_2 \times P_3 \times P_4 =$ probability of surviving 4 years
- $P_1 \times P_2 \times P_3 \times P_4 \times P_5 =$ probability of surviving 5 years

Using data from Table 3.1, we can calculate and interpret these probabilities.

<table>
<thead>
<tr>
<th>TABLE 3.1 Survival Rates After Treatment (Hypothetical Life Table of 100 Patients With No Patients Lost to Follow-Up)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NUMBER SURVIVED</strong></td>
</tr>
<tr>
<td><strong>AFTER 1 YEAR</strong></td>
</tr>
<tr>
<td>Cohort (N = 100)</td>
</tr>
</tbody>
</table>
In this example:
P1 = 88/100 = 0.88
P2 = 76/88 = 0.86
P3 = 55/76 = 0.72
P4 = 47/55 = 0.85
P5 = 33/47 = 0.70
Probability of surviving 1 year = 0.88
Probability of surviving 2 years = 0.88 × 0.86 = 0.76
Probability of surviving 3 years = 0.88 × 0.86 × 0.72 = 0.54
Probability of surviving 4 years = 0.88 × 0.86 × 0.72 × 0.85 = 0.46
Probability of surviving 5 years = 0.88 × 0.86 × 0.72 × 0.85 × 0.70 = 0.32

It is important to distinguish between the probability of surviving 5 years and the probability of surviving 5 years given that someone survived 4 years. Generally, the longer someone survives after treatment, the more likely that person will make it to the next year. Overall survival after 5 years is always a smaller number as the probability of surviving each year is multiplied against each year (Gordis, 2014).

Note that the actuarial method can be used to look at outcomes other than survival or death as it can estimate probabilities of an outcome or event occurring such as a treatment side effect (e.g., vomiting, headache) or recurrence of disease. Another important consideration is survival over time. When looking at survival rates measured over years, it is important that an APRN take into account the improvements and advances in treatments over time. APRNs should consider comparing survival rates for earlier treatment regimens with those for newer regimens, as this can affect the validity of the overall survival if not taken into consideration. In addition, certain confounders (e.g., age, gender, ethnicity, socioeconomic status) may contribute to differences in survival rates and should be examined when performing a survival analysis (see Chapter 4, “Epidemiological Methods and Measurements in Population-Based Nursing Practice,” for more on confounding). Recognition of these differences is a critical step for the evaluation of potential health disparities and is a perfect opportunity for an APRN to develop strategies to address the underlying issue causing those disparities.

In the literature, survival analysis using the actuarial method is plotted on a curve in which the x-axis represents time and the y-axis represents the number of survivors at each time interval. This is called a survival curve and represents the pattern of survival over predetermined time intervals. Using the data from the earlier example, the probabilities are plotted in a standardized survival curve (Figure 3.2).

The second type of survival analysis is the Kaplan–Meier method. This method is commonly used in medicine and is well suited for analyses of small and large populations, as well as comparisons between treatments or interventions. Although beyond the scope of this book, statistical analyses can be performed to compare treatments or interventions...
using tests of significance (log-rank test) and logistic regression (proportional hazard models [Cox models]). As with any comparison trial, it is important to take into consideration the characteristics of those patients who are lost to follow-up because if they occur more frequently in one treatment group compared to another, this can affect the results. For example, if the majority of patients lost to follow-up are receiving treatment A and most of them can be characterized as impoverished with poor access to care, then this could skew the results of the remaining patients receiving treatment A. Thus, minimizing loss to follow-up or censored patients and/or maintaining similar losses with similar characteristics in each group is paramount to reducing bias and improving the strength of the study conclusions. This reiterates the importance of randomization, which will be discussed more thoroughly in Chapter 4, Epidemiological Methods and Measurements in Population-Based Nursing Practice.

**Kaplan–Meier** curves are used to plot survival, and these plots represent a stepwise pattern of survival in which the increments of time are not standardized (e.g., 1 year, 5 years), but rather each step represents an event (e.g., time to death or an outcome of interest). Kaplan–Meier curves are seen more commonly in the literature and are a better estimate of survival as they also take into consideration patients who are lost to follow-up or are censored. These curves also allow for comparisons between different treatment regimens (Figure 3.3).

Kaplan–Meier curves are different from traditional survival curves in that they do not slope downward after each event but rather maintain a horizontal line until the next event (e.g., death) occurs, and then a downward vertical line is drawn until the new cumulative survival is reached and the steps are continued until the study is completed. At the time in which no deaths are occurring (also known as the death-free period), the cumulative survival is maintained; however, hatch marks can be seen in these plots, which represent those lost to follow-up or censored during that interval (Jekel, Katz, Elmore, & Wild, 2007).

The importance of having the knowledge and skills to interpret and calculate survival data cannot be understated. APRNs can use survival data or outcome data in various ways. Most importantly, the evidence obtained from survival or outcome data can help
APRNs to design and justify interventions to improve the quality of life for diseases such as cancer. Comparisons to other groups can be made by addressing outcomes of interest to determine whether certain interventions make a difference in the quality of life and ultimately impact the survival of those involved.

**Health Impact Assessment**

As mentioned previously, rates can be used to describe the distribution of disease and other health-related states and events, but sometimes the APRN may be more concerned with knowing how data can be used to describe the relevancy of clinical practice. *Health impact assessment* (HIA) is the assessment of the potential health effects, positive or negative, of a particular intervention on a population. HIAs can evaluate population-directed programs or interventions before they are implemented and can provide recommendations on how those programs can potentially affect the health of a population irrespective of whether positive or negative. Certain calculations can be performed to determine the efficacy of a treatment or intervention. The number needed to treat (NNT), the disease impact number (DIN), and the population impact number (PIN) are formulae that are used in HIAs. NNT is the number of patients needed to receive a treatment to prevent one bad outcome, and the lower the NNT, the better for assessing superiority.

![Figure 3.3 Hypothetical example of a Kaplan–Meier curve—comparison of treatment A to treatment B.](image-url)
of treatments. However, the NNT takes into account only those patients being treated rather than all those with disease in the population. The DIN, on the other hand, uses the number of those with the disease in question among whom one event will be prevented by the intervention. Similarly, the PIN is the number of those in the whole population among whom one event will be prevented by the intervention (Heller & Dobson, 2000). Another calculation commonly used is the YPLL, which measures premature mortality, and the productive years that are lost related to early death (Merrill, 2017). Information on years of potential life lost (YPLL) helps to magnify the importance of primary prevention measures designed to address diseases such as obesity and other risk factors such as smoking. Each of these measurements is helpful in determining the benefits or risks of new interventions or treatments. Specifically, the DIN and the PIN provide a better population-based estimate of treatment or intervention impacts on the population as a whole. (See Exhibit 3.1 for a list of these formulae.)

It is important for the APRN who is involved in population-based evaluation to be aware of these concepts. More extensive information on HIA formulae and standardization can be found in most advanced epidemiology texts and on the CDC’s Health Impact Assessment Resources page at https://www.cdc.gov/healthyplaces/hiaresources.htm.

DESCRIPTIVE STUDIES

Descriptive epidemiology is used to describe the distribution of disease and other health-related states and events in terms of personal characteristics, geographical distribution, and time. There are four types of descriptive studies: case reports, case series, cross-sectional studies, and correlation or ecologic studies. The data used in descriptive studies are often readily available and can be retrieved from such sources as hospital records, census data, or vital statistics records.

Case Reports and Case Series

Case reports are succinct written accounts of generally rare or unusual cases in which the treatment or management of the disease or condition is worth reporting. These are usually published to assist healthcare providers in the management of rare, unresearched, or undocumented cases. A case series is merely a report of a series of patients with similar diseases or conditions that describes their management or treatment in order to identify new strategies that may be helpful to treat patients with similar conditions. They also lead to future studies and can be helpful for APRNs as they can use these cases to build a case for future research of treatments or interventions that have not yet been rigorously studied.

Correlation Studies

Correlation studies are also referred to as ecologic studies and are used to conduct studies of aggregate or population characteristics. In ecologic studies, rates are calculated for characteristics that describe populations and are used to compare frequencies between different groups at the same time or the same group at different times. They are useful
for identifying long-term trends, seasonal patterns, and event-related clusters. Because data are collected on populations instead of individuals, an event cannot be linked to an exposure in individuals, and the investigator cannot control for the effect of other variables. These types of studies lead to more rigorous studies that can control for variables of interest and look at individual data to determine whether an association truly exists. Correlational studies can only report that a correlation exists and cannot show an association exists as they compare population or aggregate data. An example of a correlation study would be one that shows a correlation between high fat content and breast cancer. Countries with a high fat content correlate to countries with higher rates of breast cancer. Without knowing individual data, one cannot determine whether women with breast cancer actually also have a high fat consumption (Gordis, 2014).

A study by Pillai, Maleku, and Wei (2013) provides an example of a correlational study. The authors used data from 143 countries to study the relationship between female literacy and maternal mortality. Their analysis reveals a significant negative relationship between female literacy rates and maternal mortality. Populations with a higher prevalence of literacy have lower maternal mortality rates and populations with a lower prevalence of literacy have higher maternal mortality rates. The authors point out limitations to their study, most importantly the difficulty of controlling for known correlates with maternal mortality (such as access to healthcare services) due to a scarcity of cross-national data. These data show that a correlation exists between the variables, but not necessarily a causal one. There are many possible explanations for the relationship, including (but not exclusively) demographic and economic differences among the countries. Correlation studies must be interpreted with caution, but important information can be obtained from the trends that could identify disparities and lead to further studies and hypothesis testing.

**Cross-Sectional Studies**

In *cross-sectional* studies, also known as prevalence studies, both exposures and outcomes are collected simultaneously. These studies provide a “snapshot” at one point in time and thus exclude people who have died or who chose not to participate, which can introduce bias. Temporal relationships are difficult to determine in these studies as only prevalence can be determined and the risk of developing disease cannot be estimated. Many cross-sectional studies are surveys that sample a population and its various characteristics. They can be inexpensive and can provide timely descriptive data about a group under study, but again, they do not tell us about causality or the true risk of developing a certain outcome such as disease.

Spoelstra, Given, von Eye, and Given (2010) conducted a cross-sectional study to determine whether individuals with a history of cancer fall at a higher rate than those without cancer. They also examined whether or not the occurrence of falls in the elderly was influenced by individual characteristics. The study population consisted of 7,448 community-dwelling elderly who were 65 years or older living in one state in the Midwestern United States. The analysis of the data revealed that having cancer was not a predictor of falls in this study. Further analysis revealed that predictors of falls in this population included race, sex, activities of daily living, incontinence, depression, and pain. Although
cancer was not found to be a predictor of falls, the authors did find a high frequency of falls in that study population. The findings led the authors to conclude that it is important to develop a predictive model for fall risk in the community-dwelling elderly.

This study serves to illustrate both the advantages and disadvantages of cross-sectional studies. The study was carried out at one point in time using an existing data set (the minimum data set). One limitation of the study was that it missed people whose falls were not reported. Another limitation the authors cited was that they could not determine whether a specific cancer diagnosis, stage, or treatment was a risk factor for falls. Finally, they were unable to determine whether or not comorbidities may have placed individuals at a higher risk for falls. The inability to control for or identify the significance of potentially important variables is a disadvantage of using a cross-sectional study design. With that said, a cross-sectional study is a fairly quick method to obtain descriptive data and can be useful in identifying prevalence rates for specified populations.

ANALYTIC EPIDEMIOLOGY

Analytic epidemiology looks at the origins and causal factors of diseases and other health-related events. Analytic designs are often carried out to test hypotheses formulated from a descriptive study. The goal of analytic epidemiology is to identify factors that increase or decrease risk. Risk is the probability that an event will occur. For example, a patient who is obese might ask, “What is the likelihood that I will develop diabetes if I do not lose weight?”

Although descriptive studies allow a basis for comparison and can provide the APRN with data to identify potential risk factors and differences among groups, study designs, such as a prospective cohort, need to be carried out in order to determine whether there is an association between an exposure and a disease and to determine the strength of that association. To do this, the APRN can compare exposed and nonexposed groups and follow them over time to see who develops an outcome (such as a specific disease) and who does not. Comparison is an essential component of population studies. Case–control studies can also allow for comparisons by retrospectively looking back in time to see what exposure or risk factors are associated with being a case or a control. Comparisons can also be made by following a group using treatment A compared to treatment B or treatment A can be compared to no treatment at all. There are multiple study designs, but we will focus only on the most common study designs and discuss the advantages and disadvantages that each one poses in practice.

Cohort Studies

*Co**hort* designs can be either prospective or retrospective. In a prospective cohort design, the investigator begins with a defined population and then follows a group of individuals who were either exposed or nonexposed to a factor of interest and then follows both groups to compare the incidence of an outcome or disease. In a cohort study, one can look at multiple outcomes that develop from an exposure. In a retrospective cohort design, exposure is ascertained from past records and outcome is ascertained at the time
the study begins. If an association exists between the exposure and the outcome, then the incidence rate in the exposed group will be greater than that in the nonexposed group. The ratio of these is the relative risk (RR), which is the incidence rate in the exposed group divided by the incidence rate in the nonexposed group. RR is a measure of the strength of an association between an exposure and an outcome or disease (Table 3.2).

If the RR = 1 (the numerator equals the denominator), then the risk to the two groups is equal. If the RR >1 (the numerator is greater than the denominator), then the risk in the exposed group is greater than the risk in the nonexposed group and can be considered a positive association. If the RR <1 (the denominator is greater than the numerator), then the risk in the exposed group is less than the risk in the nonexposed group and can be considered protective. An example of a protective association may be the association between exercise and heart disease. Exercise can actually reduce the risk of heart disease and has an RR <1. Thus, it is considered a protective exposure.

Attributable risk (AR), absolute risk, or risk difference is the amount of risk that can be attributed to an exposure. For example, it is well known that smoking can cause lung cancer, but lung cancer can also occur in nonsmokers. The amount of disease that is associated with risks/exposures other than smoking is called the background risk. In order to calculate the risk attributable to a particular exposure, subtract the incidence of disease (lung cancer) in the exposed group (smokers) minus the incidence of disease (lung cancer) in the nonexposed group (background risk). This value is considered the AR due to exposure (see Table 3.2). If an APRN wants to know how much risk of disease can be reduced by removing a risk factor, one can calculate the ARR, which is synonymous with the AR. The relative risk reduction (RRR) is calculated the same as the AR proportion. This can be confusing as these terms are interchanged in medicine and epidemiology, but it is important to recognize and understand how these terms are used and interpreted. An example of RRR would be described as: What percentage of motor vehicle deaths could be reduced if we could eliminate texting while driving? This RRR percentage is what is commonly reported in the news and can be very helpful for policy makers and for

| TABLE 3.2 Calculation of RR and Attributable Risk in a Cohort Study |
|---------------------------------|-----------------|-----------------|
| **DISEASE** | **NO DISEASE** | **TOTALS** |
| Exposure | a | b | a + b | Incidence in the exposed (Inc exp) = a/(a + b) |
| No Exposure | c | d | c + d | Incidence in the nonexposed (Inc nonexp) = c/(c + d) |

Relative risk (RR) = Inc exp/Inc nonexp

Attributable risk (AR) = Inc exp – Inc nonexp

AR proportion in the exposed population = (Inc exp – Inc nonexp) / Inc exp

AR proportion in the total population = (Incidence in total population – Inc nonexp) / Incidence in total population
justification of funding. The AR can also be calculated as a proportion of the total population. For example, to determine the amount of lung cancer attributable to smoking in the total population (AR proportion), one would have to know the incidence in the total population (to review how to calculate the incidence in the total population, refer to an advanced epidemiology textbook). APRNs should be familiar with how to calculate and interpret RR and AR, as these values are reported commonly in the literature and reports such as the *MMWR*.

Cohort studies are best carried out when the investigator has good evidence that links an exposure to an outcome, when the time interval between exposure and the outcome is short, and when the outcome occurs relatively often. One of the major problems with cohort studies is that they can be time-consuming and expensive, especially if the cohort needs to be followed for a prolonged length of time. Diseases that are rare or that take many years to develop may be better suited for a case–control study as it can be difficult to follow participants for many years, especially if the outcome of interest is rare. The longer the time period, the more likely participants will be lost to follow-up, and multiple exposures can potentially confound the relationship.

A cohort study was carried out in Norway to ascertain characteristics that would predict the risk of fibromyalgia (FM). The authors examined the association among leisure time, physical exercise, body mass index (BMI), and risk of FM (Mork, Vasseljen, & Nilsen, 2010). A longitudinal study followed 15,900 women without FM or physical impairment at baseline for 11 years. At the end of the study period, there were 380 reported cases of FM, and RRs were calculated for each of the study variables (exposures). Women who reported the highest exercise level had an RR of 0.77 (95% confidence interval [CI], 0.55–1.07). In looking at exercise, the authors controlled for the potential confounding factor of BMI. Overweight or obese women (BMI > 25.0 kg/m²) had a 60% to 70% higher risk of FM compared with women of normal weight (BMI = 18.5 to 24.9 kg/m²). In their study, overweight or obese women who exercised more than 1 hour/week had an RR of 1.72 (95% CI, 1.07–2.76), compared with normal-weight women with a similar activity level. The risk for overweight or obese women who were inactive (RR, 2.09; 95% CI, 1.36–3.21) or exercised less than 1 hour/week (RR, 2.19; 95% CI, 1.39–3.46) showed a positive association between risk of developing FM and low levels of exercise. The authors concluded that being overweight or obese was associated with an increased risk of FM, especially among women who also reported low levels of physical exercise. On the basis of these findings, they recommended that community-based measures aimed at reducing the incidence of FM should emphasize maintaining a normal weight and regular exercise.

**Case–Control Studies**

In a case–control study, the APRN must first identify a group of individuals with the outcome of interest (cases). A second group is identified without the outcome of interest (controls). The proportion of those cases that have a history of exposures is then compared to the proportion of the cases that were not exposed, and the proportion of the controls that were exposed is compared to the proportion of the controls that
were not exposed. The measure of the effect of exposure is expressed as an odds ratio (OR), which is the ratio of the odds of having been exposed if you are a case to the odds of having been exposed if you are not a case. If the exposure is not related to the disease or outcome, the OR = 1. If the exposure is related to the disease or outcome, the OR >1, and if the OR <1, the exposure is considered protective. To calculate the OR, construct a 2 × 2 table in which the columns represent the cases and controls and the rows represent the exposed and nonexposed populations. It is important to set up the table correctly. If it is not set up correctly, it will affect the interpretation and conclusions. Once the table is complete, multiply the cross products to obtain the result (see Table 3.3).

In a case–control study, if there is an association between an exposure and disease, the history of exposure should be higher in persons who have the disease (cases) compared to those who do not have disease (controls). It is important to keep in mind that the OR is not a calculation of risk and cannot predict which exposures/risk factors will develop into a case or disease. The fact that a person is obese may put that person at risk for diabetes, but it does not mean that that person will get diabetes. In case–control studies, one cannot calculate RR; therefore, we cannot conclude that if you are obese you will develop diabetes, but rather, if your OR is greater than 1, you could conclude that those with diabetes (outcome) are more likely to be obese (risk factor/exposure).

Selection of cases and controls is an important step in case–control studies. Definite criteria should be used so that there is no ambiguity about how to distinguish between a case and a control. Exposure is not always all or nothing. Controls should resemble the cases as closely as possible except for the exposure to the factor under study. If the cases are drawn from a particular clinic, then ideally the controls should be drawn from the same clinic population. Matching is one method that can be used to select a sample so that potential confounders are distributed equally between the cases and controls. For example, if an APRN plans to evaluate an intervention to reduce burden among caregivers of dependent elderly in the home, it would be important to recognize the characteristics of the population studied prior to implementing the intervention. It is known that men and women have differing characteristics that affect their role as caregiver (Amankwaa, 2017). By matching for gender in the study, the APRN can eliminate this potentially confounding factor (gender). The problem with matching is that the investigator is not

TABLE 3.3 Calculation of OR in a Case–Control Study

<table>
<thead>
<tr>
<th>CASES</th>
<th>CONTROLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure history</td>
<td>a</td>
</tr>
<tr>
<td>No exposure history</td>
<td>c</td>
</tr>
<tr>
<td>Totals</td>
<td>a + c</td>
</tr>
</tbody>
</table>

Proportion of cases exposed = $a/(a + c)$
Proportion of controls exposed = $b/(b + d)$

OR = $ad/bc$
always aware of all of the potential confounding factors. It can be difficult to match each subject in a study, and in some cases, investigators can overmatch. When an investigator overmatches, one loses the ability to look at the matched variables as risk factors.

In case–control studies, the investigator begins with cases and controls and goes back retrospectively to look for exposures. In cohort studies, the investigator begins with exposed and nonexposed individuals and follows individuals over time to see who develops or does not develop an outcome or disease. Case–control studies allow the APRN to look at cases and the probability of having an exposure or risk factor. Cohort studies allow an APRN to follow a cohort over time to determine whether being exposed to a risk factor impacts the likelihood of developing a disease or diseases, or improves outcomes (as in an intervention). If associations are found, further studies are necessary to determine causal links and to prevent ecologic fallacy. When examining the results of case–control and cohort studies, it is important for the APRN to consider whether or not all other explanations for an identified association have been eliminated. No single epidemiological study can satisfy all criteria for causality. The APRN needs to look at the accumulation of evidence, as well as the strength of individual studies.

Randomized Controlled Trials

Randomized controlled trials (RCTs) or clinical trials are useful for evaluating treatments (including technology) and for assessing new ways of organizing and delivering health services. In population-based studies, the issue is often health promotion and disease prevention rather than treatment of an existing disease. Interventions can also be studied in RCTs, with the target involving defined populations rather than individuals and often involving educational, program, or policy interventions. When carefully designed, RCTs can provide the strongest evidence for evaluating treatments and interventions.

The basic design of an RCT is to assign subjects randomly to either receive the new treatment/intervention or not receive the new treatment/intervention. Inclusion and exclusion criteria for the participants must be precise and written in advance to eliminate any errors within the study or any future comparison studies. As with cohort studies, RCTs can compare more than two groups. Analysis is carried out to compare outcomes between the randomized groups. As mentioned earlier, comparisons can be made between different interventions, different treatments, or to a control group that has received no intervention or treatment.

There are many examples of how RCTs have been used to evaluate the effectiveness of interventions in specific groups of people. For example, they have been used to evaluate the effectiveness of the Orem self-care model on pain relief in people with rheumatoid arthritis (Saeedifar, Memarian, Fatahi, & Ghelichkhani, 2018), and the use of a community-based skill building intervention to improve heart failure management in community dwelling adults (Dickson, Melkus, Dorsen, Katz, & Riegel, 2015), as well as to test the impact of a nurse-delivered intervention to reduce intimate partner violence (Gupta et al., 2017). These studies illustrate how useful the randomized trial design is for testing a new intervention. RCTs are an excellent vehicle to provide evidence to enhance practice.
Sample Size

Sample selection and sample size determination are critical steps in the research process. Sample size determination is necessary to identify the minimum number of subjects needed to enroll in a study to identify true differences and associations between groups and thus has implications for the investigators as they need to allocate ample resources based on sample size to carry out the study. Power analysis is used to determine sample size. There are several factors that influence the size of the sample: variance, significance, power, and effect size.

Variance

Variance is the variation about the mean. For example, if you are looking at a continuous variable such as blood pressure, the variance away from the mean is defined as $s^2$. Variance ($s^2$) is the square of the standard deviation ($s$). The standard deviation takes into account all blood pressure measurements and essentially sums the difference of each blood pressure measurement away from the mean (see a statistic textbook for more details). If you do not have data with which to calculate the standard deviation, you can review the literature or look at a pilot data set to determine this number. A study that has very little variance (i.e., most of the values fall close to the mean) would require a smaller sample size than a study in which the blood pressure measurements have a very large range and a wider sigmoid curve.

Significance and Power

Significance is the probability that an observed difference or relationship exists and usually is defined as a $p$-value ($p < 0.05$). The smaller the $p$-value (e.g., $p < 0.01$), the larger the required sample size. Power $(1 - \beta)$ is the capacity of the study to detect differences or relationships that actually exist in the population or the capacity to correctly reject a null hypothesis, that is, prevent a type II error. The larger the power required, the less likelihood of committing a type II error. Most studies use a power of 80% or 0.80, or, if a more rigorous power is necessary (e.g., 90%), a larger sample size is required.

Effect Size

Effect size is the actual difference between groups and treatments that you hope to see in your study. One way to identify effect size is to review previous studies; another method is to conduct a pilot study. For example, if you are designing an educational intervention and want to see a 20% improvement of knowledge after the intervention, 20% is your effect size. If you want to see a smaller change in knowledge (e.g., 10%), then a larger sample will be required to detect a smaller effect or difference. In summary, effect sizes occur along a range of values. For example, if you want to see a 5% change in results of an outcome, you will need many more participants than if you want to see a 30% change.

Understanding what goes into power analysis is an important step in designing a study. Power analysis can be calculated using computer programs. There are many free software programs available on the Internet to assist with power analysis. Typing “sample size calculation” in a search engine such as Yahoo or Google will lead an APRN to many sites.
Screening

Screening is a tool used to detect disease in groups of asymptomatic individuals with the goal of reducing and/or preventing morbidity and mortality. Screening tests can be applied to groups of individuals or to high-risk populations. There are multiple examples of screening tests, including the Pap smear, the tuberculosis skin test (PPD test), the mammogram, and so on.

Determining whether a screening test is appropriate requires the APRN to address several aspects of the disease of interest. Screening is neither available nor appropriate for all diseases. In order for a screening program to be effective, certain criteria should be met. The target population needs to be identifiable and accessible and the disease should affect a sufficient number of people to make screening cost-effective. The preclinical period should be sufficient to allow treatment before symptoms appear so that early diagnosis and treatment make a difference in terms of outcome.

Finally, it is necessary for the screening test to be sensitive enough to detect most cases of the disease and to be specific enough to limit the number of false-positive tests. Screening tests should also be relatively inexpensive, easy to administer, and have minimal side effects.

The validity of a screening test refers to its ability to accurately identify those who have the disease. Sensitivity and specificity are measures of a screening test’s validity. Sensitivity is a measure of a screening test’s ability to accurately identify disease when it is present. Specificity is a measure of a screening test’s ability to correctly identify a person without disease with a negative test. The positive predictive value (PPV) is a measure of the probability of a positive test result when the disease is present. The negative predictive value (NPV) of a test is a measure of the probability that the disease is absent when there is a negative test (see Table 3.4).

Directing screening tests toward high-risk populations has many advantages. By screening populations with a higher disease prevalence, we can actually increase the PPV of that test. Screening low-prevalence populations can lead to more false positives, which can be costly and harmful to patients. Thus, selection of the disease to be tested and the patient population to be screened are both important to consider when designing a new test.

The APRN can evaluate the success of screening programs by looking at a variety of outcomes. For example, some of the outcomes that can be followed include

<table>
<thead>
<tr>
<th></th>
<th>DISEASE</th>
<th>NO DISEASE</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Test</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>− Test</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Totals</td>
<td>a + c</td>
<td>b + d</td>
<td>a + b</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>a/a + c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>d/b + d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3.4 Computing Sensitivity, Specificity, and Predictive Values in Screening Tests
the reduction in overall mortality in screened individuals, a reduction in the CFR in screened individuals, an increase in the percentage of cases detected at earlier stages, a reduction in complications, and improvement of quality of life in screened individuals.

In 2018, the U.S. Preventive Services Task Force (USPSTF) released their final recommendations for cervical cancer screening (USPSTF, 2018). The complete recommendations can be found on the USPSTF web site (www.uspreventiveservicestaskforce.org/Page/Name/us-preventive-services-task-force-issues-new-cervical-cancer-screening-recommendations). One of the recommendations for cervical cancer screening is that women aged 21 to 65 should get a Pap smear every 3 years. A second recommendation is that women aged 30 to 65 who wish to be screened less frequently can choose a combination Pap smear and human papillomavirus (HPV) testing every 5 years. The Task Force does not recommend cervical cancer screening using HPV testing in women younger than age 30. This is because evidence indicates that the expected harms (such as false positives) in this age group outweigh the potential benefits.

According to the Task Force, “since the implementation of widespread cervical cancer screening, there has been a dramatic reduction in cervical cancer deaths in the United States” (USPSTF, P4). For this reason, the Task Force urges healthcare providers to encourage women to be screened for cervical cancer, especially those who have never been screened, or who have not been screened within the past 5 years. These guidelines provide an example of how evidence on the specificity and sensitivity of a screening test can be used to create more evidence-based clinical guidelines. Therefore, screening tests need to be tailored to the disease under investigation, and many factors need to be taken into consideration; for example, How many false negatives can be missed? How many false positives are acceptable? Can screening and early detection really make a difference in the outcome of the disease? Understanding these factors and balancing them with targeted screening in high-risk populations are important considerations in screening implementation. The USPSTF provides screening guidelines on a variety of disease states with recommendations (e.g., colorectal, prostate, and breast cancers). APRNs can visit their website located at https://www.uspreventiveservicestaskforce.org for the latest recommendations.

**SUMMARY**

The natural history of disease refers to the progression of a disease from its preclinical state to its clinical state, and knowledge of these stages provides a framework for understanding approaches to the prevention and control of disease. Primary prevention refers to the process of altering susceptibility or reducing exposure to susceptible individuals and includes general health promotion and specific measures designed to prevent disease prior to a person getting a disease. Primary prevention measures are generally carried out during the stage of susceptibility. With secondary prevention, it is sometimes possible to either cure a disease at a very early stage or slow its progression to prevent complications and limit disability. Secondary prevention measures are carried out during the preclinical or presymptomatic stage of disease. Tertiary
prevention takes place during the middle or later stages of a disease (the clinical stage of disease) and refers to measures taken to alleviate disability and restore effective functioning.

The dynamic nature of disease calls for a sophisticated model for explaining causation. When designing interventions for populations, the APRN needs to keep in mind that disease develops as the result of many antecedent factors and not as a result of a single, isolated cause.

Descriptive epidemiology is used to describe the distribution of disease and other health-related states and events in terms of personal characteristics, geographical distribution, and time. It also helps APRNs design studies and measure mortality and prognosis. Analytic epidemiology looks at the origins and causal factors of diseases and other health-related events. Epidemiologic methods can be used to identify populations at risk and to evaluate interventions provided to patient populations. Population-based evaluation and planning depend on understanding the many and varied factors that influence health and disease. APRNs can use their understanding of epidemiological methods in concert with their clinical expertise to develop policies and implement and evaluate new programs and interventions to improve population outcomes.

**EXERCISES AND DISCUSSION QUESTIONS**

**Exercise 3.1** In 2018, there were 105 new cases of type 2 diabetes reported in Smithville, a city of 500,000. This brought the total number of active cases of type 2 diabetes in Smithville to 3,075. During this time, there were 105 deaths attributable to the disease.

a. What was the incidence rate per 100,000 for type 2 diabetes in 2018?

b. What was the prevalence rate of type 2 diabetes per 100,000 in 2018?

c. What was the cause-specific death rate of type 2 diabetes in 2018?

**Exercise 3.2** A city contains 100,000 people (45,000 males and 55,000 females), and 1,000 people die per year (600 males and 400 females). There were 50 cases (40 males and 10 females) of lung cancer per year, of whom 45 died (36 males and 9 females).

Using this information compute:

a. The crude mortality rate per 1,000

b. The sex-specific mortality rate per 1,000

c. The cause-specific mortality rate per 1,000 for lung cancer

d. The case fatality rate for lung cancer

e. The proportionate mortality ratio for lung cancer

**Exercise 3.3** A new rapid blood test was created to test for HPV in a rural clinic. The following is a 2 × 2 chart which describes the results of the test. Answer questions a to g using the 2 × 2 chart.
<table>
<thead>
<tr>
<th></th>
<th>HPV</th>
<th>NO HPV</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (+) test</td>
<td>95</td>
<td>37</td>
<td>132</td>
</tr>
<tr>
<td>Negative (−) test</td>
<td>39</td>
<td>278</td>
<td>317</td>
</tr>
<tr>
<td>Totals</td>
<td>134</td>
<td>315</td>
<td>449</td>
</tr>
</tbody>
</table>

a. What is the sensitivity of this test?
b. What is the specificity of this test?
c. What is the positive predictive value?
d. What is the negative predictive value?
e. Describe in words the sensitivity of this test.
f. Describe in words the negative predictive value.
g. What is the disease prevalence in this population?

**Exercise 3.4** An epidemiologic study is conducted to learn about the relationship between celiac disease and colon cancer. Suppose there are 77 cases of colon cancer in 68,000 person-years in persons with celiac disease and 54 cases of colon cancer in 215,000 person-years in those without celiac disease. (The overall rate in both groups combined = 131 cases in 283,000 person-years overall.) Use this information to answer questions a to c.

a. Calculate the rate of colon cancer in the celiac group \( R_1 \), in the no celiac group \( R_0 \), and overall \( R \). Express all rates “per 100,000 person-years.”

\[
\begin{align*}
R_1 & = \\
R_0 & = \\
R & = 
\end{align*}
\]

b. Calculate and **interpret** the relative risk of colon cancer associated with celiac disease.

c. Calculate and **interpret** the attributable risk of colon cancer associated with celiac disease.

**Exercise 3.5**

a. What is the fundamental difference between a case–control study and a cohort study?

b. What are the advantages and the disadvantages of a cross-sectional study?

c. What are the characteristics of a correlational study?
d. You read about a new protocol (HTN T) that was used successfully to improve blood pressure control in an urban clinic population. A major feature of HTN T is the use of technology (such as text messaging) to communicate with patients. You work in a rural clinic and currently use protocol HTN 1. You wonder if HTN T would improve blood pressure control in your clinic population. Explain how you could use a randomized clinical trial to evaluate whether HTN T improves blood pressure control in your clinic population.

Exercise 3.6 You are reviewing the survival statistics from your hospital using a new treatment (treatment A) compared to the old treatment (treatment B) for breast cancer. The following table lists the number of survivors after each year of treatment for both treatments A and B. Answer the following questions using the table. Assume no patients were lost to follow-up.

<table>
<thead>
<tr>
<th>TREATMENT A</th>
<th>NUMBER SURVIVED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AT 1 YEAR</td>
</tr>
<tr>
<td>Cohort (N = 1,229)</td>
<td>1,102</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TREATMENT B</th>
<th>NUMBER SURVIVED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AT 1 YEAR</td>
</tr>
<tr>
<td>Cohort (N = 1,179)</td>
<td>1,084</td>
</tr>
</tbody>
</table>

- Calculate the survival rates for treatments A and B for each year after treatment: P1, P2, P3, P4, and P5.
- Calculate the probability of surviving for 1, 2, 3, 4, and 5 years cumulatively for each of the treatments.
- Your administrator would like to know how the treatments compared to each other. You are asked for the following information:
  - What is the likelihood of surviving 5 years if you made it to 4 years of treatment for each of the treatments?
  - How does treatment A compare to treatment B for each year of survival after treatment?
  - Which treatment has the best 5-year survival rate?
- Plot the survival curve for both treatments on the same graph.
- Why might there be differences between these two treatments?
- What are some potential confounders that may contribute to one treatment working better than the other?
- Explain the advantages and disadvantages of screening tests.
- What are the limitations to performing survival analysis?
- How might patients lost to follow-up affect the validity of survival analysis?
- What are the differences between the actuarial method of survival analysis and the Kaplan–Meier method?

**Exercise 3.7**

- Perform a search to determine the incidence, prevalence, and survival rates for one of the following cancers in your state (lung cancer, breast cancer, colon cancer, prostate cancer, cervical cancer)
  - Perform a search and describe the screening recommendations for the cancer you selected.
  - Describe the advantages and disadvantages of cancer screening for the cancer you selected.
  - How does the incidence, prevalence, and survival rates in your state compared to the national rates.
  - What are potential confounders for the cancer you selected?
  - What disparities in cancer incidence, mortality, and survival were you able to determine from your search?

**3.8 What factor has the single biggest impact on mortality?**

**REFERENCES**


**INTERNET RESOURCES**

CDC, Health Impact Assessment: www.cdc.gov/healthyplaces/hia.htm

Morbidity and Mortality Weekly Report: www.cdc.gov/mmwr

U.S. Preventive Services Task Force: https://www.uspreventiveservicestaskforce.org