Mina D. Singh
Lorraine Thirsk

LoBiondo-Wood and Haber's

NURSING RESEARCH IN CANADA

Methods, Critical Appraisal, and Utilization

Fifth Edition

ELSEVIER

Evolve®

Student Resources on Evolve
Details Inside

Evolve®

YOU'VE JUST PURCHASED MORE THAN A TEXTBOOK!

Enhance your learning with Evolve Student Resources.

These online study tools and exercises can help deepen your understanding of textbook content so you can be more prepared for class, perform better on exams, and succeed in your course.

Activate the complete learning experience that comes with each NEW textbook purchase by registering with your scratch-off access code at

http://evolve.elsevier.com/Canada/LoBiondo/research

If your school uses its own Learning Management System, your resources may be delivered on that platform. Consult with your instructor.

If you rented or purchased a used book and the scratch-off code at right has already been revealed, the code may have been used and cannot be re-used for registration. To purchase a new code to access these valuable study resources, simply follow the link above.

Place Sticker Here

REGISTER TODAY!

ELSEVIER

You can now purchase Elsevier products on Evolve!
Go to evolve.elsevier.com/shop to search and browse for products.

2019v1.0

Fifth Edition

LoBiondo-Wood and Haber's

NURSING RESEARCH IN CANADA

Methods, Critical Appraisal, and Utilization

MINA D. SINGH, RN, RP, BSc, BScN, MEd, PhD, I-FCNEI
Professor
School of Nursing
Faculty of Health, York University
Toronto, Ontario
Canada

LORRAINE M. THIRSK, RN, PhD
Assistant Professor
Faculty of Health Disciplines
Athabasca University
Athabasca, Alberta

Associate Editors

SARAH STAHLKE, BScN, MHSA, PhD
Adjunct Professor/Associate Lecturer
Department of Sociology
University of Alberta; Edmonton, Alberta
Sessional Instructor
Faculty of Health Disciplines
Athabasca University
Athabasca, Alberta

RAMESH VENKATESA PERUMAL, RN, MSc(N), CCNE, CNCC®, PhD(c)
Professor
School of Community and Health Studies
Centennial College;
Assistant Professor
School of Nursing
Faculty of Health
York University
Toronto, Ontario

US Authors

GERI LOBIONDO-WOOD, PhD, RN, FAAN
Professor, Director of PhD Nursing Program
University of Texas Health Sciences Center
School of Nursing, Nursing Systems and Technology
Houston, Texas

JUDITH HABER, PhD, RN, FAAN
The Ursula Springer Leadership Professor in Nursing
Rory Meyers College of Nursing
New York University
New York, New York

ELSEVIER

Elsevier

LOBIONDO-WOOD AND HABER'S NURSING RESEARCH IN CANADA:
METHODS, CRITICAL APPRAISAL, AND UTILIZATION, FIFTH EDITION ISBN: 978-0-323-77898-5

ISBN: 978-0-323-76291-5 (softcover).

Library of Congress Control Number: 2021946978

Managing Director, Global ERC: Kevonne Holloway
Senior Content Strategist (Acquisitions, Canada): Roberta A. Spinosa-Millman
Content Development Specialist: Theresa Fitzgerald
Publishing Services Managers: Shereen Jameel/Deepthi Unni
Project Managers: Rukmani Krishnan/Manchu Mohan
Design Direction: Renee Duenow

Last digit is the print number: 9 8 7 6 5 4 3 2 1

Contents

Appendices

Author Biographies

Mina D. Singh, RN, RP, BSc, BScN, MEd, PhD, I-FCNEI, is a Professor at the School of Nursing, York University. She has a long career in nursing, having worked in acute care, mental health, psychotherapy, and public health. Her expertise is as a research methodologist, statistician, and program evaluator. In addition, she is a psychotherapist. She won the 2012 National Nursing Research Scholar award and the 2014 Accreditation Reviewer Excellence award, both granted by the Canadian Association of Schools of Nursing (CASN). She was an Accreditation Reviewer for CASN for over 18 years and has travelled internationally and nationally reviewing nursing education programs.

Lorraine M. Thirsk, RN, PhD, is an Assistant Professor in the Faculty of Health Disciplines, Athabasca University. Over the last 20 years, she has worked in rural and tertiary hospitals, home care, and palliative care. In advanced practice, she worked as a community therapist, nurse consultant, and a clinical nurse specialist. Her research program focuses on family nursing interventions in adult populations facing serious illness. As an educator, she is interested in supporting undergraduate nursing students to develop competencies in working with families and supporting graduate students from across the health disciplines in the research process. In addition, she is passionate about supporting and advancing evidence-informed nursing practice. Methodologically, she is interested in mixed methods and using qualitative research to understand complex interventions.

Sarah Stahlke, BScN, MHSA, PhD, is a sociologist and a nurse. As a clinician, her experience is in acute care, mainly in the operating room and intensive care. After moving into leadership and administrative roles, she worked in program development, hospital management, and professional regulation. She is currently an adjunct professor and lecturer in sociology at the University of Alberta and an instructor in the nursing/health studies master's degree program at Athabasca University. Her research and teaching interests relate to nursing roles, nursing practice environments, employee engagement, organizational studies, professional regulation, and interprofessional relationships. Sarah is an internationally recognized qualitative methodologist and has taught numerous courses and workshops on qualitative research methods.

Ramesh Venkatesa Perumal, RN, MSc(N), CCNE, CNCC®, PhD(c), is a Professor at Centennial College and an Assistant Professor at York University. In addition to being a Critical Care Nurse, Ramesh has more than 20 years of experience in teaching, research, and community service. He has served on the editorial boards of nursing journals and has been a peer reviewer of nursing journals. He won a prestigious award for the best teacher (Award for Excellence in Teaching) in 2009 at a public university in Muscat, Sultanate of Oman. His research and teaching interests relate to mentorship in nursing, Internationally Educated Nurses (IEN), and Relational Care in intensive care units.

Geri LoBiondo-Wood, PhD, RN, FAAN, is Professor and Director of the PhD in Nursing Program at the University of Texas Health Science Center at Houston, School of Nursing (UTHSC-Houston) and former Director of Research and Evidence-Based Practice Planning and

Development at the MD Anderson Cancer Center, Houston, Texas. She received her Diploma in Nursing at St. Mary's Hospital School of Nursing in Rochester, New York; Bachelor's and Master's degrees from the University of Rochester; and a PhD in Nursing Theory and Research from New York University. Dr. LoBiondo-Wood teaches research and evidence-based practice principles to undergraduate, graduate, and doctoral students. At MD Anderson Cancer Center, she developed and implemented the Evidence-Based Resource Unit Nurse (EB-RUN) Program, a hospital-wide program that involves all levels of nurses in the application of research evidence to practice. She has extensive national and international experience guiding nurses and other health care professionals in the development and utilization of research. Dr. LoBiondo-Wood is an editorial board member of *Progress in Transplantation* and a reviewer for *Nursing Research, Oncology Nursing Forum, Oncology Nursing,* and *Nephrology Nursing Journal*. Her research and publications focus on chronic illness and oncology nursing.

Dr. LoBiondo-Wood has been active locally and nationally in many professional organizations, including the Oncology Nursing Society, Southern Nursing Research Society, the Midwest Nursing Research Society, and the North American Transplant Coordinators Organization. She has received local and national awards for teaching and contributions to nursing. In 1997, she received the Distinguished Alumnus Award from New York University, Division of Nursing Alumni Association. In 2001, she was inducted as a Fellow of the American Academy of Nursing and in 2007 as a Fellow of the University of Texas Academy of Health Science Education. In 2012, she was appointed as a Distinguished Teaching Professor of the University of Texas System.

Judith Haber, PhD, RN, FAAN, is the Ursula Springer Leadership Professor in Nursing at the Rory Meyers College of Nursing, at New York University. She received her undergraduate nursing education at Adelphi University in New York, and she holds a Master's degree in Adult Psychiatric–Mental Health Nursing and a PhD in Nursing Theory and Research from New York University. Dr. Haber is internationally recognized as a clinician and educator in psychiatric–mental health nursing. She has extensive clinical experience in psychiatric nursing, having been an advanced practice psychiatric nurse in private practice for over 30 years, specializing in treatment of families coping with the psychosocial sequelae of acute and chronic catastrophic illness. Her NIH-funded program of research addressed physical and psychosocial adjustment to illness, focusing specifically on women with breast cancer and their partners and, more recently, breast cancer survivorship. Dr. Haber is also committed to an interprofessional program of clinical scholarship related to improving oral-systemic health outcomes and leads the *Oral Health Nursing Education and Practice (OHNEP)* program funded by the DentaQuest Foundation, as well as the HRSA-funded *Teaching Oral Systemic Health (TOSH)* program.

Dr. Haber has been active locally and nationally in many professional organizations, including the American Nurses Association, the American Psychiatric Nurses Association, and the American Academy of Nursing. She has received numerous local, state, and national awards for public policy, clinical practice, and research, including the APNA Psychiatric Nurse of the Year Award in 1998 and 2005 and the APNA Outstanding Research Award in 2005. She received the 2007 NYU College of Nursing Distinguished Alumnus Award, the 2011 NYU Distinguished Teaching Award, and the 2013 NYU Alumni Meritorious Service Award. In 1993, she was inducted as a Fellow of the American Academy of Nursing and in 2012 as a Fellow in the New York Academy of Medicine.

Contributors

Megan Aston, RN, PhD
Professor
School of Nursing
Dalhousie University
Halifax, Nova Scotia

Luisa Barton, NP-PHC, BScN, MN, DNP
Assistant Professor
Faculty of Health Disciplines
Athabasca University
Athabasca, Alberta

Caroline Foster-Boucher, RN, MN
Assistant Professor
Department of Nursing Science
MacEwan University
Edmonton, Alberta

Mélanie Lavoie-Tremblay, BScN, MScN, PhD
Associate Professor
Ingram School of Nursing
McGill University
Montréal, Quebec

Shannon MacDonald, RN, PhD
Assistant Professor
Faculty of Nursing
University of Alberta
Edmonton, Alberta

Martha MacLeod, RN, PhD
Professor
School of Nursing & School of Health Sciences
University of Northern British Columbia
Prince George, British Columbia

Christine Maheu, PN, PhD
Associate Professor
Ingram School of Nursing
McGill University
Montreal, Quebec

Ruth Rodney, RN, PhD
Assistant Professor
School of Nursing
York University
Toronto, Ontario

Evelyn Voyageur, RN, PhD (psychology)
Elder in Residence, Aboriginal Education
North Island College, Comox Valley Campus
Courtney, British Columbia

Reviewers

Miranda Bevilacqua, RN, BScN, MN, PhD(c)
Professor
School of Health, Negahneewin, & Community Services
Confederation College
Thunder Bay, Ontario

Ellen Buck-McFadyen, RN, MScN, PhD
Assistant Professor
Trent/Fleming School of Nursing
Trent University
Peterborough, Ontario

Tracy M. Christianson, RN, BSN, MN, DHEd, CCNE
Assistant Professor
School of Nursing
Thompson Rivers University
Kamloops, British Columbia

Angela J. Gillis, RN, PhD
Professor Emeritus
School of Nursing
Saint Francis Xavier University
Antigonish, Nova Scotia

Tracy Hoot RN, BScN, MSN, DHEd
Associate Dean
School of Nursing
Thompson Rivers University
Kamloops, British Columbia

Laura Anne Killam, BScN, RN, MScN
Professor
Health Sciences, Nursing & Emergency Services
Cambrian College
Sudbury, Ontario

Tsorng-Yeh Lee, RN, PhD
Associate Professor
School of Nursing
York University
Toronto, Ontario

Catherine (Cathy) Doreen MacDonald, BScN, RN, MN, PhD
Associate Dean Academic Affairs, Arts & Science
Arts & Science & Nursing
Rankin School of Nursing
Saint Francis Xavier University
Antigonish, Nova Scotia

Elise J. Matthews, BScN, BA, RN, PhD
Assistant Professor
Faculty of Nursing
University of Regina
Regina, Saskatchewan

Christie Pettipas, BSW, MPA, EdD
Instructor
School of Community Studies
Bow Valley College
Calgary, Alberta

Cheryl Pollard, RPN, RN, PhD
Dean
Faculty of Nursing
University of Regina
Regina, Saskatchewan

Nancy A. Sears, RN, BNSc, MPA, CHE, PhD
Professor
School of Baccalaureate Nursing
St. Lawrence College (Laurentian University collaborative BScN program)
Kingston, Ontario

Candice Waddell, RPN, BScPN, MPN, PhD(c)
Assistant Professor
Department of Psychiatric Nursing
Faculty of Health Studies
Brandon University
Brandon, Manitoba

Karla L. Wolsky, RN, PhD
Nursing Instructor
Nursing Centre for Health and Wellness
Lethbridge College
Lethbridge, Alberta

Acknowledgements

THIS MAJOR UNDERTAKING WAS ACCOMPLISHED WITH the help of many people, some of whom made direct contributions to this new edition and some of whom contributed indirectly. We acknowledge with deep appreciation and our warmest thanks the following people who made this fifth Canadian edition possible:

- The Inuit, First Nations, and Métis people who first lived on these lands and whose ancestors shared their knowledge with the settlers
- Nursing educators across Canada who provided valuable and insightful comments that helped to direct the revisions featured in this edition and contributed to improving the content
- Our students, particularly past and present nursing students at York University, University of Alberta, and Athabasca University, who inspired us with their feedback and ideas
- Roberta A. Spinosa-Millman, Senior Content Strategist, who got us started with encouragement, a sense of humour, and great insight
- Theresa Fitzgerald, Content Development Specialist, who encouraged us with positive feedback, made sense of the process, and understood writing during the pandemic would mean extending deadlines
- Our vignette contributors, whose willingness to share their wisdom and evidence of their innovative research made a unique contribution to this edition
- All of the reviewers, who provided thoughtful feedback not only on the previous four editions but also on the fifth Canadian edition manuscript
- Our families, who supported us and picked up the "loose ends" while we wrote and revised.

To my husband Neranjan, my daughter Sandhya, and my parents Ram and Betty Laljie, for their support and encouragement.

Mina D. Singh

To my husband Reidar Hagtvedt, my sons Benjamin and Alexander, and my parents Theo and Margot Thirsk. Completing this fifth edition, during the pandemic, would not have been possible without your support.

Lorraine M. Thirsk

Preface

THE FOUNDATION OF THE FIFTH CANADIAN edition of *LoBiondo-Wood and Haber's Nursing Research in Canada: Methods, Critical Appraisal, and Utilization* continues to be the belief that nursing research is integral to all levels of nursing education and practice. Since the first edition of this textbook, we have seen the depth and breadth of nursing research grow. More nurses are conducting research and using research evidence to shape clinical practice, education, administration, and health policy.

The Canadian Nurses Association promotes the notion that nurses must provide care that is based on the best available scientific evidence. This is an exciting challenge to meet. Nurses are using the best available evidence, combined with their clinical judgement and patient preferences, to influence the nature and direction of health care delivery and to document outcomes related to the quality and cost-effectiveness of patient care. As nurses continue to develop a unique body of nursing knowledge through research, decisions about clinical nursing practice will be increasingly evidence informed.

As editors, we believe that all nurses not only need to understand the research process but also need to know how to critically read, evaluate, and apply research findings in practice. We realize that understanding research, as a component of evidence-informed practice, is a challenge for every student, but we believe that the challenge can be accomplished in a stimulating, lively, and learner-friendly manner.

Consistent with this perspective is a commitment to advancing implementation of the evidence-informed practice paradigm. Understanding and applying nursing research must be an integral dimension of nursing education, evident not only in the undergraduate nursing research course but also throughout the curriculum. The research role of nurses calls for evidence-informed practice competencies; central to this are critical appraisal skills—that is, nurses should be competent in using research.

Preparing students for this role involves developing their critical thinking and reading skills, thereby enhancing their understanding of the research process and their ability to appraise research critically. An undergraduate course in nursing research should develop this basic level of competence, which is an essential requirement if students are to engage in evidence-informed clinical decision-making and practice. This contrasts with a graduate-level research course, in which the emphasis is on conducting research, as well as understanding and appraising it.

The primary audience for this textbook remains undergraduate students who are learning the steps of the research process, as well as how to develop clinical questions, critically appraise published research literature, and use research findings to inform evidence-informed clinical practice. This book is also a valuable resource for students at the Master's and doctoral levels who want a concise review of the basic steps of the research process, the critical appraisal process, and the principles and tools for evidence-informed practice.

This text is also a key resource for health care leaders who are preparing to be experts at evidence-informed initiatives in clinical settings. Furthermore, it is an important resource for practising nurses who strive to use research evidence as the basis for clinical decision-making and development of evidence-informed policies, protocols, and standards, rather than rely on tradition, authority, or trial and error. It is an

important resource for nurses who collaborate with nurse-scientists in the conduct of clinical research and evidence-informed practice.

Building on the success of the fourth edition, we maintain our commitment to introduce evidence-informed practice and research principles to baccalaureate students, thereby providing a cutting-edge research consumer foundation for their clinical practice.

Knowledge and language concerning sex, gender, and identity are fluid and continually evolving. The language and terminology presented in this text endeavours to be inclusive of all people and reflects what is to the best of our knowledge current at the time of publication.

LoBiondo-Wood and Haber's Nursing Research in Canada: Methods, Critical Appraisal, and Utilization prepares nursing students and practising nurses to become knowledgeable nursing research consumers in the following ways:

- Addressing the evidence-informed practice role of the nurse, thereby embedding evidence-informed competence in the clinical practice of every baccalaureate graduate.
- Demystifying research, which is sometimes viewed as a complex process.
- Using an evidence-informed approach to teaching the fundamentals of the research process.
- Teaching the critical appraisal process in a user-friendly but logical and systematic progression.
- Promoting a lively spirit of inquiry that develops critical thinking and critical reading skills, facilitating mastery of the critical appraisal process.
- Developing information literacy, searching, and evidence-informed practice competencies that prepare students and nurses to effectively locate and evaluate the best available research evidence.
- Elevating the critical appraisal process and research appreciation to a position of importance comparable to that of producing research. Before students become research producers, they must become knowledgeable research consumers.
- Emphasizing the role of evidence-informed practice as the basis for informing clinical decision-making and nursing interventions that support nursing practice, demonstrating quality and cost-effective outcomes of nursing care delivery.
- Presenting numerous examples of recently published research studies that illustrate and highlight each research concept in a manner that brings abstract ideas to life for students new to the research and critical appraisal process.
- These examples are a critical link for reinforcement of evidence-informed concepts and the related research and critiquing process.
- Showcasing, in **Research Vignettes,** the work of renowned nurse researchers whose careers exemplify the links among research, education, and practice.
- Providing numerous pedagogical chapter features, including **Learning Outcomes, Key Terms, Key Points,** new **Critical Thinking Challenges, Research Hints, Evidence-Informed Practice Tips,** new **Practical Applications**, revised **Critical Thinking Decision Paths,** and **Critical Judgement Questions,** as well as numerous tables, boxes, and figures. At the end of each chapter that presents a step of the research process, we feature a section titled **Appraising the Evidence,** which reviews how each step of the research process should be evaluated from a consumer's perspective. This section is accompanied by an updated **Critiquing Criteria** box.
- Providing a **Study Guide** that promotes active learning and assimilation of nursing research content.
- Offering an Evolve site presenting free **Evolve Resources for Instructors** that includes a Test Bank, TEACH, PowerPoint slides, critiquing exercises, an Image Collection, and critical appraisal activities. There are also Evolve

resources for both the student and faculty that include an audio glossary.

The fifth Canadian edition of *LoBiondo-Wood and Haber's Nursing Research in Canada: Methods, Critical Appraisal, and Utilization* is organized into seven parts. Each part is preceded by an introductory section and opens with an exciting "Research Vignette" by a renowned nurse researcher.

Part One, Research Overview, contains six chapters. Chapter 1, "The Role of Research in Nursing," provides an excellent overview of research and evidence-informed practice processes that shape clinical practice. This chapter introduces the role that research plays in practice and education, the roles of nurses in research activities, a historical perspective, and future directions in nursing research. The style and content of this chapter are designed to make subsequent chapters more user-friendly. Chapter 2, "Theoretical Framework," focuses specifically on how theoretical frameworks guide and inform knowledge generation through the research process. Chapter 3, "Critical Appraisal Strategies: Reading Research," addresses students directly and highlights critical thinking and critical reading concepts and strategies, thereby facilitating students' understanding of the research process and its relationship to the critical appraisal process. This chapter introduces a model evidence hierarchy that is used throughout the text.

The next two chapters address foundational components of the research process. Chapter 4, "Developing Research Questions, Hypotheses, and Clinical Questions," focuses on how research questions, hypotheses, and evidence-informed practice questions are derived, operationalized, and critically appraised. Numerous clinical examples illustrating different types of research questions and hypotheses maximize student understanding. Students are also taught how to develop clinical questions that are used to guide evidence-informed inquiry. Chapter 5, "Finding and Appraising the Literature," showcases cutting-edge information literacy content, providing students and nurses with the tools necessary to effectively search, retrieve, manage, and evaluate research studies and their findings. This chapter also develops research consumer competencies that prepare students and nurses to critically read, understand, and appraise a study's literature review and framework. The final chapter in this section, Chapter 6, "Legal and Ethical Issues," provides an overview of the increased emphasis on the legal and ethical issues facing researchers in Canada.

Part Two, Indigenous Peoples: Research, Knowledges, and Ways of Knowing, is an exciting, brand-new section in this fifth edition. We were honoured to be able to include an interview with Elder Evelyn Voyageur, and to work with Caroline Foster-Boucher, to develop Chapter 7, "Indigenous Peoples: Research, Knowledges, and Ways of Knowing." This section will introduce students to the history and significance of colonization as it relates to how Indigenous peoples have been affected by, and are affecting, nursing and health research. The chapter highlights the relevant recommendations from the Truth and Reconciliation Commission and offers a beginning exploration of Indigenous Methodologies.

Part Three, Qualitative Research, contains two interrelated qualitative research chapters. Chapter 8, "Introduction to Qualitative Research," provides a framework for understanding qualitative research designs and literature, as well as the significant contribution of qualitative research to evidence-informed practice. Chapter 9, "Qualitative Approaches to Research," presents, illustrates, and, in examples from the literature, showcases major qualitative methods. This chapter highlights the questions most appropriately answered using qualitative methods.

Part Four, Quantitative Research, contains Chapter 10 ("Introduction to Quantitative Research"), 11 ("Experimental and Quasiexperimental Designs"), and 12 ("Nonexperimental Designs"). These chapters delineate the essential steps of the quantitative research process, with

published, current clinical research studies used to illustrate each step. Links between the steps and their relationship to the total research process are examined.

Part Five, Processes Related to Research, describes the specific steps of the research process for qualitative and quantitative studies. The chapters make the case for linking an evidence-informed approach with essential steps of the research process by teaching students how to critically appraise the strengths and weaknesses of each step of the research process. Students learn how to select participants (Chapter 13, "Sampling"), gather data (Chapter 14, "Data-Collection Methods"), analyze the results (Chapter 16, "Qualitative Data Analysis," and Chapter 17, "Quantitative Data Analysis"), and present their results (Chapter 18, "Presenting the Findings"). Chapter 15, "Rigour in Research," gives students the tools for assessing the quality and trustworthiness of a study.

Part Six, Critiquing Research, makes the case for linking an evidence-informed approach with essential steps of the research process by teaching students how to critically appraise the strengths and weaknesses of each step of the research process. Each chapter critiques two examples of actual published research. Chapter 19, "Critiquing Qualitative Research Articles," focuses on qualitative research, whereas Chapter 20, "Critiquing Quantitative Research," is based on the quantitative research process.

Part Seven, Application of Research: Evidence-Informed Practice, contains the final chapter in the book. Chapter 21, "Developing an Evidence-Informed Practice," provides a dynamic review of evidence-informed models. These models can be applied—step by step, at the organizational or individual patient level—as frameworks for implementing and evaluating the outcomes of evidence-informed health care.

The Evolve website that accompanies the fifth Canadian edition provides interactive learning activities that promote the development of critical thinking, critical reading, and information literacy skills designed to develop the competencies necessary to produce informed consumers of nursing research. Instructor resources are available at a passcode-protected website that gives faculty access to all instructor materials online, including the TEACH for Nurses Lesson Plans, Image Collection, PowerPoint Slides, a Test Bank that allows faculty to create examinations through the use of the ExamView test generator program, and more.

The development and refinement of an evidence-informed foundation for nursing practice is an essential priority for the future of professional nursing practice. The fifth Canadian edition of *LoBiondo-Wood and Haber's Nursing Research in Canada: Methods, Critical Appraisal, and Utilization* will help students develop a basic level of competence in understanding the steps of the research process that will enable them to critically analyze research studies, evaluate their merit, and judiciously apply evidence in practice. To the extent that this goal is accomplished, the next generation of nursing professionals will include a cadre of clinicians who inform their practice by using theory and research evidence, combined with their clinical judgement, and specific to the health care needs of patients and their families in health and illness.

Mina D. Singh
minsingh@yorku.ca

Lorraine M. Thirsk
lthirsk@athabascau.ca

To the Student

EVIDENCE-INFORMED PRACTICE is integral to meeting the challenge of providing quality health care in partnership with patients and their families and significant others, as well as with the communities in which they live. As you read this fifth Canadian edition of *LoBiondo-Wood and Haber's Nursing Research in Canada: Methods, Critical Appraisal, and Utilization*, we hope you develop an appreciation of the importance of nursing research to practice. Whether you are a student or a practising nurse whose goal is to use research evidence as the foundation of your practice, you will discover that nursing research and a commitment to evidence-informed practice positions our profession at the forefront of change. As you learn about nursing research and evidence-informed practice, you will develop a foundation of knowledge and skills that will equip you for clinical practice today and into the future.

The fifth Canadian edition of *LoBiondo-Wood and Haber's Nursing Research in Canada: Methods, Critical Appraisal, and Utilization* reflects cutting-edge trends for developing evidence-informed nursing practice. The seven-part organization and special features in this text are designed to help you develop your critical thinking, critical reading, information literacy, and evidence-informed clinical decision-making while providing a user-friendly approach to learning that expands your competence to deal with these new and challenging experiences. The companion *Study Guide,* with its chapter-by-chapter activities, will serve as a self-paced learning tool to reinforce the content of the text. The accompanying Evolve website offers "summative" review material to help you reinforce the concepts discussed throughout the book.

Remember that evidence-informed practice skills are used in every clinical setting and can be applied to every patient population or nursing practice issue. Whether your practice involves primary care or specialty care and provides inpatient or outpatient treatment in a hospital, clinic, or home, you will be challenged to use research as the foundation for your evidence-informed practice. The fifth Canadian edition of *LoBiondo-Wood and Haber's Nursing Research in Canada: Methods Critical Appraisal, and Utilization* will guide you as you discover your ability to play a vital role in contributing to the building of an evidence-informed professional nursing practice.

Mina D. Singh
minsingh@yorku.ca

Lorraine M. Thirsk
lthirsk@athabascau.ca

A NOTE ABOUT THE NEXT-GENERATION NCLEX

The National Council for the State Boards of Nursing (NCSBN) is a not-for-profit organization whose members include nursing regulatory bodies. In empowering and supporting nursing regulators in their mandate to protect the public, the NCSBN is involved in the development of nursing licensure examinations, such as the NCLEX-RN®. In Canada, the NCLEX-RN® was introduced in 2015 and is, as of the writing of this text, the recognized licensure exam required for practising RNs in Canada.

The NCLEX-RN® will, as of 2023, be changing in order to ensure that its item types adequately measure clinical judgement, critical thinking, and problem-solving skills on a consistent basis. The

NCSBN will also be incorporating into the examination what they call the Clinical Judgement Measurement Model (CJMM), which is a framework that the NCSBN has created to measure a novice nurse's ability to apply clinical judgement in practice.

These changes to the examination come as a result of research findings that indicated that novice nurses have a much higher than desirable error rate with patients (i.e., errors that cause patient harm) and, upon NCSBN's investigation, the discovery that the overwhelming majority of these errors were caused by failures of clinical judgement.

Clinical judgement has been a foundation underlying nursing education for decades, based on the work of a number of nursing theorists. The theory of clinical judgement that most closely aligns to what NCSBN is basing their CJMM is the work by Christine A. Tanner.

The new version of the NCLEX-RN® is loosely being identified as the "Next-Generation NCLEX" or "NGN" and will feature the following:

- Six key skills in the CJMM: recognizing cues, analyzing cues, prioritizing hypotheses, generating solutions, taking actions, and evaluating outcomes.
- Approved item types as of June 2020: multiple response, extended drag and drop, cloze (drop-down), enhanced hotspot (highlighting), and matrix/grid. More question types may be added.
- All new item types are accompanied by mini case studies with comprehensive patient information—some of it relevant to the question, and some of it not.
- Case information may present a single, unchanging moment in time (a "single-episode" case study) or multiple moments in time as a patient's condition changes (an "unfolding" case study).
- Single-episode case studies may be accompanied by one to six questions; unfolding case studies are accompanied by six questions.

For more information (and detail) regarding the NCLEX-RN® and changes coming to the exam, visit the NCSBNs website: https://www.ncsbn.org/11447.htm and https://ncsbn.org/Building_a_Method_for_Writing_Clinical_Judgment_It.pdf.

For further NCLEX-RN® examination preparation resources, see *Silvestri's Canadian Comprehensive Review for the NCLEX-RN® Examination*, Second Edition, ISBN 9780323709385.

Prior to preparing for any nursing licensure examination, please refer to your provincial or territorial nursing regulatory body to determine which licensure examination is required in order for you to practise in your chosen jurisdiction.

Next-Generation NCLEX™ (NGN)-Style Case Studies can be found on this text's accompanying Evolve site.

PART ONE

Research Overview

RESEARCH **VIGNETTE**

Nursing Research to Improve Immunization in Canada

Shannon MacDonald, PhD, RN
Assistant Professor
Faculty of Nursing
University of Alberta
Edmonton, Alberta

We have all seen social media posts that try to scare people about vaccines. There are claims that vaccines are unnecessary, ineffective, or unsafe. Unfortunately, the average Canadian doesn't see the other side of the story, namely the people who have come to harm because they weren't protected from vaccine-preventable diseases. As nurses, we see both sides because we are the ones caring for people infected with these diseases. We also have the scientific literacy to critically read and evaluate the evidence on this topic. Thus, nurses are in the ideal position to promote vaccine uptake; to advocate for equitable vaccine access; and to ask important research questions that ensure that safe and effective vaccine programs continue to be provided to Canadians.

My own interest in the immunization field started when I worked as a nurse in the Pediatric Intensive Care Unit. I cared for children who were on ventilators as a result of a whooping cough infection contracted because they were too young to be vaccinated; teenagers who lost hands or legs to meningococcemia because the vaccine wasn't yet available to them; and children who had measles encephalopathy because their parents had decided not to give them the vaccine. Later in my nursing career, I travelled to less developed regions of the world, where it wasn't uncommon for mothers to report losing multiple children to vaccine-preventable diseases. In those cases, it wasn't due to a choice not to vaccinate, but instead due to limited access to health services. The cumulative effect of those clinical experiences led to my lifelong passion and program of research focused on improving immunization practices, policies, and parent decision-making.

The World Health Organization (WHO, 2014) has a conceptual framework to explain the influences on vaccine hesitancy and uptake. They identify 'the 3 Cs': **Complacency** (vaccines are not perceived as necessary), **Confidence** (lack of trust in the vaccines/providers/policy-makers), and **Convenience** (physical availability, affordability, geographical accessibility, ability to understand, and appeal/ease of immunization services). Any one, or a combination, of these factors can act as an impediment to someone accepting vaccines for themselves or their children.

When considering the WHO '3 Cs' framework, I have always found it striking that 'convenience' can prevent someone from vaccinating, even if complacency or confidence are not a concern. Thus, I have chosen to focus my research on identifying and improving system-level factors that influence convenience, such as immunization access, accurate immunization records, and reminder systems. Through our research, my team has found that these system-level barriers disproportionately impact vulnerable populations, including children receiving child intervention services (Hermann et al., 2019), First Nations children (MacDonald et al., 2016), and children in single-parent, low-income families that have unstable housing (Bell et al., 2015). Identifying these at-risk populations allows us as nurses to develop strategies to reach them, and to advocate for public health policies to provide system-level supports for these populations. I will share with you two examples of my research that seeks to achieve these goals.

Under-immunization of children receiving child intervention services: Children who receive intervention from the child welfare system (sometimes referred to as "children in care") typically do so to ensure their safety and security (Canadian Child Welfare Research Portal, 2020). However, the circumstances in their home, the processes of removal from their home, and their subsequent placement in one or more other care settings (e.g., foster care homes) may result in interruptions in provision of their immunizations (Hermann et al., 2019). In our study (Hermann et al., 2019), we wanted to assess vaccine coverage (i.e., the proportion of the eligible population who have received a vaccine) of children who had spent time in care of

the child welfare system. What we discovered was that children in care have consistently lower vaccine coverage than children who had never been in care. For instance, at age 2, vaccination coverage for children in care ranged from 54.3% to 81.4%, depending on vaccine, while coverage for those not in care ranged from 74.2% to 87.4%. So, our study revealed that vaccine coverage for this already vulnerable population was significantly below the rest of the population and vastly below the target level for vaccine coverage (target levels for most vaccines are >95%). As a result of our study, we have advocated for improvements in tracking immunizations for children in care, as well as increased collaboration between health and children's services ministries to ensure these vulnerable children are better served. We also identified the need to conduct more research on this topic, specifically, qualitative studies to identify the specific barriers that are preventing equitable access to immunizations for this population.

Rotavirus immunization for pre-term infants: Rotavirus vaccine is provided to infants, starting when they are 2 months old, with additional booster doses provided before the child turns 8 months old. Unlike some other vaccines, there is an upper age limit to when children can receive the vaccine, due to safety concerns for older children (Parashar et al., 2018). Our team conducted a study (Rafferty et al., 2019) to determine whether there were some sub-populations of children in Alberta that had lower vaccine coverage for rotavirus vaccine, in order to identify areas for improved service delivery. What we found was that pre-term infants were less likely to be vaccinated, despite being at higher risk of becoming seriously ill from a rotavirus infection. Other studies from the USA (Fathima et al., 2019; Dahl et al., 2018), found similarly that infants that were pre-term and/or low birth weight had low rotavirus vaccine coverage, and that these infants were more likely to have spent their first days/weeks/months in a Neonatal Intensive Care Unit (NICU). Our study raised a red flag in Alberta regarding whether infants in the NICU were receiving their recommended vaccines while hospitalized. It led us to collaborate with public health and hospital-based clinicians to assess and remedy the situation.

These are just two examples of the type of research that I and my research team have engaged in to improve immunization service delivery, and to improve the convenience for parents seeking immunizations for their children. The fact that I am a nurse leads me to approach this research through a nursing lens. I also approach research from a very applied, rather than theoretical, perspective. I am always eager to move beyond acquisition of knowledge, to identify ways that this new knowledge can improve clinical care and health policies in the real world.

Nurses, and nurse researchers, have the potential to play important roles in supporting immunization best practices, policies, and parent decision-making. As one of the largest groups of health care providers, and the most trusted by the public (Milton, 2018), we are in a prime position to provide clear, evidence-informed guidance to families who are struggling with the immunization decision. In many provinces and territories of Canada, nurses are also the main provider of immunizations. Thus, the role we can play is significant. Nurses are also ideally positioned to lead research in the field of immunization. We have the clinical perspective, the experience of interacting with families, and the education to engage in critical inquiry. Currently, there are only a handful of nurses in Canada who are working in this field. My dream is for nurses and emerging nurse researchers to accept the challenge to improve the care provided to families in Canada and beyond through improved evidence-informed immunization services. ■

REFERENCES

Bell, C. A., Simmonds, K. A., & MacDonald, S. E. (2015). Exploring the heterogeneity among partially vaccinated children in a population-based cohort. *Vaccine, 33*(36), 4572–4578.

Canadian Child Welfare Research Portal (2020). *Frequently Asked Questions* (FAQs). http://cwrp.ca/faqs.

Dahl, R. M., Curns, A. T., Tate, J. E., & Parashar, U. D. (2018). Effect of rotavirus vaccination on acute diarrheal hospitalizations among low and very low birth weight us infants, 2001-2015. *The Pediatric Infectious Disease Journal, 37*(8), 817–822. https://doi.org/10.1097/INF.0000000000001930.

Fathima, P., Gidding, H. F., Snelling, T. L., McIntyre, P. B., Blyth, C. C., Sheridan, S., & Moore, H. C. (2019). Timeliness and factors associated with rotavirus vaccine uptake among Australian Aboriginal and non-Aboriginal children: A record linkage cohort study.

Vaccine, 37(39), 5835–5843. https://doi.org/10.1016/j.vaccine.2019.08.013.

Hermann, J. S., Featherstone, R. M., Russell, M. L., & MacDonald, S. E. (2019). Immunization coverage of children in care of the child welfare system in high-income countries: A systematic review. *American Journal of Preventive Medicine, 56*(2), e55–e63. https://doi.org/10.1016/j.amepre.2018.07.026.

Hermann, J. S., Simmonds, K. A., Bell, C. A., Rafferty, E., & MacDonald, S. E. (2019). Vaccine coverage of children in care of the child welfare system. *Canadian Journal of Public Health, 110*(1), 44–51. https://doi.org/10.17269/s41997-018-0135-5.

MacDonald, S. E., Bell, C. A., & Simmonds, K. A. (2016). Coverage and determinants of uptake for privately funded rotavirus vaccine in a Canadian birth cohort, 2008–2013. *The Pediatric Infectious Disease Journal, 35*(6), e177–e179. https://doi.org/10.1097/INF.0000000000001125.

Milton, C. L. (2018). Will nursing continue as the most trusted profession? An ethical overview. *Nursing Science Quarterly, 31*(1), 15–16. https://doi.org/10.1177/2F0894318417741099.

Parashar, U. D., Cortese, M. M., & Offit, P. A. (2018). *Rotavirus vaccines* (7th ed.). Philadelphia: Elsevier.

Rafferty, E., Guo, X., McDonald, B., Svenson, L. W., & MacDonald, S. E. (2019). Measurement of coverage, compliance and determinants of uptake in a publicly funded rotavirus vaccination programme: A retrospective cohort study. *BMJ Open*, *9*(11). http://dx.doi.org/10.1136/bmjopen-2019-031718.

World Health organization (WHO) (2014). *Report of the SAGE working group on vaccine hesitancy*. http://www.who.int/immunization/sage/meetings/2014/october/SAGE_working_group_revised_report_vaccine_hesitancy.pdf?ua=1. Accessed May 14, 2020.

The Role of Research in Nursing

Lorraine Thirsk

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- State the significance of research to the practice of nursing.
- Recognize that theory, research, and practice are related.
- Describe the history of nursing research.
- Identify the roles of the research user and producer.
- Identify trends and priorities in health care research.

KEY TERMS

critical appraisal
data
dissemination
evidence-based practice
evidence-informed practice
phenomena
quality improvement
research

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

HOW DO YOU KNOW WHAT YOU know? Is it less painful for a patient if you remove a burn dressing slowly or quickly? What are the challenges that pregnant women face in obtaining prenatal care, and how is this complicated if they lack transportation or housing? How do you assess pain in someone with cognitive impairment like dementia? What is the best way to talk with a family who is hesitant to vaccinate their child? Does cannabis have a therapeutic effect for the treatment of anxiety? These are all questions that can be answered with research and are important to nurses.

Research is a systematic way to acquire knowledge. We also acquire knowledge through our own experiences, from our family, culture, traditions, and authorities. Historically, many nursing practices were based on received wisdom or tradition. Some of the first documentation discussing the importance of research evidence in nursing practice is from Florence Nightingale, who, in the 1850s, noted there was a connection between poor sanitary conditions and death rates among wounded soldiers (Nightingale, 1863).

In Europe, in the early 20th century, a movement began to define criteria that would separate science from non-sciences such as pseudoscience, metaphysics, ideology, and religion (Hansson, 2017). The scientific method was developed as a

systematic way to determine knowledge and understand the world. While human beings have an incredible capacity for thinking and understanding the world around them, the mind does sometimes make mistakes and is influenced by beliefs and ideology that do not always reflect reality. Over the last century, psychological studies have revealed that people have biases in their thought processes that can be problematic and give us an incorrect or a skewed view of the world around us (Buetow, 2019; Kahneman, 2011; Paley et al., 2007). Our hunches about patterns and probabilities are insufficient to ensure good decision-making. Health care resources are scarce, and thus it is prudent for nurses and other health care professionals to ensure that scarce resources are used wisely. In other words, we need to know that our decisions are based on the best evidence available (Melnyk & Fineout-Overholt, 2011).

Research is integral to achieving the goal of providing quality outcomes in partnership with patients, their families and significant others, and the communities in which they live. As you progress through your educational program you will be taught how to ensure quality and safety in practice by acquiring knowledge of various sciences and health care principles. Research is a critical foundation of an **evidence-informed** approach to nursing practice, positioning nurses at the cutting edge of change and improvement in patients' outcomes.

The aim of this book is to prepare you to critically appraise research and incorporate research into your practice. Throughout this text you will find special features that will help refine and develop your competence in using research. Each chapter contains a *Critical Thinking Decision Path* related to each step of the research process; these will sharpen your decision-making skills as you critique research articles. *Internet* resources in the chapters will also enhance your research user skills. *Critical Thinking Challenges*, which appear at the end of each chapter, are designed to reinforce your critical thinking and judgment skills in relation to the steps of the research process. *Research Hints*, designed to reinforce your understanding and critical thinking, appear at various points throughout the chapters. *Evidence-Informed Practice Tips* will help you apply evidence-informed practice strategies in your clinical practice. Finally, *Practical Application* boxes offer examples of translating principles and methods of nursing research into real-life nursing situations and interventions.

Your critical thinking, critical reading, and clinical decision-making skills will expand as you develop clinical questions, search the research literature, evaluate the research evidence found in the literature, and make clinical decisions about applying the best available evidence. In this book you will discover the "who, what, where, when, why, and how" of research and develop a foundation of knowledge, evidence-informed practice, and competencies that will equip you for 21st-century nursing practice. To begin, this chapter provides an overview of the importance of research to evidence-informed practice, the role that research plays in practice, the roles of nurses in research activities, and future directions of health care research.

SIGNIFICANCE OF RESEARCH AND EVIDENCE-INFORMED PRACTICE

The health care environment is changing at an increasingly rapid pace. The challenges associated with these changes and with nursing's rapid pace of growth can best be met by integrating evidence-informed knowledge into nursing practice. Nursing research provides scientific knowledge that enables nurses to keep up with these changes.

In learning about research, it is important to differentiate between the terms *research, evidence-based practice,* and *evidence-informed practice*. **Research** is systematic, rigorous, logical investigation with the aim of answering questions about nursing phenomena. **Phenomena** can be defined as occurrences, situations, or facts that

are perceptible by the senses. Although the origin of the term *phenomena* refers to events that are observable and/or measurable, nurses are also interested in experiences that are not easily observed, such as the experiences of pain, loss, or anxiety.

In the past 25 years, many health care disciplines have adopted the tenets of evidence-informed practice to provide better health care for their patients. The roots of modern evidence-informed practice stem from Dr. Archie Cochrane's investigation of the efficacy of health care interventions, particularly in medicine. His work resulted in the establishment of the Cochrane Collaboration, which provides systematic reviews of health care interventions.

In 1996, Sackett and colleagues defined "evidence-based medicine" as the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (p. 312). This was considered the beginning of **evidence-based** practice, which most health professions have now adopted. Research is completed, published in academic journals, and then assessed to determine application to clinical practice—this results in practice that is evidence-based. The evidence-based practice movement has not been without challenges and problems. The strict application of algorithms and guidelines, along with selective trials that overlook multiple morbidities, means that individual patients are not always receiving the most appropriate care (Greenhalgh et al., 2014). Evidence-based practice can become rigid and not consider patient preferences, individuality, and contexts.

Evidence-informed practice extends beyond the early definitions of evidence-based practice. With evidence-informed practice, the methods for gathering evidence are the same as the processes used for evidence-based practice; however, the evidence also incorporates expert opinion, clinical expertise, patient preference, and other resources (CNA, 2018). It is important to remember that evidence-informed practice focuses on a more inclusive and interactive process of decision-making that pertains to all nurses—whether they are clinicians, educators, researchers, administrators, or policy-makers (CNA, 2018). Building on the foundation of evidence-based practice, evidence-informed practice also involves acknowledging and considering the myriad factors that constitute decision-making, taking into account patient preference, culture, history, and local context.

When you first read about the research and the evidence-informed practice process, you will notice that both processes may seem similar. Each begins with a question. The difference is that in a research study, the question is tested with a design appropriate for the question and with specific methods (sample, instruments, procedures, and **data** analysis). In the evidence-informed practice process, a question is used to search the literature for studies already completed and then you critically appraise this literature in order to answer your clinical question.

Broadly, there are two types of research: quantitative and qualitative. Increasingly, many researchers use mixed methods—in other words, they utilize both types of research in one project, or in examination of one phenomenon in a program of research. You will be introduced to these types of research in more depth in Chapter 2. In addition, the *Research Vignettes* included throughout the text will introduce you to nurses who use a variety of research methods to study phenomenon important to health care and nursing practice. The methods used by nurse researchers are the same methods used in other disciplines; the difference is that nurses study questions relevant to nursing practice. Nurse researchers also conduct research collaboratively with researchers from other disciplines. Through the conducting of research, they produce knowledge that is reliable and useful for nursing practice. The methods and findings of studies provide evidence that is evaluated, and their applicability to practice is used to inform decisions.

Throughout this text, the steps of the research and evidence-informed practice processes are

described. Understanding the step-by-step process that researchers use will help you develop the assessment skills necessary to judge the soundness of research studies and participate in or lead research projects someday. Chapter 21 will further describe how you can implement evidence into practice to improve patient outcomes.

RESEARCH: THE ELEMENT THAT LINKS THEORY, EDUCATION, AND PRACTICE

Research links theory, education, and practice. Theoretical formulations supported by, or developed from, research findings may become the foundations of theory-informed practice in nursing and inform further research studies. Your educational setting, whether a nursing program or the health care organization where you are employed, provides an environment in which you, as a student or an employee, can learn about the research process. In the setting of a nursing program or a health care organization, you can also explore different theories and practices and begin to evaluate them based on research findings. The knowledge you gain through your educational program, whether theory or research based, will inform the decisions that you make in your practice. See the Practical Application box for an example of how theory and research influence health care practices.

The example in the Practical Application box is an attempt to answer a question that you may have asked before taking this course: "How will the theory and research content of this course relate to my nursing practice?" This example demonstrates how theory informs practice, how knowledge based only on experience can be biased and limited, and how approaching clinical problems with systematic, scientific research methods can improve patient outcomes. Dan Ariely was not a nurse, but this anecdote demonstrates how nursing practice could be drastically changed by research. In this example you can see how theory, research, and practice are connected. Theory is used to explain causal relationships (e.g., if I remove the dressing quickly, it will be less painful). Theory needs to be tested through research, and then this new evidence needs to be incorporated into practice. Often in the absence of evidence, theoretical knowledge will guide practice. Research can also be used to generate new theory. The relationships between theory, practice, and research will be further explored in Chapter 2.

Practical Application

Dan Ariely was badly burned when he was 18 years old (Ariely, 2009)—70% of his body experienced 3rd degree burns. During daily dressing changes, he noticed that most of the nurses would grab the bandages and rip them off as quickly as possible. He recalls thinking the nurses had ***theorized*** that quick, sharp bursts of pain were better for the patient than slowly pulling off the bandages. In addition, he noticed that there was no rationale as to whether the dressing changes were started at the most painful part of his body or the least painful part. As a patient, he had opinions about which methods were better, but there did not seem to be any ***evidence*** to help guide the nurses on the best methods. When he later attended university, Ariely began working as a research assistant and eventually started to test some of his theories about pain and the removal of burn dressings. The ***research*** he conducted showed that slowly removing burn dressings would result in the least amount of pain for the patient. He wondered how these kind and experienced nurses could be so wrong. "I knew that their behaviour was not due to maliciousness, stupidity, or neglect. Rather, they were most likely the victims of inherent biases in their perceptions of their patients' pain—biases that apparently were not altered even by their vast experience" (Ariely, 2009, p. xvi). Interestingly, when he reported his results back to the nurses at the burn unit, one nurse explained that perhaps removing the dressings quickly lessened the nurse's *psychological pain*, which they experienced when they inflicted pain on patients.

Source: Based on Ariely, D. (2009). *Predictably Irrational: The Hidden Forces That Shape Our Decisions.* New York:. Harper Perennial.

Evidence-Informed Practice Tip

What is the current evidence on how to remove burn dressings? Given that the example provided was over a decade old, has this research been incorporated into practice? Has more recent research been done? Health care leaders have an important role to play in implementing research in practice areas. Implementing research in practice requires support from leaders who are champions of evidence (see Chapter 21 for more about implementing evidence-informed practice).

Learning about research will provide you with an appreciation and understanding of the research process so that you can more easily become a participant in research activities and an intelligent consumer of research. A research user actively uses and applies research. To be a knowledgeable research user, you must have knowledge about the relevant subject matter, the ability to discriminate and to evaluate information logically, and the ability to apply the knowledge gained. You need not actually conduct research to be able to appreciate and use research findings in practice. Rather, you must understand the research process and develop the critical evaluation skills needed to judge the merit and relevance of evidence before applying it to practice. The success of evidence-informed practice depends on your ability to understand the research process and to evaluate the evidence. Nurses in practice, who understand research and its contribution to knowledge, are ideally suited to identify phenomena and issues to be studied by asking relevant research questions.

ROLES OF THE NURSE IN THE RESEARCH PROCESS

Every nurse practising in the 21 century has a role to play in the research process.

The Canadian Nurses Association (2017) declares that "nurses support, use and engage in research and other activities that promote safe, competent, compassionate and ethical care, and they use guidelines for ethical research that are in keeping with nursing values" (p. 9). What does this mean for you? There is a consensus that effective use of research calls for the skills of **critical appraisal**; that is, you can appraise research evidence and use existing standards to determine the merit and readiness of research for use in clinical practice:

> Sources of evidence need to be critically appraised before their findings are incorporated into decision-making and practice. Sources that meet this standard include systematic reviews, research studies and peer-reviewed journals that summarize valid and clinically useful published studies. (CNA, 2018, p. 1)

Therefore, to use research for evidence-informed practice, you may not necessarily be conducting research, but you can understand and appraise the steps of the research process in order to read the research literature critically and use it to inform your clinical decisions. Even as students you can participate by completing surveys, attending research conferences, and asking questions.

At a provincial level, each province in Canada has its own standards for entry into nursing practice, and many of these standards have specific related research competencies. For example, the College and Association of Registered Nurses of Alberta (2019) outlined the following competencies for nurses in their role as scholars:

> Registered nurses are scholars who demonstrate a lifelong commitment to excellence in practice through critical inquiry, continuous learning, application of evidence to practice, and support of research activities. (p. 15)

Nurses must be intelligent users of research; that is, they must understand all steps of the research process and their interrelationships. Frisch et al. (2013) have developed a useful description of a *Health Services Researchers Pathway* that explains the five levels of nurses' roles in research (Table 1.1). The nurse interprets, evaluates, and determines the credibility of research findings. The nurse discriminates between interesting findings for which further investigation is required and those that are sufficiently supported by evidence before applying findings to practice. The nurse should then use these competencies to advance nursing or interdisciplinary evidence-informed practice projects (e.g., developing clinical standards, tracking **quality improvement** data, or coordinating implementation of a pilot project to test the efficacy of a new wound care protocol) of the workplace committees to which he or she belongs. Nurses are also responsible for generating clinical questions to identify nursing issues that necessitate

TABLE 1.1

HEALTH SERVICES RESEARCHER PATHWAY

	RESEARCH PROCESS	DATA ANALYSIS LITERACY	KNOWLEDGE TRANSLATION
Level 1 Research User: Learning about research use in care delivery settings	• Defines and distinguishes between research and quality improvement • Follows agency policy and clinical practice guidelines; collaborates on QI activities • Curious and willing to learn about research	• Understands and values statistics and quantitative and qualitative research methods • Reads research reports	• Identifies credible and reliable resources • Performs literature searches, integrating evidence into EIP • Interested in and advocates for practice improvement
Level 2 Research User: Using research in care delivery settings	• Describes research and QI processes; explains QI processes and models • Interprets protocols for relevancy, conducts literature reviews, participates in policy development and QI • Appreciates relationship between research and practice, values active engagement of front-line staff in QI and research	• Understands application of statistics and steps of research process • Collects and uses accurate data, uses basic statistics and qualitative methods • Appreciates the process of conducting research	• Identifies opportunities for knowledge sharing, understands concept of strength of evidence, distinguishing between single studies and systematic reviews • Collaborates with team to change practice and support KT • Aware of and appreciates research activities in the workplace and willing to lead KT activities
Level 3 Research User: Facilitating and leading research use in care delivery settings	• Describes emerging knowledge, best practices and priorities; facilitates research	• Interprets qualitative and quantitative data and can conduct simple analysis	• Describes KT practices and facilitates KT projects, translates projects
Level 4 Research Producer: Beginning researcher	• Understands research designs and theoretical frameworks, manages research projects, contributes to research teams	• Understands advanced analysis techniques, critically and accurately analyzes research data	• Uses research findings to support policy and practice, carries out KT plans
Level 5 Research Producer: Research scientist leading a program of research	• Expertise in at least one method, understands various research approaches • Leads a program of research	• Expert in analysis methods in own research program, manages and supervises use of data, values rigorous analysis	• Builds and implements KT as part of own research program

Source: Adapted from Frisch & Hamilton (2013). *Health Services Researcher Pathway*. Michael Smith Foundation for Health Research and the BC Nursing Research Initiative.

investigation and for participating in the implementation of scientific studies. Nurses often generate research ideas or questions from hunches, gut-level feelings, intuition, or observations of patients or nursing care. These ideas often become the seeds of research investigations.

Nurses may participate in research projects as members of research teams in one or more phases of a project. For example, a staff nurse may work on a clinical research unit in which a research project is underway to test a new type of nursing care (e.g., for pain management, prevention of falls, or treatment of urinary incontinence). In situations such as these, the nurse administers care according to the format described in the research protocol. The nurse may also be involved in collecting and recording data relevant to the administration of, and the patient's response to, nursing care.

After new knowledge is generated, it is important to share findings widely. This is called **dissemination**. Examples of dissemination include publishing an article or presentation at a conference. It may involve joining a health care agency's research committee or its quality assurance or quality improvement committee, in which research articles, integrative reviews of the literature, and clinical practice guidelines are evaluated for evidence-informed clinical decision-making.

Nurses who have graduate degrees are further prepared to conduct research as co-investigators or primary investigators. With a master's degree, nurses can focus on being more active members of research teams. Although master's degrees may focus on advanced clinical practice, advanced practice nurses are still champions for research. They can assume the role of clinical expert, collaborating with an experienced researcher in proposal development, data collection, data analysis, and interpretation. Nurses with master's degrees enhance the quality and relevance of nursing research by providing clinical expertise and evidence-informed knowledge about the way clinical services are delivered. Nurses with master's-level training also facilitate the investigation of clinical problems by enabling a climate that is open to nursing research and by engaging in evidence-informed practice projects. A clinical nurse specialist prepared with a master's or doctoral degree in nursing who has clinical expertise in a specific practice area can be the primary researcher or act as a collaborator to "ensure their practice applies evidence-based care most effectively while being a leader in every aspect of research" (CNA, 2020, Roles section).

To achieve the greatest expertise in appraising, designing, and conducting research, nurses must complete PhDs. Nurses with doctoral degrees develop theories for phenomena relevant to nursing, develop methods of scientific inquiry, and use a variety of methods to modify or extend existing knowledge so that it is relevant to nursing (or to other areas of health care). In addition to their role as researchers, nurses with doctoral-level training act as role models and mentors to guide, stimulate, and encourage other nurses who are developing their research skills. Nurses with doctoral degrees also collaborate and consult with social, educational, and health care institutions or governmental agencies in their respective research endeavours. These nurses then disseminate their research findings to the scientific community, clinicians, and—as appropriate—the general public through scientific journal articles and presentations at research conferences.

An essential responsibility of all nurses is to pay special regard to the ethical principles of research, especially the protection of human participants (see Chapter 6). For example, nurses caring for patients who are participating in research on antinausea chemotherapy must ensure that patients have signed the informed consent form and that all their questions are answered by the research team before they begin participation. Furthermore, if patients have an adverse reaction to the medication, nurses must not administer more doses until they have notified an appropriate member of the research team. Regardless of

their role, nurses need to view the research process as integral to the growing professionalism in nursing.

As a professional, you must take time to read research studies and evaluate them, using the current standards for scientific research. Also, you will need to use the critiquing process to identify the strengths and weaknesses of each study. Bearing in mind that each study has its limitations, you should consider whether sound and relevant evidence from one study can be used in other settings as well. Chapter 21 will expand on how to bring research into your nursing practice.

HISTORICAL PERSPECTIVE[1]

During the Crimean War, Florence Nightingale's detailed and systematic observation of nursing actions and outcomes resulted in major changes in nursing practice. Her work demonstrated the importance of systematic observational research to nursing practice.

In Canada, the establishment of university nursing courses starting in 1918, followed by master's degree programs in the 1950s and 1970s and by doctoral programs in the 1990s and 2000s, was crucial to the development of nursing research. Since the 1970s and 1980s, the two major factors in the development of nursing research have been the establishment of research training through doctoral programs and the establishment of funding to support nursing research. Throughout the 1970s and 1980s, university faculties and schools of nursing built their research resources so that they could establish doctoral programs. The first provincially approved doctoral nursing program was established at the University of Alberta Faculty of Nursing in 1991. Another was established at the University of British Columbia School of Nursing later that year, and programs at McGill University and the University of Toronto followed in 1993. Now there are many doctoral nursing programs in universities across Canada. In addition, there has been growth of university-based Registered Psychiatric Nursing programs, with the first master's degree established at Brandon University in 2011.

Growing awareness of the importance of nursing research gradually led to the availability of research funds. The year 1964 marked the first time that a federal granting agency funded nursing research in Canada (Good, 1969). In 1999, the Canadian government established the Nursing Research Fund, budgeting $25 million for nursing research ($2.5 million over each of the following 10 years). The research areas targeted for support included nursing policies, management, human resources, and nursing care. Although this funding is no longer available, nurse researchers have been successful at obtaining Tri-Council funding nationally and international funding to support their programs of research. Tri-Council is a term referring to three federal research agencies: Canadian Institutes of Health Research, National Science and Engineering Research Council of Canada, and the Social Sciences and Humanities Research Council.

CURRENT STATE AND FUTURE DIRECTIONS

While the last 30 years has seen an increase in the number of nurse researchers in Canada, there are still challenges. A global shortage of nursing faculty, particularly of nurses who hold PhDs, impacts the education of the next generation of nurses as well as the capacity for nursing research (Vandyk et al., 2017). In Canada, there are numerous faculty vacancies and an insufficient number of PhD graduates every year to fill these spaces; recruitment and retention of faculty is a concern with an aging nursing faculty workforce (Canadian Association of

[1]This section (i.e. Historical Perspectives p. 12) is adapted with permission from Duggleby, W., & Astle, B.J. (2019). The development of nursing in Canada. In Potter, P., Duggleby, W., Stockert, P. Astle, B., Perry A., & Hall, A. (Eds.), *Canadian fundamental of nursing* (6th ed., pp. 75–80). Elsevier.

Schools of Nursing, 2016). Nursing faculty are needed to teach in undergraduate and graduate programs and mentor the new generation of nurse researchers. To further the body of nursing knowledge, nurses will need to develop programs of research, increase research on nursing interventions and outcomes, and be aware of national and international trends and issues in health and health care.

Developing Programs of Research

To build robust research knowledge, nurses need to be recruited early in their careers to pursue graduate education and develop programs of research. Developing a program of research can take years. Researchers need time in their careers to establish their expertise and develop the necessary collaborations and funding streams to support their investigations.

Research programs that include a series of studies in a similar area, each of which builds on a prior investigation, promote depth and credibility in nursing science. An example of a research program can be seen in the Practical Application box. To maximize use of resources and to prevent duplication, researchers must develop intradisciplinary, interdisciplinary, and international networks in similar areas of study. Researchers from a variety of health professions (e.g., medicine, nursing, and respiratory therapy) and other disciplines such as psychology, law, and business can come together to delineate common and unique aspects of patient care. Interdisciplinary health research may be "a team of researchers who come together to research an important and challenging health issue" (Hall et al., 2006, p. 764). Interdisciplinary research is increasingly becoming a mandate of research funders, as it is recognized that expertise is required from many disciplines to solve complex health and social problems (Clarke et al., 2012).

Dr. Stajduhar's work illustrates the value of building a program of research and highlights the importance of research teams. A large cadre of nurse researchers, who begin their research careers at a young age, is important for the development of research programs like Dr. Stajduhar's. The goal is to increase the longevity of research careers, enhance the discipline's scientific development, promote mentoring opportunities, prepare the next generation of researchers, and provide leadership in health care. The *Research Vignettes* included in this book have further examples of Canadian nurse researchers who have developed programs of research addressing a variety of current health care trends and issues.

Practical Application

Dr. Kelli Stajduhar is a professor at the Institute for Aging and Lifelong Health and the School of Nursing at the University of Victoria. With a research career spanning more than 20 years, Dr. Stajduhar leads a team of researchers studying palliative and end-of-life care issues. The work of this team of researchers spans from palliative care in vulnerable populations (Stajduhar et al., 2019), supporting family caregivers (Sutherland et al., 2016), and understanding family's experiences (Stajduhar et al., 2017). Dr. Stajduhar collaborates with researchers from around the world, as well as policy makers and health care providers. This work has influenced and will continue to shape a palliative approach to care in Canada.

Interventions and Outcomes

Globally, there is a need for more research on nursing interventions (Richards, Hanssen, & Borglin, 2018) and fundamental nursing care (Kitson et al., 2019). Quality research is still needed to address essential nursing care tasks including managing elimination, hygiene, nutrition, and mobility with patients (Richards, Hilli, et al., 2018).

Strategies that enhance nurses' focus on outcomes management through evidence-informed quality improvement activities and the use of research findings for effective clinical decision making also are being refined and identified as research priorities (see Chapter 21).

Evidence-informed practice guidelines, standards, protocols, decision tools, and critical pathways are becoming benchmarks for cost-effective, high-quality clinical practice. For example, the Registered Nurses' Association of Ontario (RNAO) (2016) has developed 50 best practice guidelines to support nurses in their efforts to provide the best possible patient care.

Evidence-Informed Practice Tip

The COVID-19 pandemic in 2020 created many challenges and opportunities for nursing research. While delaying the conduct of some research projects and funding decisions, the pandemic also resulted in a plethora of research on new topics. For example, research is examining the impact of personal protective equipment on nursing workflow in emergency departments (Government of Canada, 2020), standardized nursing care models for COVID-19 patients (Richards, 2020), and nursing leadership in acute and long term care settings during the pandemic (Baxter, 2020).

An International Perspective

The continuing development of a national and international research environment is essential to the nursing profession's mission to "improve the health and well-being of all world citizens" (National Institute of Nursing Research, 2015, n.p.). The CNA has been partnering with many international networks in more than 45 countries to strengthen the nursing profession's contribution to global health through study, research, and practice (CNA, 2012). Because of nursing's emphasis on the cultural aspects of care and the influence of such factors on practice, international research is likely to increase. Access to multiple populations as a function of globalization allows the testing of nursing science from various perspectives.

International research projects are often focused on comparative research in which a phenomenon is studied in more than one country. Ideally, relationships are formed with researchers from the international sites, resulting in collaborative research projects. Despite the financial and logistical limitations of this method, the number of international collaborative research projects has increased. Nurse researchers participating in collaborative international research projects are well positioned to play a large role in improving health care globally (CNA, 2012; Grady, 2015). An example of international collaboration can be seen in Chapter 19 (Harvey et al., 2019).

International organizations committed to the goal of health care for all help create natural research partnerships. For example, the World Health Organization (WHO, n.d.) has established a series of collaboration centres to advance health care for the global community. One such centre works toward maximizing the contribution of nursing and midwifery and provides relevant research and clinical training to nurses worldwide.

Research Priorities Reflecting Trends and Issues

Funding agencies often determine research priorities based on their needs and interests. These priorities are often reflective of trends and issues in health and health care. In 2018, the Canadian Association of Schools of Nursing identified seven priorities for nursing research:

- Indigenous and other vulnerable and/or equity seeking communities
- Chronic disease management and care delivery across space and time
- Home care and primary health care nursing
- Care of older adults across diverse care contexts
- Roles, scopes of practice, and value of RNS, and/or NPs to health care
- Nursing care, quality improvement, and patient safety
- Nursing education outcomes

The Canadian Foundation for Healthcare Improvement (CFHI, 2020a) supports spreading health care innovations throughout Canada by bringing together patients, families, health and social services providers, governments, and other organizations from across the country to solve persistent health care problems. Two priority

health challenges that have been identified are improving access to addiction and mental health services, and home and community care (CFHI, 2020b). These two areas were identified by the Government of Canada as shared health priorities between federal, provincial, and territorial governments (Health Canada, 2018).

The Canadian Institute for Health Research (CIHR) is one of the largest funders of health research, although the application process is highly competitive, with as few as 13% of applicants being successful in procuring funding (Semeniuk, 2016).

In 2016, CIHR developed a strategy for patient-oriented research to help engage patients in research as more than just participants and to promote research that addressed patients' concerns (CIHR, 2018). This resulted in research units being developed across the country to help researchers increase and improve patient-oriented research. It has impacted how researchers get funding for research and offers support to help facilitate research that is focused on patient outcomes.

Reducing health disparities in underserviced communities and vulnerable populations is another major topic that will shape the focus of future nursing- and interdisciplinary-related research agendas, particularly among Indigenous peoples. The CIHR (2020) has an Institute of Indigenous Peoples' Health and the health of First Nations, Inuit and Metis Peoples is a priority for the other institutes of CIHR that offer research funding. In 2019, CIHR created six research awards for Indigenous Research Chairs in Nursing. These awards, totalling close to $6 million, support Indigenous and non-indigenous nurses to conduct research focused on Indigenous health.

Health research will continue to occur across the lifespan. For example, the health concerns of mothers and infants will continue to spur research that deals effectively with the maternal–neonatal mortality rate. Individuals of all ages who have sustained life-threatening illnesses will live with the help of new life-sustaining technology that will in turn create new demands for self-care and family support. Cancer, heart disease, arthritis, asthma, chronic pulmonary disease, diabetes, and Alzheimer's disease, prevalent during middle age and later life, will be responsible for expenditures of large proportions of the available health care resources. The impacts of the COVID-19 pandemic on the health of individuals and communities will likely not be fully understood for years. HIV/AIDS, a chronic illness that affects men, women, and children, will continue to have a significant effect on health care delivery. Access to quality palliative care services and groundbreaking research on medical assistance in dying will be prevalent.

Another vulnerable population, persons with mental health illness and addictions, will be served by a better understanding of mental disorders, which will emerge because of advancements in psychobiological knowledge and research initiatives. Mental health illnesses will continue to be a major public health issue; "depression is a leading cause of disability and a major contributor to the overall global burden of disease" (WHO, 2020, Key Facts). Alcohol and drug abuse will continue to be responsible for significant individual suffering and health care expenses as well as significant social and economic losses (WHO, 2018).

Nurse researchers will have an increasingly strong voice in shaping public policy relating to health care. Disciplines such as nursing—because of its focus on treatment of chronic illness, health promotion, independence in health, and care of the acutely ill, all of which are heavily emphasized values for the future—will be central to the shaping of health care policy in the future. Research evidence that supports or refutes the merit of health care needs and programs focusing on these issues will be timely and relevant. Thus, nursing and its scientific base is well placed to shape

health policy decisions (Turale & Kunaviktikul, 2019).

Data analytics has incredible potential to improve health care. Nurse leaders will need competencies in analyzing, managing, and using data analysis tools. Using this unprecedented amount of information, nurses can improve patient care and mitigate risk by informing decisions about patient flow, interventions, workforce modelling, cost drivers, and workplace safety (Solman, 2017). Knowing how to ethically access and successfully analyze this data will be key if nurses—regardless of their role—are to make the best use out of this increase in technology and computing power.

Communication of nursing research has also become increasingly important. Research findings continue to be disseminated in professional arenas (e.g., international, national, regional, and local electronic and print publications and conferences) as well as in consultations and staff development programs implemented on site through webinars and websites. Dissemination of research findings in the public sector has also gained importance.

Increasingly, nurse researchers are being asked to testify at governmental hearings and to serve on commissions and task forces related to health care. Nurses are quoted in the media when health care topics are addressed, and their visibility has expanded significantly.

As opportunities are recognized and gaps in science are observed, nurses will conduct, critique, and use nursing research in ways that publicly demonstrate how nursing care makes a difference in patients' lives. Nurses have a research heritage to be proud of. They also have a challenging and exciting future ahead of them. Both researchers and users of research need to engage in a united effort to gather and assess research findings that make a difference in the care that is provided and in the lives that are touched by their commitment to evidence-informed nursing practice.

CRITICAL THINKING CHALLENGES

- What research roles are you interested in?
- What effects will evidence-informed patient outcome studies have on the practice of nursing?
- Have you had any experiences, like Dan Ariely's, that make you question nursing practice?
- Why is it important to have interdisciplinary and international research perspectives?
- What topics in nursing do you think require further research?

CRITICAL JUDGEMENT QUESTIONS

1. What is the most appropriate source of information for evidence-informed practice?
   - A. Charge nurse
   - B. Attending physician
   - C. Clinical practice guideline
   - D. Nightingale's notes on nursing
2. Why are interdisciplinary networks important in research?
   - A. Collaboration can help solve complex problems
   - B. Nurses do not do independent research
   - C. Research funding needs to be spread across disciplines
   - D. There is overlap in the scopes of practice
3. What drives the priorities for health care research?
   - A. Political agenda
   - B. Changes in values in society
   - C. Trends and issues in health care
   - D. The United Nations

KEY POINTS

- Nursing research expands the body of scientific knowledge that forms the foundation of evidence-informed nursing practice.
- Nurses gain research literacy through education and practical experience. As users of

research, nurses must have a basic understanding of the research process and must demonstrate critical appraisal skills to evaluate the strengths and weaknesses of research before applying the research to clinical practice.
- All nurses, whether they possess baccalaureate, master's, or doctoral degrees, have a responsibility to participate in the research process.
- Programs of research studies and replication of studies will become increasingly valuable.
- Research studies will emphasize clinical issues, problems, and outcomes. Priority will be given to research studies that focus on health promotion, care for the health needs of vulnerable groups, and the development of cost-effective health care systems.
- Both users of research and nurse researchers will engage in a collaborative effort to further the growth of nursing research and accomplish the profession's research objectives.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Ariely, D. (2009). *Predictably irrational*. New York: Harper.

Baxter, P. (2020). *Nursing leadership during the pandemic*. McMaster University School of Nursing. Retrieved from https://nursing.mcmaster.ca/news-events/news/news-item/2020/09/08/nursing-leadership-during-the-pandemic.

Buetow, S. (2019). Apophenia, unconscious bias and reflexivity in nursing qualitative research. *International Journal of Nursing Studies, 89*, 8–13. https://doi.org/10.1016/j.ijnurstu.2018.09.013.

Canadian Association of Schools of Nursing [CASN]. (2016). *Registered nurses education in Canada statistics*. Retrieved from http://www.casn.ca/wp-content/uploads/2016/11/2014-2015-SFS-FINAL-REPORT-suppressed-updated.pdf.

Canadian Association of Schools of Nursing [CASN]. (2018). *National research priorities for nursing*. Retrieved from https://www.casn.ca/wp-content/uploads/2018/12/Research-Priorities-2018-EN.pdf.

Canadian Foundation for Healthcare Improvement [CFHI]. (2020a). About us. Retrieved from https://www.cfhi-fcass.ca/AboutUs.aspx.

Canadian Foundation for Healthcare Improvement [CFHI]. (2020b). *Priority health innovation challenge.* Retrieved from https://www.cfhi-fcass.ca/WhatWeDo/challenges/priority-challenge.

Canadian Institutes of Health Research [CIHR]. (2018). *About SPOR*. Retrieved from https://cihr-irsc.gc.ca/e/51036.html.

Canadian Institutes of Health Research [CIHR]. (2020). *Indigenous health research at CIHR*. Retrieved from https://cihr-irsc.gc.ca/e/50339.html.

Canadian Nurses Association [CNA]. (2012). CNA global health partnerships retrospective. Retrieved from https://nurseone.ca/, http://media/cna/page-content/pdf-fr/global_health_partnership_program_2012_e.pdf?la=en.

Canadian Nurses Association [CNA]. (2017). Code of ethics for registered nurses. Retrieved from https://www.cna-aiic.ca/html/en/Code-of-Ethics-2017-Edition/files/assets/basic-html/page-1.html.

Canadian Nurses Association [CNA]. (2018). Evidence-informed decision-making and nursing practice. Retrieved from https://www.cna-aiic.ca/-/media/cna/page-content/pdf-en/evidence-informed-decision-making-and-nursing-practice-position-statement_dec-2018.pdf.

Canadian Nurses Association [CNA]. (2020). *Clinical nurse specialists,*.Retrieved from https://www.cna-aiic.ca/en/nursing-practice/the-practice-of-nursing/advanced-nursing-practice/clinical-nurse-specialists.

Clarke, D., Hawkins, R., Sadler, E., Harding, G., Forster, A., McKevitt, C., Godfrey, M., Monaghan, J., & Farrin, A. (2012). Interdisciplinary health research: perspectives from a process evaluation research team. *Quality in primary care, 20*(3), 179–189. https://pubmed.ncbi.nlm.nih.gov/22828672/.

College and Association of Registered Nurses of Alberta. (2019). *Entry-level competencies for the practice of registered nurses.* Retrieved from https://nurses.ab.ca/docs/default-source/document-library/standards/entry-to-practice-competencies-for-the-registered-nurses-profession.pdf?sfvrsn=15c1005a_16.

Frisch, N., Hamilton, S., Borycki, E., Lawrie, B., MacPhee, M. Mallidou, A., Mickelson, G., Redekopp, M., & Young, L. (2013). *Health services researcher pathway.* Retrieved from https://www.msfhr.org/health-services-researcher-pathway-0.

Good, S. R. (1969). *Submission to the study of support of research in universities for the Science Secretariat of the Privy Council.* Ottawa: Canadian Nurses Association and Canadian Nurses Foundation.

Government of Canada. (2020). *Government of Canada funds 49 additional COVID-19 research projects – details of the funded projects.* Retrieved from https://www.canada.ca/en/institutes-health-research/news/2020/03/government-of-canada-funds-49-additional-covid-19-research-projects-details-of-the-funded-projects.html.

Grady, P. (2015). Questions and answers. *Global Health Matters, 14*(1), 5.

Hall, J. G., Bainbridge, L., Buchan, A., et al. (2006). A meeting of the minds: Interdisciplinary research in the health sciences in Canada. *Canadian Medical Association Journal, 175*(7), 753–761.

Greenhalgh, T., Howick, J., & Maskrey, N. (2014). Evidence based medicine: A movement in crisis? *BMJ, 348*, g3725.

Hansson, S. O. (2017). Science and pseudo-science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition). Stanford, CA: Stanford University. https://plato.stanford.edu/archives/sum2017/entries/pseudo-science.

Harvey, G., Gifford, W., Cummings, G., Kelly, J., Kislov, R., Kitson, A., Pettersson, L., Wallin, L., Wilson, P., & Ehrenberg, A. (2019). Mobilising evidence to improve nursing practice: A qualitative study of leadership roles and processes in four countries. *International Journal of Nursing Studies, 90*, 21–30. https://doi.org/10.1016/j.ijnurstu.2018.09.017.

Health Canada. (2018). *A common statement of principles on shared health priorities.* Retrieved from https://www.canada.ca/en/health-canada/corporate/transparency/health-agreements/principles-shared-health-priorities.html.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Macmillan.

Kitson, A., Carr, D., Conroy, T., et al. (2019). Speaking up for fundamental care: The ILC Aalborg Statement. *BMJ Open*, *9*(12). http://dx.doi.org/10.1136/bmjopen-2019-033077.

Melnyk, B. M., & Fineout-Overholt, E. (Eds.). (2011). *Evidence-based practice in nursing & healthcare: A guide to best practice*. Philadelphia, PA: Lippincott Williams & Wilkins.

National Institute of Nursing Research. (2015). *Overview of global health research*. Retrieved from https://www.ninr.nih.gov/researchandfunding/globalhealth.

Nightingale, F. (1863). *Notes on hospitals*. London: Longman Group.

Paley, J., Cheyne, H., Dalgleish, L., Duncan, E. A., & Niven, C. A. (2007). Nursing's ways of knowing and dual process theories of cognition. *Journal of Advanced Nursing, 60*(6), 692–701. https://doi.org/10.1111/j.1365-2648.2007.04478.x.

Potter, P., Duggleby, W., Stockert, P., Astle, B., Perry, A., & Hall, A. (Eds.). (2019). *Canadian fundamentals of nursing* (6th ed.). Toronto: Elsevier Canada.

Registered Nurses' Association of Ontario [RNAO]. (2016). *Best practice guidelines*. Retrieved from http://rnao.ca/bpg.

Richards, D. A. (2020). *COVID-NURSE*. Retrieved from http://blogs.exeter.ac.uk/covid-nurse/.

Richards, D. A., Hanssen, T. A., & Borglin, G. (2018). The second triennial systematic literature review of European nursing research: Impact on patient outcomes and implications for evidence-based practice. *Worldviews on Evidence-Based Nursing, 15*(5), 333–343. https://doi10.1111/wvn.12320.

Richards, D. A., Hilli, A., Pentecost, C., Goodwin, V. A., & Frost, J. (2018). Fundamental nursing care: A systematic review of the evidence on the effect of nursing care interventions for nutrition, elimination, mobility and hygiene. *Journal of Clinical Nursing, 27*(11-12), 2179–2188. https://doi.org/10.1111/jocn.14150.

Sackett, D. L., Rosenberg, W., Gray, J. A., et al. (1996). Editorial: Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*(7023), 71.

Semeniuk, I. (2016). Pushback from scientists forces rethink of grant competition process. *The Globe and Mail*. Retrieved from http://www.theglobeandmail.com/news/national/pushback-from-scientists-forces-overhaul-of-funding-system/article30931215/.

Solman, A. (2017). Nursing leadership challenges and opportunities. *Journal of Nursing Management, 25*(6), 405–406. https://doi.org/10.1111/jonm.12507.

Stajduhar, K., Mollison, A., Giesbrecht, M., et al. (2019). "Just too busy living in the moment and surviving": Barriers to accessing health care for structurally vulnerable populations at end-of-life. *BMC Palliative Care*, *18*(11). https://doi.org/10.1186/s12904-019-0396-7.

Stajduhar, K., Sawatzky, R., Cohen, S. R., et al. (2017). Bereaved family members' perceptions of the quality of end-of-life care across four types of inpatient care settings. *BMC Palliative Care, 16*(1), 59. https://doaj.org/article/d2caf13917824bc58502028154915d72.

Sutherland, N., Ward-Griffin, C., McWilliam, C., & Stajduhar, K. (2016). Structural impact on gendered expectations and exemptions for family caregivers in hospice palliative home care. *Nursing Inquiry, 24*(1), e12157. https://doi.org/10.1111/nin.12157.

Turale, S., & Kunaviktikul, W. (2019). The contribution of nurses to health policy and advocacy

requires leaders to provide training and mentorship. *International Nursing Review, 66*(3), 302–304. https://onlinelibrary.wiley.com/doi/full/10.1111/inr.12550.

Vandyk, A., Chartrand, J., Beké, É., Burlock, L., & Baker, C. (2017). Perspectives from academic leaders of the nursing faculty shortage in Canada. *International Journal of Nursing Education Scholarship, 14*(1). https://doi.org/10.1515/ijnes-2017-0049.

World Health Organization. (n.d.). *Networks of WHO collaborating centres.* Retrieved from http://www.who.int/collaboratingcentres/networks/networksdetails/en/index1.html.

World Health Organization. (2018). *Alcohol*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/alcohol.

World Health Organization. (2020). *Depression.* Retrieved from https://www.who.int/en/news-room/fact-sheets/detail/depression.

CHAPTER 2

Theoretical Frameworks

Sarah Stahlke

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Define key concepts in the philosophy of science.
- Identify and differentiate between theoretical/empirical, aesthetic, personal, sociopolitical, and ethical ways of knowing.
- Identify assumptions underlying the post-positivist, critical social, and interpretive/constructivist views of research.
- Compare inductive and deductive reasoning.
- Describe how a framework guides research.
- Differentiate between conceptual and operational definitions.
- Describe the relationships among theory, research, practice, and leadership.
- Discuss levels of abstraction related to frameworks guiding research.
- Describe the points of critical appraisal used to evaluate the appropriateness, cohesiveness, and consistency of a framework guiding research.

KEY TERMS

aim of inquiry
concept
conceptual definition
conceptual framework
constructivism
constructivist paradigm
context
critical social theory
critical social thought
deductive reasoning
epistemology
hypothesis
inductive reasoning
methodology
model
ontology
operational definition
paradigm
philosophical beliefs
positivism
post-positivism
qualitative research
quantitative research
text
theoretical framework
theory
values
worldview

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

THE NATURE OF KNOWLEDGE

AS YOU LEARNED IN CHAPTER 1, NURSES DEVISE clinical questions, based on their daily practice experiences, that, if answered, can improve the care they provide to individuals, families, and communities. Each question requires that clinicians and nurse researchers engage in a knowledge development process (Fig. 2.1). The process begins with the identification of *knowledge gaps:* the absence of theoretical or scientific knowledge relevant to the phenomenon of interest. *Knowledge generation* occurs next, with the conduct of research that provides answers to well-thought-out research questions. This *knowledge* is then *distributed* through journal articles, textbooks, and public presentations to nurses. Next, the *knowledge* is *adopted,* as nurses alter their practice based on published information or as health care organizations develop policies and protocols that are informed by newly generated knowledge. Finally, *knowledge is reviewed and revised* as new health issues arise, advances in clinical practice occur, or knowledge becomes outdated. In this chapter, we focus specifically on theoretical frameworks and how they guide and inform *knowledge generation.*

Nursing knowledge is created and interpreted at various **levels of abstraction**, ranging from the most abstract to the most concrete thinking (Butts, 2015). Fawcett has identified five components of nursing knowledge, which span a range of abstraction levels. These include metaparadigm, philosophy, conceptual model, theory, and empirical indicator (Butts, 2015). The metaparadigm is the most abstract level of knowledge in nursing. It is the worldview of the discipline, which distinguishes its focus (Butts, 2015). Philosophy addresses questions about existence, reality, knowing, and ethics as they pertain to nursing (Butts, 2015). Conceptual models are a set of concepts that address broad, general ideas of interest to the discipline, while theories translate concepts into testable questions that can be explored using empirical indicators such as instruments, experiments, or procedures (Butts, 2015).

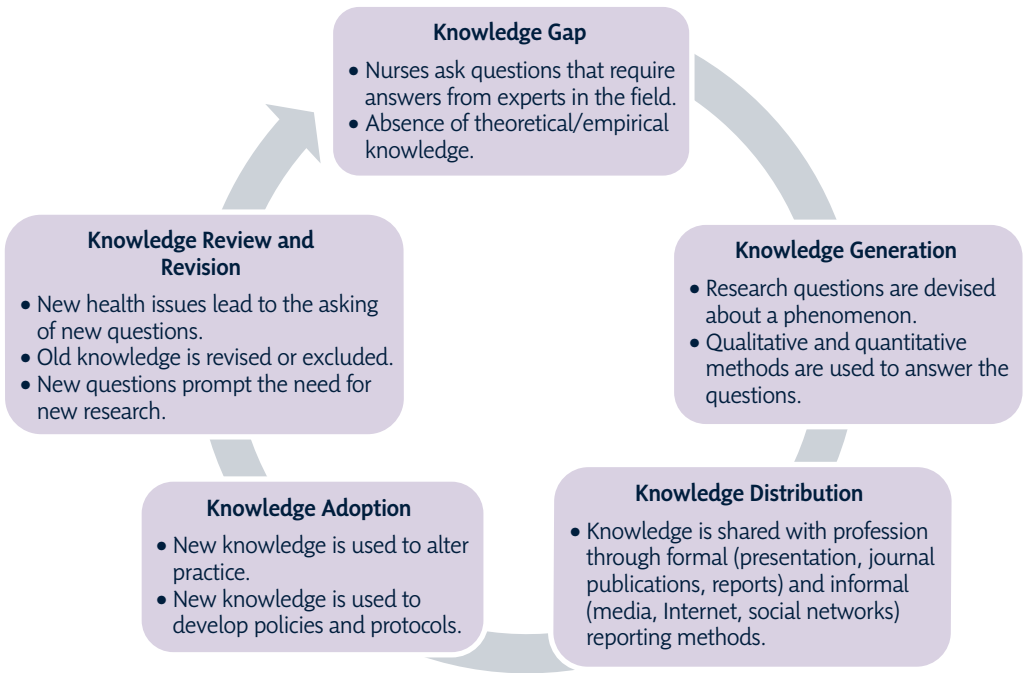


FIG. 2.1 Knowledge development process.

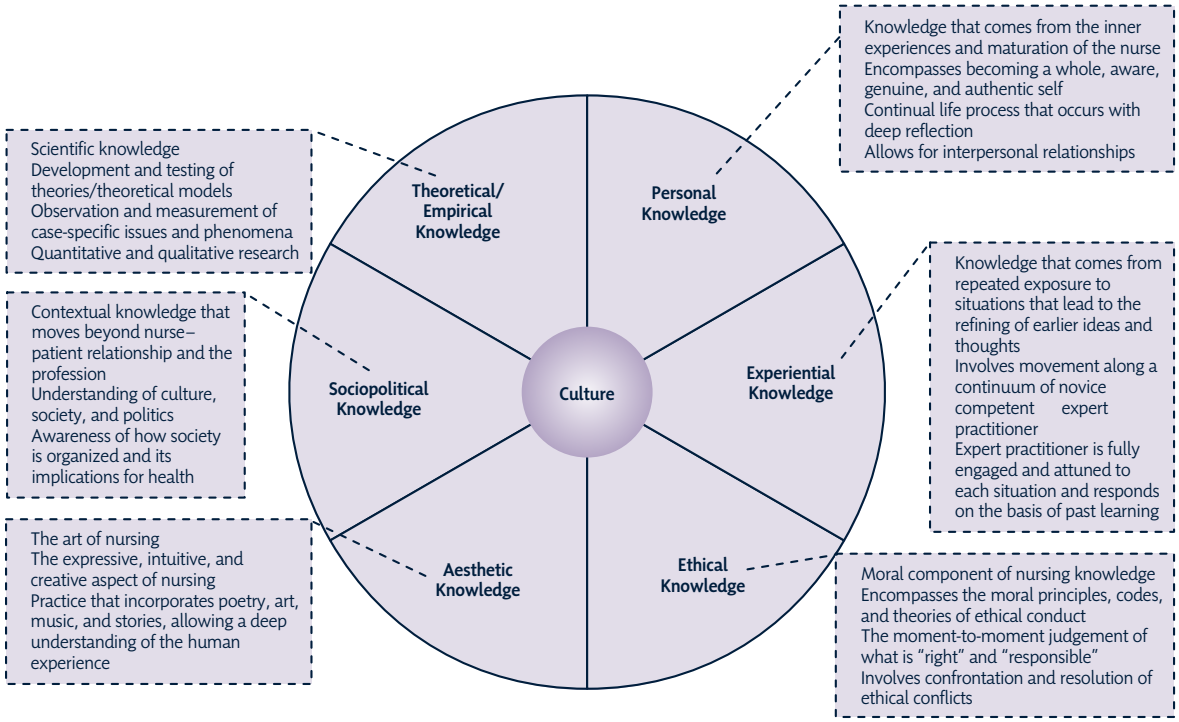


FIG. 2.2 Nursing knowledge.
(Adapted from Chapter 2 of the previous edition of this book, written by Cherylyn Cameron.)

Figure 2.2 outlines the various ways by which nurses inform their practice. These include *theoretical/empirical, personal, experiential, ethical, aesthetic,* and *sociopolitical/cultural* ways of knowing (Chinn & Kramer, 2015; Zander, 2007). Empirical knowing is acquired through the scientific process, which results in tested and justifiable knowledge for practice (Zander, 2007). In contrast, aesthetics is the art of nursing in which the nurse interprets the patient's behaviour and needs and addresses the bigger picture of patient care (Zander, 2007). Ethical knowing is the moral component of nursing and is concerned with justifying and judging actions. This knowing depends on an understanding of philosophical concepts about what is good as well as rights and obligations (Zander, 2007). Personal knowing involves an existential awareness of self and others in relationship (Zander, 2007). Experience and intuition constitute other forms of knowing in nursing (Zander, 2007). Understanding the sociopolitical context of practice is an overarching way of knowing; nurses can situate their practice and knowledge within a societal context (Zander, 2007).

It is important to remember that the way we understand these aspects of knowledge is based on our cultural perspectives, such as Western ideologies or Indigenous ways of knowing. Indigenous perspectives are covered in depth in Chapter 7, but a brief introduction is warranted here. Indigenous knowledge is "a learned way of looking at the world that may have very different forms of acquisitions, transmission, and manifestation for Indigenous peoples" (Battiste & Henderson, 2000, p. 48). Indigenous knowledge is transmitted through oral traditions, which provides a way of understanding Indigenous people's experiences and promotes a continued existence of their culture. Battiste (2000) rightly

notes that Indigenous knowledge is essential for transformation toward a just society that will "engage with and react to multiple circumstances and shapes of oppression, exploitation, assimilation, colonization, racism, genderism, ageism, and the many other strategies of marginalization" (p. xxi). Battiste (2002) claimed that Indigenous epistemology is located in theories, philosophies, histories, ceremonies, and stories as ways of knowing. The guiding tenets of Indigenous epistemology encompass a legitimate way of knowing, storytelling as a method for sharing knowledge, and the relationship between the researcher and participants as a natural component of the research process (Kovach, 2005). Indigenous epistemology is derived from multiple sources including traditional knowledge from previous generations, empirical knowledge from careful observation, personal knowledge and experiences, oral transmission, holistic knowledge, and experimental knowledge (Castellano, 2000).

From the Western perspective that is common in Canada, theoretical/empirical knowledge is most commonly referred to as *scientific knowledge*. Recently, theoretical/empirical knowledge has gained prominence in nursing with the increasing focus on evidence-informed practice. Theoretical and empirical knowledge really cannot be separated; however, theoretical knowing is concerned with developing or testing theories (possible explanations or ideas) that nurse researchers have about how the world operates. Theoretical knowing is informed by empirical knowing, which involves observations of reality. Observations may include the following:

1. Speaking with people about their life experiences (e.g., living with Alzheimer's disease) and using their responses to specific and general questions to understand the phenomenon
2. Observing social or cultural interactions (e.g., homeless individuals interacting with service providers) as they naturally occur, interpreting what the interactions might mean for both parties, and using those interpretations to develop theories about health service delivery for that population
3. Delivering an intervention (e.g., a school health program for obese children) and assessing changes in health care–related behaviours (e.g., type of foods consumed, amount of daily exercise, weight and other biomedical parameters) after the delivery of the intervention
4. Using surveys or a questionnaire to ask a large group of people questions about health experiences or their current symptom levels with regard to, for example, pain, digestive problems, or depression.

Taking an example of published work, Nahm and colleagues (2010) hypothesized that using social cognitive theory (involving self-efficacy, mastery, and modelling) as the foundation for a hip fracture prevention website would be more effective in changing individuals' health behaviours than conventional educational strategies; this **hypothesis** is an example of *theoretical knowing*. However, it was only through developing a clinical trial that tested the intervention that Nahm and colleagues could observe the effectiveness of the theory-based educational website (e.g., differences in knowledge and health behaviours) compared to the effectiveness of a conventional prevention website. Their results provided support for the hypothesis; this support is an example of *empirical knowing*.

PHILOSOPHIES OF RESEARCH

Thus far, we have used a number of terms that may be new to you. Every discipline has characteristic terminology for communicating important features of the work in that field. Learning new terminology is part of what nursing students do when they learn research methods and skills. Each research method and all philosophies of science have specialized language that nursing students will encounter in the literature. Thus, to help you comprehend the research you will read, it is important to clarify a few terms.

All research is based on **philosophical beliefs** about the world; these beliefs are the motivating values, concepts, principles, and the nature of human knowledge of an individual, group, or culture, and they are the basis of a **worldview**, or **paradigm**. *Paradigm* is from the Greek word *paradeigma,* meaning "pattern." Paradigms are "different ways of viewing the world and often form the foundation from which research is undertaken. They consist of a set of assumptions about what is reality, how knowledge is created and what is valuable to learn" (Davies & Fisher, 2018, p. 21). Therefore, knowing and comprehending these views and practices is important in understanding and using research findings. They are not right or wrong; rather, they represent different ways of viewing the world and the way things operate within it. Nursing researchers may tend to gravitate toward a particular worldview but it is important to remember that different kinds of research questions call for different approaches to understanding and studying reality, so it is possible for researchers to value different worldviews at different times. For example, a researcher can know that a certain drug has side effects or that turning bedridden patients regularly will reduce the risk of skin breakdown but also understand that people's narratives about their experiences will yield multiple and sometimes conflicting explanations. This shows how different paradigms can be held simultaneously.

It is important to note that there can be overlap between paradigms and they are not always exclusive of one another (Davies & Fisher, 2018), although categorizations can simplify learning about the basic worldviews. In nursing, there are three key paradigmatic perspectives that are commonly held and that generally underpin various approaches to research. These are positivism/post-positivism, constructivism, and critical theory. These three paradigms are compared in Table 2.1; however, first you need to understand the philosophical language used in the table. **Ontology** (from the Greek word *onto,* meaning "to be") is the science or study of being or existence and its relationship to nonexistence. Ontology addresses two primary questions: (1) What exists or what is real? and (2) Into what categories can existing things be sorted? **Epistemology** addresses four key questions: (1) What is knowledge? (2) What are the sources of knowledge? (3) What are the ways we come to know something, in contrast to believing, wondering, guessing, or imagining? and (4) What is truth and what role does it play in knowledge? **Methodology** refers to the principles, rules, and procedures that guide the process through which knowledge is acquired. The **aim of inquiry** refers to the goals or specific objectives of the research. **Context** refers to the personal, social, organizational, cultural and/or political environment in which a phenomenon of interest (that "thing of interest") occurs. The context of research studies can include physical settings, such as the hospital or home, or less concrete "environments," such as the context that cultural understandings bring to an experience. **Values** are the things that the nurse researcher holds to be important.

Positivism is a philosophical orientation that arose in the 18th century, as part of a movement away from religious knowing to reasoning and science. Positivistic research aims for objectivity and impartiality, with a goal of producing unbiased, generalizable research. **Post-positivism** emerged in response to the realization that such objectivity is usually not possible and our observations cannot always be relied upon because they are subject to error and human bias—we all have different values, cultures, and life experiences and, thus, generate different interpretations.

Constructivism, which is also referred to as interpretivism or relativism, points to the centrality of human experiences, social and cultural constructs, values, perspectives, and language (Clark, 2008). It is a philosophical orientation that suggests that reality and the way in which we understand our world are largely dependent on our perceptions and context. Truth about life experience is regarded as relative and multiple rather than absolute. The value of

knowledge development lies in the ability to understand how people perceive their world. Knowledge development occurs through observation, dialogue with people, or both, and as a result of paying attention to the language people use to describe life experiences. Constructivists tap into personal experience rather than seeking measurement and elusive objectivity. This form of research is aimed at creating an understanding of people and their life experiences from their point of view.

Critical realism is a philosophical orientation that has become of great interest in nursing research. It harnesses the strengths and addresses the weaknesses of positivism and constructivism (Clark, 2008) by acknowledging that, although there is a single reality, we cannot know it for sure (Davies & Fisher, 2018). This perspective offers a middle ground that "does not reduce the world to unknowable chaos [as might be the case with relativism] or a positivistic universal order" (Clark, 2008, p. E68). In critical realism, social entities exist independently of human understanding. For example, discrimination and power imbalances exist regardless of whether humans recognize their influence.

TABLE 2.1

BASIC RESEARCH PARADIGMS

ITEM/QUESTIONS	POST-POSITIVISM	CRITICAL THEORY	CONSTRUCTIVISM
ONTOLOGY			
What can be said to exist? Into what categories can we sort existing things?	A material world exists. Not all things can be understood, sensed, or placed into a cause-and-effect relationship. The senses provide us with an imperfect understanding of the external/material world.	Reality is constructed by those with the most power at particular points in history. Over time, reality is shaped by numerous social, political, economic, and cultural forces. Stories/discourses shaped by the powerful become accepted reality.	Reality is constructed by individual perception within a social context. Truth is relative and based on perception or some particular frame of reference.
EPISTEMOLOGY			
What is knowledge? How is knowledge acquired? How do we know what we know?	Researchers are naturally biased. Objectivity (controlled bias) is the ultimate goal, although pure objectivity is not attainable. Uses triangulation and replication of findings across multiple perspectives.	Research is a transaction that occurs between the researcher and research participant. The perceptions (standpoint) of the researcher and the research participants naturally influence knowledge generation/creation. Perceptions (standpoints) are determined by context, and so contextual awareness and its relationship to the participant's understanding of reality is the focus of the research.	Research is a transaction that occurs between the researcher and research participant. Research emphasizes the meaning ascribed to human experiences. Objectivity is not possible; knowledge is co-created.
RESEARCHER'S VALUES			
How do the researcher's values influence the knowledge development process?	All attempts are made to exclude researcher bias. Influence is denied.	Researcher perspectives are acknowledged and recognized as influential. Researcher values drive research questions and purpose; researcher manages own perspectives.	Researcher perspectives, values, experiences are recognized as potentially influential. Influence is managed with reflection and bracketing.

Continued

TABLE 2.1

BASIC RESEARCH PARADIGMS—cont'd

ITEM/QUESTIONS	POST-POSITIVISM	CRITICAL THEORY	CONSTRUCTIVISM
METHODOLOGY			
Within a particular discipline, what principles, rules, and procedures guide the process through which knowledge is acquired?	Inquiry generally involves quantitative methods and is viewed as a series of logically related steps. Research questions/hypotheses are proposed and subjected to empirical testing. Research is characterized by careful accounting for and control of factors that may influence research findings. Qualitative research may be used to develop hypotheses.	Inquiry requires dialogue between the investigator and research participant. Research purposes are transformative, emancipatory, or consciousness raising. Dialogue brings to the forefront the historical context behind experiences of suffering, conflict, and collective struggles. Dialogue increases participants' awareness of actions required to incite change.	Inquiry requires dialogue between the investigator and research participant. Focus is on interpretation of interview data, documents, and artifacts, including written texts, art, pictures. Interpretation brings to the forefront the varying ways in which people construct their understanding of their social world and how their interpretation shifts as they interact with others.
AIM OF INQUIRY			
What is the goal of research?	Explanation, prediction, and control.	Critique, change, reconstructed reality, and emancipation.	Understanding of social realities and interpretations of meaning.
CONTEXT			
What biographical, life, social, and political factors may influence the research findings?	Attempt is made to set aside researcher histories, politics, ideologies, and experiences.	Focus is on historical, social, and political context. *Context* refers to the social and political climate in which an event or process occurred. Social context highlights how structural, economic, representational, and institutional factors of the past influence how people understand an issue today. Political context highlights how political dialogue and opinions, legal directives, and government policies of the past influence how people understand an issue.	Focus is on life context, including significant conditions and demands that provide greater understanding of the phenomena being studied; focus also emphasizes time and place.
SUB-PARADIGMS/ THEORIES/ METHODS	Critical realism Mixed methods	Feminism Critical Race Theory Intersectionality Disability Studies Critical Policy Analysis	Symbolic Interactionism Philosophy Qualitative approaches

Adapted from Denzin, N. K., & Lincoln, Y. S. (2000). *Handbook of Qualitative Research.* SAGE Publications Inc.

However, human perceptions and experiences can still be incorporated into an understanding of reality (Clark, 2008). From a research perspective, truth and reality are ascertained through varied approaches to data collection and analysis to increase the accuracy with which social phenomena are understood (Davies & Fisher, 2018).

Critical social theory emphasizes that reality and our understanding of reality are constructed by people with the most power in a particular time and place. This perspective supports the understanding that health and other aspects of reality are shaped by numerous social, political, economic, and cultural factors. Such factors include gender, sexual orientation, class and economic status, race and ethnicity, ability, and geographic location. In nursing research, a strong emphasis is placed on understanding how power imbalances associated with these factors influence health and well-being, access to health care, and patient experiences and outcomes. The goal of critical research is emancipation and social change (Davies & Fisher, 2018).

Some specific types of critical theory include feminist theory, critical race theory, queer theory, disability theory, and intersectionality. Feminist theories are numerous and varied (e.g., Marxist, poststructuralist, cultural, eco) but generally focus on the experiences of women, although this can be broadened to study the experiences and perspectives of all genders, especially in terms of stereotyping, marginalization (exclusion), and emancipation. The primary intent of feminist theories is to dismantle systems of oppression and to raise awareness of gender disparities in a more meaningful way. Rolls and Young (2012) assert that feminist nursing research ought to benefit women; value women's experiences, ideas, and needs; recognize the structural, interpersonal, and ideological conditions that oppress women; and include a portrayal of women's strengths (p. 18). To apply this, they conducted a review of the research literature about older women's experiences of heart failure. They found that there were very few studies on this topic because men have been the subjects of most research about cardiovascular disease; they call for more women-focused studies to inform nursing practice more broadly.

Critical race theory provides a foundation for "studying and transforming the relationship among race, racism, and power" (Delgado & Stefancic, 2001, p. 2). Critical race theorists point out that racism is an everyday experience that is difficult to address because it is not acknowledged. They also note that race is an idea that benefits the dominant group so they have no incentive to eliminate racism (Wesp et al., 2018). Wesp and colleagues (2018) evaluated models and guidelines for cultural competency in nursing from a critical race perspective. They noted that most nurses would agree that there is a need to reduce culturally based disparities in health care, which cultural competency guidelines seek to address. However, they show that these guidelines are flawed because they fail to consider the prejudicial attitudes and imbalances of power that affect patient care and nursing practice. Instead of promoting cultural competence, these guidelines can ironically perpetuate stereotypes and fail to produce change.

Queer theory offers a critique of normative understandings of sex, gender, and sexuality, combining these with other aspects of social difference such as race and class (Hall & Jagose, 2013). Disability theory critiques the medical view of disability as an individual defect that requires cure or elimination. Critical disability theories allow scholars to explore and challenge social injustice and oppression, which equate ability with human worth (Siebers, 2008). Intersectionality is a critical theoretical perspective that combines various critical theories. This shifts the focus from single aspects of identity (such as gender or race) toward a recognition that these aspects work together in people's lived experience. The way that these social forces, identities, and ideologies combine in people's lives has an impact on how much power they have or lack (Crenshaw, 2017). This has implications for people's expectations and

experiences of care and for the ways that nurses relate to patients/clients (Van Herk et al., 2011).

Van Herk and colleagues (2011) point out that nurses do not use intersectionality and critical theories as much as they should. They point out that nurses need to have a critical awareness of how power works to create privilege for some and oppression for others in order to provide safe, meaningful, and high-quality care.

Research Hint

Values are involved in all research. For the positivist/post-positivist, it is assumed that values will be held at a distance to minimize their influence. For critical social and constructivist researchers, values and their potential influences on the research results are incorporated more overtly in the research process. However, it is important to remember that all researchers have values and these influence research, regardless of whether they are acknowledged or made explicit.

RESEARCH METHOD\S: QUALITATIVE AND QUANTITATIVE

Research methods are the techniques, procedures, and processes used by researchers to organize a study so that it provides answers to the research question. Research methods can be classified into two major categories: qualitative and quantitative. A researcher chooses between these categories primarily on the basis of the question the researcher is asking. If a researcher wishes to test a cause-and-effect relationship, such as how different levels of social support (cause) leads to high blood pressure (effect), quantitative methods are most appropriate. If, however, a researcher wishes to discover and understand the meaning of an experience or process, such as death and dying, a qualitative approach would be appropriate. A researcher can also design a study that combines both categories, which is known as mixed-methods.

Qualitative research is a systematic, interactive research method used to describe and interpret life experiences. The emphasis is on capturing the personal perceptions of the study participants. Figure 2.3 outlines the qualitative research process. A researcher would choose to conduct a qualitative research study if the question to be answered concerns the illumination and understanding of human experience, such as illness, loss, or life change.

A study by Smith and colleagues (2018) demonstrates the qualitative research process. They investigated mothers' experiences of supporting their adolescent children though treatment for substance use disorder. The research process as described in Figure 2.3 was used in this study as follows:

Step 1: The researchers began with a broad research question so that they could freely explore the experience rather than narrow it from the start. They state that "the purpose of this study was to gain insights into the women's experiences, to explore how they composed themselves as mothers in the midst of dominant and competing stories of motherhood, family, and substance abuse" (p. 512). The researchers supported their research question with a literature review that covered statistics about the prevalence of substance abuse, the impact of substance abuse on families, and the perceptions and self-perceptions that exist about mothers of children with substance use disorders.

Step 2: A purposive sample of four mothers was selected from one family-oriented, long-term treatment centre for adolescents with alcohol and/or drug addictions. Mothers from different stages of the treatment program were included to explore their experiences at different stages.

Step 3: Each of the mothers participated in six one-on-one interviews as well as six group conversations. Meeting locations included the treatment centre, restaurants, and participants' homes. Data collection strategies were varied, "including conversations, genograms, talking about items of significance, telling personal and family stories, reviewing photos, and writing Haikus, journals, and letters" (p. 514).

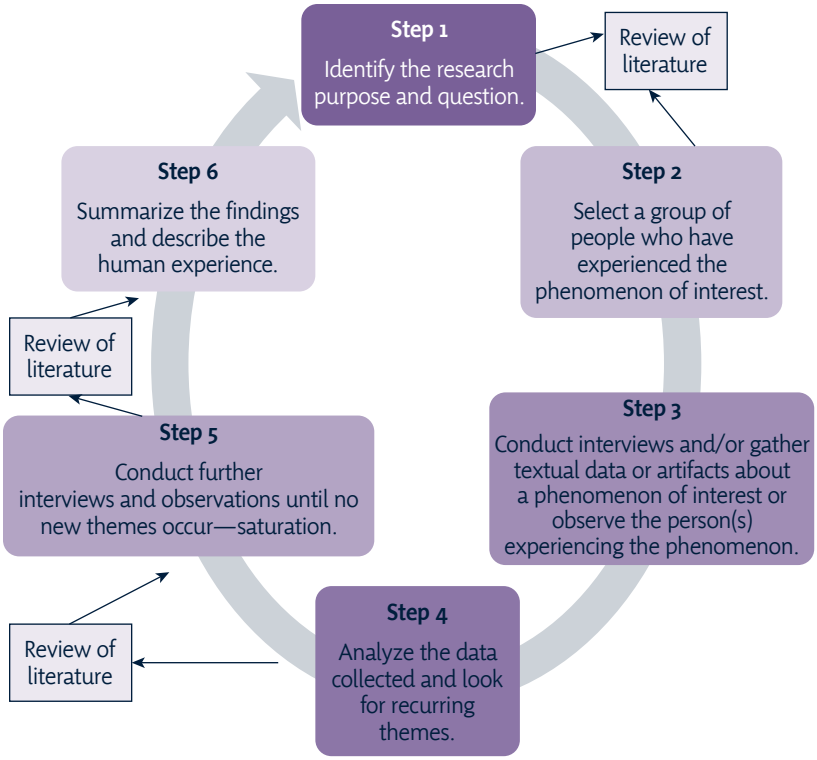


FIG. 2.3 Qualitative research process.

Step 4: After the interviews, "narrative accounts were co-composed with each participant" (p. 515). Four narrative threads (like themes) were theorized from the narratives.

Step 5: Several interviews were held with each participant over a year-long period to generate rich stories of the experience over time.

Step 6: The findings are presented and then linked to the existing literature.

As illustrated by this study, qualitative research is generally conducted in natural settings. Data are usually words and/or images rather than numbers. Sample sizes are often relatively small when compared to quantitative research because the goal is to study a phenomenon in depth with people who can speak knowledgeably about an experience. Thus, data from qualitative studies help nurses understand experiences or phenomena that affect patients, and this information in turn leads to improved care and stimulates further research. Chapters 8 and 9 provide an in-depth overview of the underpinnings, designs, and methods of qualitative research.

Where the purpose of qualitative research is to explore, describe, and/or explain a phenomenon or generate theory, a researcher would choose to conduct a **quantitative** research study if the question to be answered concerned testing for the presence of specific relationships, assessing for group differences, clarifying cause-and-effect interactions, or explaining how effective a nursing intervention was. Quantitative methods entail the use of precise and controlled measurement techniques to gather data that can be analyzed and summarized statistically. Figure 2.4 outlines the quantitative research process. Like the qualitative research process, the quantitative research process begins with the development of a research question and a purpose statement that highlight a relationship between two things.

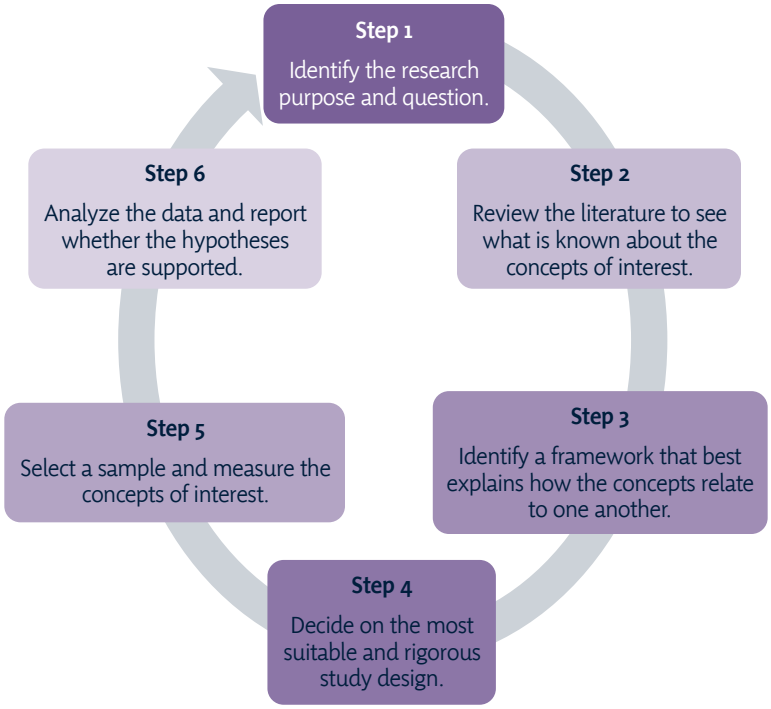


FIG. 2.4 Quantitative research process.

A study completed by El-Masri and colleagues (2014) demonstrates the quantitative research process. They explored the impact of cancer radiation treatment on men with prostate cancer. The research process as described in Figure 2.4 was used in this study as follows:

Step 1: They defined the purpose of the study as being "to compare the effects of 3 types of radiation treatment on functions, bother, and well-being in men with prostate cancer at 1, 6, and 12 months after completion of treatment" (p. 42).

Step 2: The research process began with a review of the literature, which can include journal articles, books, government documents, and even Internet sources, to determine what is known about the phenomenon of interest and theories that explain the phenomenon. In their introduction and review of the literature, the authors noted that although there were several studies on urinary, sexual, and bowel function after treatment for prostate cancer, there was a dearth of studies that followed patients through the first year of recovery after the radiation treatments. They noted that studies focusing on the well-being and psychosocial impacts were inconsistent.

Step 3: The study was to compare the effect of three different types of radiation on the functions, bother, and well-being of men with prostate cancer.

Step 4: The researchers established a baseline survey prior to treatment, followed by a series of subsequent surveys to measure patients' self-perceived functions, bother, and well-being. The surveys, with the exception of the demographic data collected during the baseline survey (i.e., age), were identical. The authors described the survey tool in detail and provided rationale for the choice of the surveys, including reliability and validity.

Step 5: The researchers recruited participants during an orientation class that all patients attended before commencement of their radiation treatment. To be considered for inclusion, patients had to (1) have a confirmed case of localized prostate cancer; (2) be about to undergo radiation treatment (one of the three types identified); (3) be able to read and understand English; and (4) be able to provide consent.

Step 6: After conducting the statistical analysis, the authors concluded that there were "no differences among the three radiation treatments with regard to any of the outcome variables" (p. 50). They found that there was improvement over time in all of the variables measured; however, the researchers noted that although sexual function improved over time, sexual function continued to be a concern at one year post-treatment.

An important part of the quantitative research process is to decide which design is most appropriate for answering the research question. The numerous choices include descriptive, correlational, longitudinal, quasi-experimental, and experimental designs. El-Masri and colleagues chose a longitudinal design, specifically a repeated survey method, to compare the function, bother, and well-being in men with prostate cancer who had undergone three different radiation regimens.

As demonstrated in the article by El-Masri and colleagues (2014), quantitative research techniques are systematic, and the methodology emphasizes control of the research process, the environment in which the study is conducted, and how each variable is measured. In contrast to qualitative approaches—in which a question is asked and the participant is responsible for providing an in-depth response—quantitative responses are restricted to a preselected set of responses.

When you read research articles, remember that researchers may vary the steps slightly, depending on the nature of the research problem, but all of the steps should be addressed systematically.

INTRODUCTION TO FRAMEWORKS FOR RESEARCH

Frameworks provide a general orientation to understanding a phenomenon of interest and identify what factors are most significant as we examine various aspects of health. A **concept** is an image or symbolic representation of an abstract idea. Chinn and Kramer (2015, p. 160) define concept as a "complex mental formulation of experience." Concepts offer ways to convey abstract ideas. A **conceptual framework** is a structure or assembly of concepts that is used as a map or a scaffolding of ideas for the study. Generally, a conceptual framework is a synthesis or integration of existing views and knowledge on a topic, developed from a review of the literature (Imenda, 2014). It defines the main ideas and the network of relationships between them, grounding the study in the knowledge bases that are relevant to the study. A specific theory may not be guiding the study but concepts always are (Rocco & Plakhotnik, 2009).

A **theoretical framework** is similar but distinct; a **theory** is a unified set of interrelated concepts that serves the purpose of explaining or predicting phenomena. A theory is like a blueprint, which depicts the elements of a structure and the relationship of each element to the other, just as a theory unifies and depicts both the concepts that compose it and how they are related (Imenda, 2014). Chinn and Kramer (2015) define theory as an "expression of knowledge… the creative and rigorous structuring of ideas that project a tentative, purposeful, and systematic view of phenomena" (p. 255). Theory guides practice and research; practice enables testing of theory and generates questions for research; and research contributes to theory building. Thus, what is learned through practice, theory, and research constitutes the knowledge of the discipline of nursing. When a researcher uses a theoretical framework, they are presenting a specific, unified, highly developed theory, such as systems theory or self-efficacy, and any previous empirical work

about or development of that theory (Imenda, 2014; Rocco & Plakhotnik, 2009).

The social determinants of health framework is one example that is often used in nursing, based on the understanding that the primary factors that shape the health of Canadians are not medical treatments or lifestyle choices but rather the inequitable living conditions they experience (Mikkonen & Raphael, 2010). This framework depicts how people's health and well-being are determined by 14 factors:

1. Aboriginal status
2. Gender
3. Disability
4. Housing
5. Early childhood development
6. Income and income distribution
7. Education
8. Race
9. Employment and working conditions
10. Social exclusion
11. Food insecurity
12. Social safety net
13. Health services
14. Unemployment and job security

A framework does not necessarily reveal how every possible factor relates to one another. The social determinants of health framework guides the researcher in addressing the relationship between health and any of the 14 factors just listed. The researcher can determine how many factors to focus on and in what way the factors are related to each other and, in turn, to health. In some cases, the researcher may develop a diagram or a pictorial representation of these relationships.

A **model** is a symbolic representation of a set of concepts that is created to depict relationships. Figure 2.5 shows Stewart et al.'s (2009) model of social support. It highlights the process through which support from peers and professionals influences the stressful life situations, coping behaviours, and health care–related behaviours of homeless youths. In this model, arrows are used to depict a process that explains how social support is related to the social network, stress, and health functioning. For example, the arrow from "social support" to "processes" suggests that social support has an effect on social network comparison, exchange, and learning. Whether this is positive or negative is unknown; however, the social network then influences coping behaviours (problem focused, support seeking, and emotion focused), which in turn influence health care–related functioning (loneliness, depression, drug use, and health behaviours).

Qualitative and quantitative research rely on different forms of reasoning to arrive at their conclusions. Inductive reasoning is the pattern of "figuring out what is there" from the details of the nursing practice experience and is the foundation for most qualitative inquiry. Research questions related to the issue of the meaning of experience for the patient can be addressed with the inductive reasoning of qualitative inquiry. Inductive reasoning is a process of starting with the details of experience and building toward a general picture. Deductive reasoning, followed by quantitative researchers, involves a process of starting with the general picture—in this case, the theory—and moving toward the specific. In deductive reasoning, the researcher measures concepts that, when combined, enable the researcher to suggest relationships between the concepts. Deductive reasoning begins with a structure that guides searching for "what is there." Inductive and deductive reasoning can both be used within a field of study or even within a single study to explore concepts and test variables, thereby building a comprehensive body of knowledge within a topic area.

Variables are factors that can take on a range of values (e.g., temperature, heart rate, pain, job satisfaction). The key empirical aspects of a study—its concepts and variables—are generally articulated through conceptual and operational definitions. A **conceptual definition** is much like a dictionary

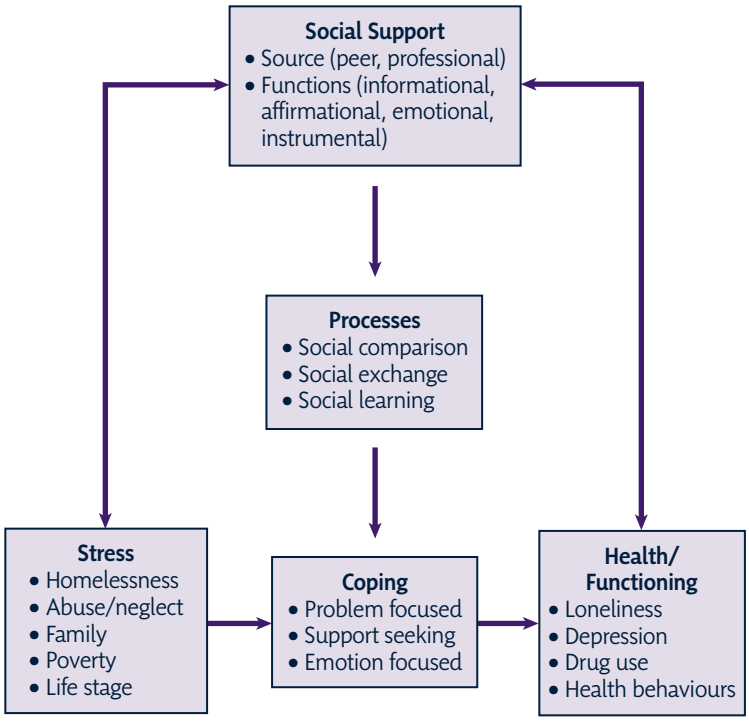


FIG. 2.5 Conceptual model.
(From Stewart, M. et al. (2009). A support intervention to promote health and coping among homeless youths. *Canadian Journal of Nursing Research, 41*(2), 54–77.) SAGE Publications Inc.

definition, conveying the general meaning of the concept. However, for research, the conceptual definition goes beyond the general meaning found in the dictionary; the concept is defined as it is rooted in the theoretical literature. The **operational definition** specifies how the concept will be measured—that is, what instruments will be used to capture the features of the variable.

The Critical Thinking Decision Path on p. 34 takes you through the thinking of a researcher who is about to begin conducting research. You can expect to find some, but not all, of the phases of decision making addressed in a research publication. Beginning with the research question, the researcher is inclined to approach a research problem from the perspective of inductive or deductive reasoning. Qualitative researchers, who pursue an inductive reasoning approach, generally do not present a framework before beginning the discussion of methods. This is not to say that the literature will not be reviewed before the methods are introduced. Qualitative researchers may use a conceptual or theoretical framework to inform or sensitize them to existing thinking and theorizing about their topic (Bowen, 2006) but avoid using a framework to structure their analysis so that they are able to see freely what their data contain.

Conversely, researchers who use deductive reasoning must choose between a conceptual and a theoretical framework. In the theory literature, these terms are used interchangeably (Chinn & Kramer, 2015), although, in the case presented in the Critical Thinking Decision Path, each term is distinguished from the other on the basis of whether the researcher is creating the structure or whether the structure has already been created by someone else. In general, each of these terms refers to a structure that provides guidance for research by assisting the researcher in

determining study variables and operational definitions. In other words, conceptual and then operational definitions will emerge from the framework.

Research Hint

Some research reports embed conceptual definitions in the literature review. The reader should find the conceptual definitions so that the logical fit between the conceptual and the operational definitions can be determined.

Research Hint

When researchers have used conceptual frameworks to guide their studies, you can expect to find a system of ideas, synthesized for the purpose of organizing thinking and providing study direction.

APPRAISING THE EVIDENCE

The Framework

The framework for research provides guidance for the researcher as study questions are fine-tuned, methods for measuring variables are selected (for quantitative research), and analyses are planned. Once data are collected and analyzed, the framework is used as a basis for comparison. The reader of research needs to know how to critically appraise a framework for research (see the Critiquing Criteria box below).

Evaluating frameworks for research requires skill that can be acquired only through repeated critique and discussion with other nurses who have critiqued the same publication. The novice reader of research must be patient while developing these skills. With continuing education and a broader knowledge of potential frameworks, you will build a repertoire of knowledge to enable you to judge the foundation of a research study, the framework for research.

CRITIQUING CRITERIA

1. Is the framework for research clearly identified?
2. Is the framework consistent with a nursing perspective?
3. Is the framework appropriate to guide research on the subject of interest?
4. Are the concepts and variables (if doing a quantitative study) clearly and appropriately defined?
5. Did the study present sufficient literature to support the selected concepts?
6. Is there a logical, consistent link between the framework, the concepts being studied, and the data collection strategies?
7. Are the study findings examined in relation to the framework?

CRITICAL THINKING CHALLENGES

- Explain the difference between research that is based on a constructivist paradigm and research that is based on a positivist paradigm.
- Discuss how a researcher's values can influence the results of a study. Include an example in your answer.
- You are taking an elective course in advanced pathophysiology. The professor compares the knowledge of various disciplines and states that nursing is an example of a nonscientific discipline, declaring in support of this position that nursing's knowledge has been generated with unstructured methods, such as intuition, trial and error, tradition, and authority. What assumptions has this professor made? How would you counter or support this position?
- How would you argue against the following statement: "As a beginning consumer of research, it is ridiculous to expect me to determine whether a researcher's study has an appropriate theoretical framework."

CRITICAL JUDGEMENT QUESTIONS

1. Nurses inform their practice through various ways of knowing. Which of the following is NOT true about how nurses use forms of knowledge in practice?
   a. Theoretical knowing is relevant for practice.
   b. Empirical knowing is acquired through the scientific process.
   c. Personal knowing is based on opinion rather than fact.
   d. Ethical knowing depends on an understanding of what is good.
2. Which of the following statements about research methods is true?
   a. Qualitative research is not systematic.
   b. Quantitative research eliminates bias.
   c. Quantitative research always produces better evidence.
   d. Qualitative research is more appropriate than quantitative research for questions about the meaning of an experience.
3. Critical social theory influences nursing because it:
   a. Shows nurses how power imbalances influence health.
   b. Focuses on women at the expense of men.
   c. Makes nurses view things negatively.
   d. Brings theories into nursing research that are interesting but not directly relevant to nursing practice.

KEY POINTS

- The scientific approaches used to generate nursing knowledge reflect both inductive and deductive reasoning.
- The interaction among theory, practice, and research is central to knowledge development in the discipline of nursing.
- Conceptual and theoretical frameworks can be created by the researcher as a result of the study findings or found in the existing literature and used to support the study.
- The use of a framework for research is important as a guide to systematically identify concepts and, for quantitative research, study variables.
- Conceptual and operational definitions are critical in the evolution of a study, regardless of whether they are explicitly stated in a research report.
- In developing or selecting a framework for research, knowledge may be acquired from other disciplines or directly from nursing. In either case, that knowledge is used to answer specific nursing questions.
- When you critique a framework for research, examine the logical, consistent link between the framework, the concepts for study, and the methods of data collection.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Battiste, M. (2000). *Reclaiming indigenous voice and vision*. Vancouver, BC, Canada: UBC Press.

Battiste, M. (2002). *Indigenous knowledge and pedagogy in First Nations education: A literature review with recommendations*. Ottawa, ON, Canada: National Working Group on Education and the Minister of Indian Affairs, Indian and Northern Affairs Canada.

Battiste, M., & Henderson, J. Y. (2000). *Protecting indigenous knowledge and heritage*. Saskatoon, SK, Canada: Purich Publishing Ltd.

Bowen, G. A. (2006). Grounded theory and sensitizing concepts. *International Journal of Qualitative Methods, 5*(3), 12–23.

Butts, J. B. (2015). Components and levels of abstraction in nursing knowledge. In J. B. Butts, & K. L. Rich (Eds.), *Philosophies and theories for advanced nursing practice* (pp. 87–108). Burlington, MA: Jones & Bartlett Learning.

Castellano, M. (2000). Updating Aboriginal traditions of knowledge. In G. Dei, B. Hall, & D. Rosenberg (Eds.), *Indigenous knowledges in global contexts* (pp. 21–36). Toronto, ON, Canada: University of Toronto Press.

Chinn, P. L., & Kramer, M. K. (2015). *Knowledge development in nursing. Theory and process* (9th ed.). St. Louis, MO: Elsevier.

Clark, A. (2008). Complex critical realism: Tenets and application in nursing research. *Advances in Nursing Science, 31*(4), E67–E79.

Crenshaw, K. (2017). *On intersectionality: Essential writings*. New York: The New Press.

Davies, C., & Fisher, M. (2018). Understanding research paradigms. *Journal of the Australasian Rehabilitation Nurses' Association, 21*(3), 21–25.

Delgado, R., & Stefancic, J. (2001). *Critical race theory: An introduction*. New York: New York University Press.

El-Masri, M. M., Fox-Wasylyshyn, S. M., Springer, C. D., et al. (2014). Exploring the impact of prostate cancer radiation treatment on functions, bother, and well-being. *Canadian Journal of Nursing Research, 46*(2), 42–56.

Hall, D. E., & Jagose, A. (Eds.). (2013). *The Routledge queer studies reader*. Abingdon: Routledge.

Imenda, S. (2014). Is there a conceptual difference between theoretical and conceptual frameworks? *Journal of Social Sciences, 38*(2), 185–195.

Kovach, M. (2005). *Emerging from the margins: Indigenous methodologies*. Toronto, ON: Canadian Scholars' Press/Women's Press.

Mikkonen, J., & Raphael, D. (2010). *Social determinants of health: The Canadian facts*. Toronto: York University School of Health Policy and Management.

Nahm, E., Barker, B., Resnick, B., Covington, B., Magaziner, J., & Brennan, P. F. (2010). Effects of a social cognitive theory-based hip fracture prevention website for older adults. *Computers, Informatics, Nursing, 28*(6), 371–379.

Rocco, T. S., & Plakhotnik, M. S. (2009). Literature reviews, conceptual frameworks, and theoretical frameworks: Terms, functions, and distinctions. *Human Resource Development Review, 8*(1), 120–130.

Rolls, T. P., & Young, L. E. (2012). Disrupting the biomedical discourse: Older women's lived experiences with heart failure: A feminist review of the literature. *Canadian Journal of Cardiovascular Nursing, 22*(1), 18–25.

Siebers, T. (2008). *Disability theory*. Ann Arbor, MI: University of Michigan Press.

Smith, J. M., Estefan, A., & Caine, V. (2008). Mothers' experiences of supporting adolescent children through long-term treatment for substance use disorder. *Qualitative Health Research, 28*(4), 511–522.

Stewart, M., Reutter, L., Letourneau, N., & Makwarimba, E. (2009). A support intervention to promote health and coping among homeless youths. *Canadian Journal of Nursing Research, 41*(2), 54–77.

Van Herk, K. A., Smith, D., & Andrew, C. (2011). Examining our privileges and oppressions: Incorporating an intersectionality paradigm into nursing. *Nursing Inquiry, 18*(1), 29–39.

Wesp, L. M., Scheer, V., Ruiz, A., Walker, K., Weitzel, J., Shaw, L., Kako, P. M., & Mkandawire-Valhmu, L. (2018). An emancipatory approach to cultural competency: The application of critical race, postcolonial, and intersectionality theories. *Advances in Nursing Science, 41*(4), 316–326.

Zander, P. E. (2007). Ways of knowing in nursing: The historical evolution of a concept. *Journal of Theory Construction and Testing, 11*(1), 7–11.

Critical Appraisal Strategies: Reading Research

Lorraine Thirsk

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Explain the importance of critical appraisal of research articles.
- Summarize the steps associated with appraising research articles.
- Use identified strategies to critically read research articles.
- Identify the format and style of research articles.
- Develop a strategy to read research articles effectively.

KEY TERMS

abstract
assumptions
critical appraisal
critical reading
critical thinking
critique
critiquing criteria
reliability
validity

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

AS YOU READ THIS TEXT, YOU will learn details of how the steps of the research process unfold. The steps are systematic and orderly, and they relate to the development of nursing knowledge. Understanding the step-by-step process that researchers use will help you develop the critiquing skills necessary to judge the soundness of research studies you will encounter in the literature. Throughout the chapters in this book, research terms pertinent to each step are identified, defined, and illustrated with many examples from the research literature. Five published research studies are featured in the appendices (A–E), and they are used as examples to illustrate significant points in each chapter. Judging not only a study's soundness but also its applicability to practice is a key skill. This chapter provides an overview of the format of research articles. It also introduces you to the critical appraisal skills you will need to be a knowledgeable research user. The chapter is designed to help you

begin to read research articles more effectively and with greater understanding.

Before you can assess a study, you need to understand the differences between and among studies. As you read the chapters and the appendices, you will encounter many different study designs, as well as standards for critiquing the soundness of each step of a study and for judging both the strength of evidence provided by a study and its applicability to practice. While the presentation of the research may vary between articles and types of research, all the steps should be addressed systematically. In general, the steps of the research process are reflected in the layout of the research article. The sections of the research article where these steps of the research process are described are presented in Table 3.1 for qualitative research and Table 3.2 for quantitative research.

As you continue to use and perfect your critical appraisal skills, remember that these very skills are an expected competency for delivering evidence-informed nursing practice.

CRITICAL APPRAISAL SKILLS

Your critical appraisal, also called a critique of the literature, is an organized, systematic approach to evaluating a research study or group of research studies. It involves the use of a set of established critical appraisal criteria to objectively determine the strength, quality, and consistency of evidence; these characteristics help you determine the applicability of the evidence to research, education, or practice. As a research user, you will become skilled at critically appraising research studies, combining the evidence with your clinical experience and the patient population you are caring for, to make an evidence-informed decision about the applicability of a particular nursing intervention for your patient or for the patient population in your practice setting.

TABLE 3.1

STEPS OF THE RESEARCH PROCESS AND JOURNAL FORMAT: QUALITATIVE RESEARCH

RESEARCH PROCESS STEPS OR FORMAT ISSUES	USUAL LOCATION IN JOURNAL HEADING OR SUBHEADING
Identification of the phenomenon	In abstract, introduction, or both
Purpose of research study	In abstract, at beginning or end of introduction, or in more than one of these locations
Literature review	In introduction, discussion, or both
Design	In abstract, "Introduction" section, "Methods" subsection titled "Design," "Methods" section in general, or more than one of these locations
Sample	In "Methods" subsection titled "Sample," "Subjects," or "Participants"
Legal–ethical issues	In section on data collection, in "Procedures" section, or in description of sample
Data-collection procedure	In "Data Collection" or "Procedures" section
Data analysis	In "Methods" subsection titled "Data Analysis" or "Data Analysis and Interpretation"
Results	In abstract (briefly), in separate section titled "Results" or "Findings"
Discussion and recommendations	In separate "Discussion" or "Discussion and Implications" section
References	At end of article

TABLE 3.2

STEPS OF THE RESEARCH PROCESS AND JOURNAL FORMAT: QUANTITATIVE RESEARCH

RESEARCH PROCESS STEPS OR FORMAT ISSUES	USUAL LOCATION IN JOURNAL HEADING OR SUBHEADING
Research problem	In abstract, introduction (not labelled as a research problem), or separate subsection titled "Problem"
Purpose	In abstract or introduction or both; at end of literature review or discussion of theoretical framework; or in separate section titled "Purpose"
Literature review	At end of introduction but not labelled as a literature review; in separate section titled "Literature Review," "Review of the Literature," or "Related Literature" Variables reviewed may appear as titles of sections or subsections
Theoretical framework, conceptual framework, or both	In "Literature Review" section (combined) or in separate sections titled "Theoretic Framework" and "Conceptual Framework"; or each concept or definition used in theoretical or conceptual framework may appear as title of separate section or subsection
Hypothesis/research questions	Stated or implied near end of "Introduction" section, which may be labelled; in separate sections or subsection titled "Hypothesis" or "Research Questions"; or, for first time, in "Results" section
Research design	In abstract or introduction (stated or implied) or in section titled "Methods" or "Methodology"
Sample: type and size	Size: may be stated in abstract, in "Methods" section, or in separate "Methods" subsection as "Sample," "Sample/Subjects," or "Participants" Type: may be implied or stated in any of previous headings described under size
Legal–ethical issues	In section titled "Methods," "Procedures," "Sample," "Subjects," or "Participants" (in all cases, stated or implied)
Instruments (measurement tools)	In section titled "Methods," "Instruments," or "Measures"
Validity and reliability	In section titled "Methods," "Instruments," "Measures," or "Procedures" (specifically stated or implied)
Data-collection procedure	In "Methods" subsection titled "Procedure" or "Data Collection" or in separate section titled "Procedure"
Data analysis	In "Methods" subsection under subheading "Procedure" or "Data Analysis"
Results	In separate section titled "Results"
Discussion of findings and new findings	Combined with results or in separate section titled "Discussion"
Implications, limitations, and recommendations	Combined with discussion or presented in separate or combined major sections
References	At end of article
Communicating research results	In research articles, poster, and paper presentations

As you read articles, you may notice the difference in style or format between research articles and theoretical or clinical articles. The terms in a research article may be new to you. Reading research articles is a new skill that gets easier with practice. As a student, you are not expected to completely understand a research article; nor are you expected to develop critiquing skills on your own. A primary objective of this book is to help you acquire the skills needed for critical appraisal. No perfect critique exists; your interpretation will be based on your current knowledge, experience, and understanding. Remember that becoming a competent critical thinker and research user takes time, patience, and experience.

The best way to become a knowledgeable research user is to use critical thinking skills when you read research articles. Critical thinking is the rational examination of ideas, inferences, assumptions, principles, arguments, conclusions, issues, statements, beliefs, and actions (Paul & Elder, 2008). As applied to reading research, this means that you are engaged in the following:

- Systematic understanding of the research process
- Thinking that displays a mastery of the criteria for critiquing research and evidence-informed practice
- The art of being able to make your thinking better (i.e., clearer, more accurate, or more defensible) by clarifying what you understand and what you do not know

Being a critical thinker means that you are consciously thinking about your own thoughts and what you say, write, read, or do, as well as what other people say, write, or do. While thinking about all of this, you are also questioning the appropriateness of the content, applying standards or criteria, and seeing how the information measures up. Take the time to reflect so that you more thoroughly consider your own thoughts and feelings and how they impact the decisions you make—your decisions affect not only you but the patients and clients you are caring for (Aveyard et al., 2015). As you read research articles, there are some key critical thinking questions, presented in Box 3.1, to keep in mind.

BOX 3.1

CRITICAL THINKING QUESTIONS

Who	Benefits? Is harmed? Makes decisions? Is most affected? Deserves recognition?
What	Is another perspective? Another alternative? A counter-argument? Is most/least important?
Where	Are there similar concepts/situations? In the world is this a problem? Can we get more information?
When	Is this acceptable/unacceptable? Would this benefit society? Cause problems? Has this played a part in our history?
Why	Is this a problem/challenge? Is this relevant? Should people know about this? Have we allowed this to happen?
How	Is this similar to _____? Does this disrupt things? Do we know the truth? Does this benefit/harm us/others? Can we change this?

Source: From Wabisabi Learning. https://wabisabilearning.com/.

Critical reading is "an active, intellectually engaging process in which the reader participates in an inner dialogue with the writer" (Paul & Elder, 2008, p. 461). A critical reader actively looks for assumptions (accepted truths), key concepts and ideas, reasons and justifications, supporting examples, parallel experiences, implications and consequences, and any other structured features of the text so as to interpret and assess the text accurately and fairly (Paul & Elder, 2008).

You will find that critical thinking and critical reading skills used in the nursing process can be transferred to understanding the research process and reading research articles. You will gradually be able to read an entire research article and reflect on it by identifying and challenging assumptions, identifying key concepts, questioning methods, and determining whether the conclusions are soundly based on the study's findings. Once you have obtained this competency in critiquing research, you will be ready to synthesize the findings of multiple research studies to use in developing evidence-informed practice.

STRATEGIES FOR READING AND CRITIQUING RESEARCH STUDIES

Critiquing a research study may require several readings, especially as you are developing critical reading skills. As you begin, you may need to read an article three or four times; this will get faster as you practice. Critical reading is a process that involves the following levels of understanding and allows you to critically assess a study's validity:

- Preliminary: familiarizing yourself with the content (skimming the article)
- Comprehensive: understanding the researcher's purpose or intent
- Analytical: understanding the parts of the study and developing a critique
- Synthesized: understanding the whole article and understanding how it fits with the cumulative body of knowledge

BOX 3.2

PRELIMINARY LITERATURE FILTERING PROCESS

TITLE

Look at the title to decide whether it addresses the subject matter you areinterested in.

ABSTRACT

Read the abstract in full to compare its content with the topic and your inclusion and exclusion criteria to see if the article addresses these.

FULL TEXT

Read the full text to compare the content with the topic, purpose of the study, and your inclusion and exclusion criteria to see if the article meets allyour requirements.

TYPE OF ARTICLE

In this final stage, you will need to decide if the article is what you want. For example, if you want to use only empirical research data in your review.

Source: Wakefield, A. (2014). Searching and critiquing the research literature. *Nursing Standard*, *28*(39), 49-57. https//:doi.10.7748/ns.28.39.49.e8867

Preliminary Reading

In this first step, keep your research textbook at your side so you can clarify unfamiliar terms and concepts as you read. The goal of a preliminary reading is to ensure the article is relevant for your purposes. Wakefield (2014) suggests a filtering process during this initial reading where you read through the title, then abstract, and then full article (Box 3.2). In addition, you may want to highlight the purpose statement or research question, as well as the identified steps of the research process.

You should also identify the type of article you are reading in this stage. The articles that you read may be broadly divided into research and non-research literature. Research literature includes quantitative studies, qualitative studies, mixed methods studies, and systematic reviews; non-research literature includes theoretical or methodological articles, review articles, dissertations and theses, as well as others (American Psychological Association, 2019). Definitions and explanations of the various types of articles are presented in Table 3.3.

Comprehensive Reading

The purpose of this next reading is to understand the aim or intent of the article.

You should be able to identify the main theme of the article in your own words, as well as identify the main steps of the research process, including the phenomenon under study or variables included in the research question (see Chapter 4). Continue to clarify terms and concepts that you are not familiar with. A list of questions is provided in Table 3.4 to help you with this comprehensive reading.

Analytical Reading

As you continue to read an article, you are ready to begin the appraisal process that will help determine a study's value. Critique is the process of

TABLE 3.3

TYPES OF SCHOLARLY ARTICLES

BROAD CATEGORY	SPECIFIC TYPE OF ARTICLE	DISTINGUISHING FEATURES
Research	Quantitative	Look for explicit use of the words research or study; evidence of ethics approval if it involves human or animal participants; format reflects Introduction, Methods, Results, Discussion; methods of research stated.
	Qualitative	
	Mixed-methods	
	Systematic review	
Non-research	Theoretical	Furthers a theory or theoretical argument.
	Methodological	Present new approaches to research processes or propose modifications to existing processes. These types of articles may discuss research but are not presenting results of research.
	Literature reviews	Not as rigorous or restrictive as a systematic review; may present numerous articles in a table format; offers a summary of previous literature as well as gaps, inconsistencies and relationships between the articles reviewed.
	Other	Often shorter articles such as letters to the editor, editorials, book reviews, comments or responses.

Source: Based on American Psychological Association (2019). Scholarly writing and publishing principles. In *Publication Manual of the American Psychological Association* (7th ed., pp. 3–26). American Psychological Association.

critical appraisal in which a person critically evaluates the content of a research report for scientific validity or merit and applicability to practice. It requires some knowledge of the subject matter, as well as knowledge of how to read critically and use critiquing criteria.

Research Hint

People often think that "being critical" is the same thing as "being negative." When offering a critique, it really means reading with a questioning mindset. While no research study is perfect, many articles will still be useful, and some will be great.

Critiquing criteria are the standards, appraisal guides, or questions used to judge (assess) an article. Guidelines for conducting a critique are presented in several of the following chapters of this book. In addition, there are many critical appraisal tools available, some of which are listed in Table 3.5. While these are helpful tools to provide structure and consistency in reviewing research articles, you still need to develop further understanding of the research process for them to be useful.

It is important to note that using a critical appraisal tool will not help you if you do not understand the fundamental principles of the research design of the study you are critiquing. It is therefore important to become familiar with the basic research methods of the research papers you have identified. If you do not understand the research methods used by the authors of the studies incorporated in your literature review, you will not be able to critique the studies with any confidence. (Aveyard, 2014, p.111)

In analyzing a research report, you must evaluate each step of the research process and ask whether the author's description of each step of the process meets the criteria. For instance, the

TABLE 3.4

QUESTIONS FOR COMPREHENSIVE READING

DESCRIPTIVE QUESTIONS	ANALYTICAL QUESTIONS
What is the question I want to address?	Does this article address my question in full or in part, and is it based on empirical research?
What is the quality of the source?	Is the journal and, therefore, the article credible? How do I know and on what am I basing my decision?
Who has written the article?	Is the author a subject expert or novice, and does this matter? How do I know if the author is credible?
In what type of setting has the study taken place?	Is it possible to transfer the findings from this study to my own setting? Is the setting equivalent to my own setting and does this matter? If so, why? If not, why not and can I articulate this?
What was the sample and how was it generated?	Who are the participants? Are they the same or similar to those I want to include? Does this matter and, if so, why? If not, why not and does this matter? Am I able to articulate my rationale for including this study in my review?
What study method was used?	Was the method appropriate and fit for purpose? Was it robust? How do I know? What am I basing my decision on? Has it helped me clarify what methods I want to use as part of my own study?
What were the findings?	Are the findings relevant to what took place in the study and are they relevant to my own needs? Do I understand what the researchers have deduced from the findings? Is it clear how the findings have been generated? Is there a clear trail outlining how, where, when and why the data have been managed?
How were the data analysed?	Were the correct statistical tests used as and where appropriate? How do I know and what am I basing my decision on? Were any themes generated from the qualitative data? How was this done? Is it clear that the ideas expressed came from the data? Was the process of analysing such data robust? How do I know this?
How have the researchers reported or discussed their findings?	Are the findings and modes of analysis transparent, or are they so brief I do not understand what took place? Can I trust the findings and how do I know this?
What are the conclusions?	How realistic and how appropriately derived are the study conclusions? Are they based on the data or do they appear tangential to the study? How applicable are they to my own setting or proposed study?
Is this article worth including and if so, in what context?	Will it form one of the themes of the literature review or will it only be worth using as background information?

Source: Wakefield, A. (2014). Searching and critiquing the research literature. *Nursing Standard, 28*(39), 49-57. https//:doi.10.7748/ns.28.39.49.e8867.

critiquing criteria for assessing a literature review, discussed in Chapter 5, include whether the literature review identifies gaps and inconsistencies in the literature about a subject, concept, or problem, and whether all of the concepts and variables are included in the review. These two criteria relate to critiquing the research question and the literature review. Remember when you are doing a critique you are pointing out strengths as well as weaknesses.

TABLE 3.5

QUALITY APPRAISAL CHECKLISTS

ORGANIZATION	TYPES OF APPRAISAL CHECKLISTS AVAILABLE
British Medical Journal (BMJ, 2021) https://bestpractice.bmj.com/info/toolkit/ebm-toolbox/critical-appraisal-checklists/	• Two-armed Randomized Controlled Trials • Multiple-Armed Randomized Controlled Trials • Diagnostic Test studies • Systematic Reviews
Centre for Evidence-Based Medicine, University of Oxford (2021) https://www.cebm.ox.ac.uk/resources/ebm-tools/critical-appraisal-tools	• Systematic Reviews • Diagnostics • Prognosis • Randomized Controlled Trials • Qualitative studies • Individual Patient Data Review
Critical Appraisal Skills Programme (CASP UK, 2020) https://casp-uk.net/casp-tools-checklists/	• Randomized Controlled Trials • Systematic reviews • Qualitative studies • Cohort study • Diagnostic study • Case Control study • Economic Evaluation • Clinical Prediction
Joanna Briggs Institute (JBI, n.d.) https://jbi.global/critical-appraisal-tools	• Analytical Cross Sectional studies • Case Control studies • Case Reports • Case Series • Cohort studies • Diagnostic Test Accuracy • Economic Evaluation • Prevalence studies • Qualitative Research • Quasi-Experimental studies • Randomized Controlled Studies • Systematic Reviews • Text and Opinion

Research Hint

If you still have difficulty understanding a research study after using the steps related to skimming and comprehensive reading, make another copy of your marked-up research article, include your specific questions or area of difficulty, and ask your professor to read this copy. Comprehensive understanding and synthesis are necessary for analyzing a research article. Understanding the author's purpose and methods for the study reflects critical thinking and facilitates evaluation of the study.

Another useful tool when analysing and critiquing research studies is to determine the level of evidence (Levin & Chang, 2014). Figure 3.1 depicts a model for determining the levels of evidence associated with the design of a study, ranging from systematic reviews of randomized clinical trials to expert opinions. This reflects the relative strength of different research designs to *testing hypotheses,* which is only one reason for doing research. As suggested in Chapter 2, different research methods provide different types of evidence and answer different research questions. Although the hierarchy suggests that evidence provided by qualitative studies ranks lower (i.e., levels V and VI), it is important to remember that qualitative methods are *not* used to test hypotheses and thus would not provide appropriate evidence for hypothesis-testing research or research about intervention effectiveness. However, qualitative research makes equally important contributions to the overall body of knowledge for nursing and

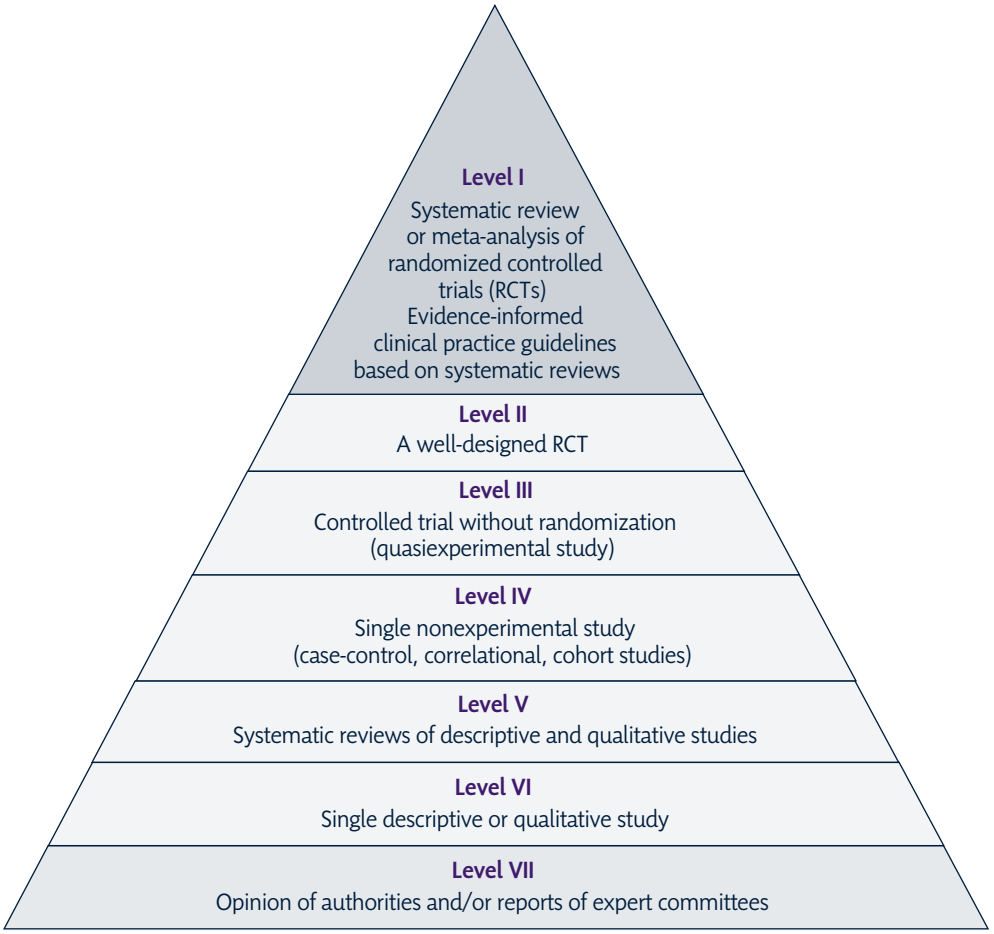


FIG. 3.1 Levels of evidence: hierarchy for rating levels of evidence, associated with a study's design. Evidence is assessed at a level according to its source.
(Based on Melynk, B. M., & Finoult-Overholt, E. (2011). *Evidence-based practice in nursing & literature: A guide to best practice* (2nd ed.). Philadelphia: Lippincott, Williams and Wilkins.)

health care and plays a different role than randomized controlled trials (i.e., level II). Qualitative research can be used to capture experiences of interventions, understand complex causal pathways, generate new theory, explain quantitative findings, or generate hypotheses (Moore et al., 2015).

The importance of an evidence rating system will become clearer to you as you read the chapters on quantitative research. For example, the Jackson and Dennis study (2017; Appendix B) is level II because it is a randomized controlled trial, whereas the study by Pesut et al. (2020; Appendix A) is level VI because of its qualitative design. The level of evidence, by itself, does not reveal the full worth of a study but is another tool that helps you think about the strengths and weaknesses of research designs for particular purposes and the nature of the evidence provided in the findings and conclusions.

Even though a study may be ranked as a level I or II study, it still needs thorough appraisal to determine its quality and credibility; even a randomized controlled trail can be biased (Djulbegovic & Guyatt, 2017). Assessing the strength of scientific evidence or potential research bias provides a vehicle to guide nurses in evaluating research studies for their applicability in clinical decision-making.

Synthesis of Understanding

After you have completed a critique of the article, compose a one-pag.8(a)0e,

comprehensive picture, or is a piece out of place? Are all the pieces connected logically together? Are there holes or gaps in the picture that is presented? In the case of reading several studies for synthesis, you need to consider how interrelated each of the studies are and determine the overall strength and quality of evidence and its applicability to practice. Reading for synthesis is essential in critiquing research studies. Box 3.3 summarizes tips for these levels of critical reading.

EVIDENCE-INFORMED PRACTICE AND RESEARCH

Along with gaining confidence while reading and critiquing research studies, you need to undertake a final step of reading and appraising the research literature: deciding how, when, and whether to apply a study or studies to your practice so that

APPRAISING THE EVIDENCE

Qualitative Research

Although general criteria for critiquing qualitative research are proposed in the following Critiquing Criteria box, each qualitative method has unique characteristics that influence what you may expect in the published research report, and journals often have page restrictions that penalize qualitative researchers because it can be difficult to fully explain the research process in a few pages. The criteria for critiquing are formatted to evaluate the selection of the phenomenon, the structure of the study, data gathering, data analysis, and description of the findings. Each question of the criteria focuses on factors discussed throughout the chapter. Appraising qualitative research is a useful activity for learning the nuances of this research approach. You are encouraged to identify a qualitative study of interest and apply the criteria for critiquing. Keep in mind that qualitative methods are the best way to examine questions that previously have not been addressed in research studies or to explain and explore phenomenon in a different way. The answers provided by qualitative data reflect important evidence that may provide the first insights into a patient population or clinical phenomenon.

CRITIQUING CRITERIA

Qualitative Approaches

IDENTIFYING THE PHENOMENON

1. Is the phenomenon clearly described and consistent throughout?

STRUCTURING THE STUDY

Research Question

2. Does the question specify a distinct process, phenomenon, or experience to be studied?
3. Does the question identify the context (participant group/place) that will be studied?
4. Does the choice of a specific qualitative method fit with the research question?

Researcher's Perspective

5. Is the researcher's relationship to participants described?
6. What is the role of the researcher as instrument?

Sample Selection

7. Is it clear that the selected sample is experiencing the phenomenon of interest?
8. Are there other people who would offer differing perspectives on the phenomenon?

DATA GATHERING

9. Are data sources and methods for gathering data specified?
10. Is there evidence that participant consent is an integral part of the data-gathering process?

DATA ANALYSIS

11. Can the dimensions of data analysis be identified and logically followed?
12. Is the participant's reality clearly described?
13. Is participant data (e.g., quotes) shown to substantiate the researcher's interpretation?
14. How did the research team make decisions and proceed through analysis?

DESCRIBING THE FINDINGS

15. Are examples provided to guide the reader from the raw data to the researcher's synthesis?
16. Does the researcher link the findings to existing theory or literature, or is a new theory generated?
17. Are the conclusions and implications for practice appropriate and justified based on the scope of the study?

In summary, the term *qualitative research* is an overriding description of multiple methods with distinct origins and procedures. In spite of distinctions, each method shares a common nature that guides data collection from the perspective of the participants to create a story that synthesizes disparate pieces of data into a comprehensible whole that provides evidence and promises direction for building nursing knowledge. An example of how three of these methods are used to study topics in palliative care is provided in Table 9.1. When reading these abstracts, consider how the research question fits the method that was selected, the different perspectives or worldviews that might be represented by the particular approach, and the different types of knowledge that each research project generated.

TABLE 9.1

EXAMPLES OF QUALITATIVE METHODS EXPLORING PALLIATIVE CARE

AUTHOR AND TITLE	METHOD	ABSTRACT
Kaasalainen et al. (2019)	Mixed methods with participatory action research	The goal of this study was to examine current rates of resident deaths, emergency department (ED) use within the last year of life, and hospital deaths for long-term care (LTC) residents. Using a mixed-methods approach, we compared these rates across four LTC homes in Ontario, Canada and explored potential explanations of variations across homes to stimulate staff reflections and improve performance based on a quality improvement approach. Chart audits revealed that 59% of residents across sites visited EDs during the last month of life and 26% of resident deaths occurred in hospital. Staff expressed surprise at the amount of hospital use during end of life (EOL). Reflections suggested that clinical expertise, comfort with EOL communication, clinical resources (i.e., equipment), and family availability for EOL decision making could all affect nondesirable hospital transfers at EOL. Staff appeared motivated to address these areas of practice following this reflective process.
Schick-Makaroff & Sawatzky (2020)	Interpretive description	The Edmonton Symptom Assessment System (Revised) (ESAS-r) contains nine questions pertaining to symptoms/well-being. It is a standardized patient-reported assessment instrument, but inconsistently used in palliative care. Thus, a problem exists in knowledge translation regarding routine use of the ESAS-r in palliative practice. The objective was to understand clinicians' perspectives on the use of the ESAS-r in palliative care in hospitals and at home. Qualitative focus groups (n = 14 with 46 clinicians) and interviews (n = 24) elicited views regarding use of the ESAS-r in palliative practice. Interpretive description was used as a general approach to this qualitative analysis focused on understanding clinicians' views. Palliative clinicians presented multiple perspectives of the ESAS-r pertaining to their (1) underlying values, (2) disparate purposes, and (3) incommensurate responses toward use in daily practice. Benefits and challenges supported diversity within these themes, highlighting divergence among perspectives and complexity of integrating a standardized tool in patient care. Integration of the ESAS-r in palliative care requires (1) educational support for developing competence; (2) consideration of clinicians' existing, heterogeneous beliefs regarding the use of standardized assessment instruments; and (3) consultation with multidisciplinary practitioners about optimal ways that ESAS-r results can be used in a person-centered approach to palliative care. (p.692)
Sutherland et al. (2017)	Critical ethnography	Evidence of gender differences in the amount and type of care provided by family caregivers in hospice palliative home care suggests potential inequities in health and health care experiences. As part of a larger critical ethnographic study examining gender relations among clients with cancer, their family caregivers and primary nurses, this article describes gendered expectations and exemptions for family caregivers within the sociopolitical context of end-of-life at home. Data were collected from in-depth interviews (n = 25), observations of agency home care visits (n = 9) and analyses of policy and home care agency documents (n = 12). Employing a critical feminist lens, a gender-based analysis revealed that structural discourses emphasizing an artificial divide between public and private spheres constructed end-of-life at home as private and apolitical. Associated with care of home and family, women were most impacted by these public/private discourses underpinning neoliberal values of cost-efficiency. Findings suggest that a critical perspective is needed to assist policy makers and healthcare providers to view how caregiver experiences are shaped by structures that control the availability of resources. Thus, instead of focusing on caregivers' deficits, interventions should be directed at the social, political and economic conditions that shape gendered experiences. (p.1)

CRITICAL THINKING CHALLENGES

- How does the researcher select a specific type of qualitative research method to answer the research question?
- Do findings from qualitative research studies need to be validated in subsequent studies?
- How can a nurse researcher select a qualitative research method when he or she is attempting to accumulate evidence regarding a new topic about which little is known?
- How can a focused ethnography approach to research be applied to evidence-informed practice?

CLINICAL JUDGEMENT QUESTIONS

1. What needs to be considered when engaging patients in research?
   a. All methods can engage patients throughout the entire research process
   b. Patient engagement can only occur with qualitative methods
   c. Patients routinely participate in qualitative research
   d. All methods use patients at a particular step in the research
2. Which method would be most appropriate to study processes?
   a. Participatory action research
   b. Grounded theory
   c. Phenomenology
   d. Institutional ethnography
3. What is a common critique of generic qualitative methods?
   a. They are too complicated for novice researchers
   b. They can lack the rigour found in other methods
   c. The interpretations stay close to the data
   d. There is no data saturation.

KEY POINTS

- Qualitative research is the investigation of human experiences in naturalistic settings, pursuing knowledge that informs theory, practice, instrument development, and further research.
- Qualitative research studies are guided by research questions.
- Data saturation occurs when the information being shared with the researcher becomes repetitive.
- Qualitative research methods include five basic elements: identifying the phenomenon, structuring the study, gathering the data, analyzing the data, and describing the findings.
- The phenomenological method is a process of learning and constructing the meaning of human experience through intensive dialogue with persons who are living the experience.
- The grounded theory method is an inductive approach that implements a systematic set of procedures to arrive at theory about basic social processes.
- The ethnographic method focuses on systematic descriptions of cultural groups.
- CBPR is a method that systematically accesses the voice of a community to plan context-appropriate action.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Agar, M. H. (1986). *Speaking of ethnography*. Beverly Hills, CA: Sage.

Boyle, J. S. (1991). Field research: A collaborative model for practice and research. In J. M. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue*. Newbury Park, CA: Sage.

Braun, V., & Clarke, V. (2014). What can "thematic analysis" offer health and wellbeing researchers? *International Journal of Qualitative Studies on Health and Wellbeing*, *9*(1), 26152. doi:10.3402/qhw.v9.26152.

Brookfield, S., Fitzgerald, L., Selvey, L., & Maher, L. (2019). The blind men and the elephant: Meta-ethnography 30 years on. *Qualitative Health Research*, *29*(11), 1674–1681. https://doi.org/10.1177/2F1049732319826061.

Caelli, K., Ray, L., & Mill, J. (2003). "Clear as mud": Toward greater clarity in generic qualitative research. *International Journal of Qualitative Methods*, *2*(2), 1–24.

Campbell, M., & Gregor, F. (2002). *Mapping social relations: A primer in doing institutional ethnography*. Aurora, ON: Garamond Press.

Campbell, M. L., & Rankin, J. M. (2017). Nurses and electronic health records in a Canadian hospital: Examining the social organisation and programmed use of digitised nursing knowledge. *Sociology of Health & Illness*, *39*(3), 365–379. https://doi.org/10.1111/1467-9566.12489.

Chambliss, D. F. (1996). *Beyond caring: Hospitals, nurses, and the social organization of ethics*. Chicago: University of Chicago Press.

Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509–535). Thousand Oaks, CA: Sage.

Charmaz, K. (2014). *Constructing grounded theory: A practical guide through qualitative analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Cohen, M. Z. (1987). A historical overview of the phenomenological movement. *Image*, *19*(1), 31–34.

Colaizzi, P. (1978). Psychological research as a phenomenologist views it. In R. S. Valle, & M. King (Eds.), *Existential phenomenological alternatives for psychology* (pp. 48–71). New York: Oxford University Press.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research*. Los Angeles: Sage.

Currie, G., & Szabo, J. (2019). "It would be much easier if we were just quiet and disappeared": Parents silenced in the experience of caring for children with rare diseases. *Health Expectations*, *22*(6), 1251–1259. https://doi.org/10.1111/hex.12958.

Denzin, N. K (1998). The practices and politics of interpretation. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 458–498). Thousand Oaks, CA: Sage.

Denzin, N. K., & Lincoln, Y. S. (1998). *The landscape of qualitative research*. Thousand Oaks, CA: Sage.

Dinmohammadi, M., Peyrovi, H., & Mehrdad, N. (2013). Concept analysis of professional socialization in nursing. *Nursing Forum*, *48*, 26–34. https://doi.org/10.1111/nuf.12006.

Duffy, M. (2011). Narrative inquiry: The method. In P. Munhall (Ed.), *Nursing research: A qualitative perspective* (5th ed., pp. 421–440). Mississauga, ON: Jones & Bartlett Learning.

El Hussein, M., Hirst, S., & Osuji, J. (2019). Professional socialization: A grounded theory of the clinical reasoning processes that RNs and LPNs use to recognize delirium. *Clinical Nursing Research*, *28*(3), 321–339. https://doi.org/10.1177/1054773817724961.

Ellis, C., Adams T. E., & Bochner, A. P. (2011). Autoethnography: An overview historical social research. *Historische Sozialforschung*, *36*(138), 273–290.

Estefan, A., Moules, N. J., & Laing, C. M. (2019). Composing sexuality in the midst of adolescent cancer. *Journal of Pediatric Oncology Nursing*, *36*(3), 191–206. https://doi.org/10.1177/1043454219836961.

Fetterman, D. M. (1998). Ethnography. In L. Bickman, & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 473–504). Thousand Oaks, CA: Sage Publications.

Fontana, J. S. (2004). A methodology for critical science in nursing. *Advances in Nursing Science*, *27*, 93–101.

Gadamer, H. G. (2004). *Truth and method* (2nd rev. ed.). (Weinsheimer, J., Marshall, D. G., Trans.). New York: Continuum.

Giorgi, A., Fischer, C. L., & Murray, E. L. (1975). *Duquesne studies in phenomenological psychology*. Pittsburgh, PA: Duquesne University Press.

Glaser, B., & Strauss, A. (1965). *Awareness of dying*. Chicago: Aldine de Gruyter.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glesne, C. (2011). *Becoming qualitative researchers: An introduction* (4th ed.). Toronto: Pearson.

Holkup, P. A., Tripp-Reimer, T., Salois, E. M., et al. (2004). Community-based participatory research: An approach to intervention research with a Native American community. *Advances in Nursing Science*, *27*, 162–175.

Jardine, D. W. (1992). The fecundity of the individual case: Considerations of the pedagogic heart of interpretive work. *Journal of Philosophy of Education*, *26*, 51–61. doi:10.1111/j.1467-9752.1992.tb00264.x.

Kahlke, R. M. (2014). Generic qualitative approaches: Pitfalls and benefits of methodological mixology. *International Journal of Qualitative Methods*, *13*, 37–52.

Kaasalainen, S., Sussman, T., Durepos, P., McCleary, L., Ploeg, J., & Thompson, G. (2019). What are staff perceptions about their current use of emergency departments for long-term care residents at end of life? *Clinical Nursing Research*, *28*(6), 692. https://doi.org/10.1177/1054773817749125.

LeCompte, M. D., & Schensul, J. J. (2010). *Designing and conducting ethnographic research: An introduction*. Lanham, MD: AltaMira Press.

Lewin, K. (1948). *Resolving social conflicts*. New York: Harper.

Lundy, K. S. (2011). Historical research. In P. Munhall (Ed.), *Nursing research: A qualitative perspective* (5th ed., pp. 381–397). Mississauga, ON: Jones & Bartlett Learning.

MacDonald, C., Martin-Misener, R., Steenbeek, A., et al. (2015). Honouring stories: Mi'kmaq women's experiences with Pap screening in Eastern Canada. *Canadian Journal of Nursing Research*, *47*(1), 72–96.

Macquire, P. (1996). Considering more feminine participatory research: What's congruency got to do with it? *Qualitative Inquiry*, *2*, 106–118.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation*. San Francisco: Jossey-Bass.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., ... & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *British Medical Journal*, *350*. https://doi.org/10.1136/bmj.h1258.

Morse, J. M., & Richards, L. (2002). *Readme first for a user's guide to qualitative methods*. Thousand Oaks, CA: Sage.

Moules, N. J. (2002). Hermeneutic inquiry: Paying heed to history and Hermes an ancestral, substantive, and methodological tale. *International Journal of Qualitative Methods*, *1*(3), 1–21. https://journals.sagepub.com/doi/full/10.1177/160940690200100301.

Patton, M. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. (2015). *Qualitative research & evaluation methods* (4th ed.). Thousand Oaks, CA: Sage.

Pauly, B., McCall, J., Browne, A. J., et al. (2015). Toward cultural safety: Nurse and patient perceptions of illicit substance use in a hospitalized setting. *Advances in Nursing Science*, *38*(2), 121–135.

Polit, D., & Beck, C. (2017). *Nursing research: Generating and assessing evidence for nursing practice* (10th ed.). Riverwoods, IL: Wolters Kluwer.

Roper, J., & Shapira, J. (2000). *Ethnography in nursing research*. Thousand Oaks, CA: Sage.

Sandelowski, M. (2000). What ever happened to qualitative description? *Research in Nursing and Health*, *23*, 334–340.

Sandelowski, M. (2010). Whats in a name?: Qualitative description revisited. *Research in Nursing and Health*, *1*, 77. https://0-doi-org.aupac.lib.athabascau.ca/10.1002/nur.20362.

Schick-Makaroff, K., & Sawatzky, R. (2020). Divergent perspectives on the use of the Edmonton Symptom Assessment System (Revised) in palliative care. *Journal of Hospice & Palliative Nursing*, *22*, 75–81. https://doi.org/10.1097/NJH.0000000000000617.

Smith, D. E. (2005). *Institutional ethnography: A sociology for people*. Lanham, MD: AltaMira Press.

Smith, V. (2001). Ethnographies of work and the work of ethnographers. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography* (pp. 220–233). Thousand Oaks, CA: Sage.

Spiegelberg, H. (1976). *The phenomenological movement* (Vols. I–II). The Hague: Martinus Nijhoff.

Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt, Rinehart, & Winston.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. New York: Cambridge University Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Stringer, E. T. (1999). *Action research* (2nd ed.). Thousand Oaks, CA: Sage.

Stringer, E. T., & Genat, W. J. (2004). *Action research in health*. Thousand Oaks, CA: Sage.

Sutherland, N., Ward, G. C., McWilliam, C., & Stajduhar, K. (2017). Structural impact on gendered expectations and exemptions for family caregivers in hospice palliative home care. *Nursing Inquiry*, *24*(1). n/a. https://doi.org/10.1111/nin.12157.

Thorne, S. (2018). What can qualitative studies offer in a world where evidence drives decisions? *Asia-Pacific Journal of Oncology Nursing*, *5*(1), 43–45. https://doi.org/10.4103/apjon.apjon_51_17.

Thorne, S. (2020). Beyond theming: Making qualitative studies matter. *Nursing Inquiry*, *27*, e12343. https://doi.org/10.1111/nin.12343.

Thorne, S., Reimer Kirkham, S., & MacDonald-Emes, J. (1997). Interpretive description: A noncategorical qualitative alternative for developing nursing knowledge. *Research in Nursing & Health*, *2*, 169–177.

van Kaam, A. (1969). *Existential foundations in psychology*. New York: Doubleday.

Wall, S. (2013). Nursing entrepreneurship: Motivators, strategies and possibilities for professional advancement and health system change. *Nursing Leadership*, *26*(2), 29–40. https://doi.org/10.7939/R3VF1G.

Wolcott, H. F. (1994). *Transforming qualitative data: description, analysis and interpretation*. Thousand Oaks, CA: Sage.

Wolcott, H. F. (1999). *Ethnography: A way of seeing*. Lanham, MD: AltaMira Press.

Wuest, J. (2011). Grounded theory: The method. In P. Munhall (Ed.), *Nursing research: A qualitative perspective* (5th ed., pp. 225–256). Mississauga, ON: Jones & Bartlett Learning.

Zahavi, D. (2019). Getting it quite wrong: Van Manen and Smith on phenomenology. *Qualitative Health Research*, *29*(6), 900–907.

Zahavi, D., & Martiny, K. M. (2019). Phenomenology in nursing studies: New perspectives. *International Journal of Nursing Studies*, *93*, 155–162.

PART **FOUR**

Quantitative Research

RESEARCH **VIGNETTE**

Why I Focus on Violence Against Women and Girls (VAWG)

Ruth Rodney
Assistant Professor
School of Nursing
York University
Toronto, Ontario

VAWG is a human rights and public health issue that affects one in three women globally (WHO, 2013). Unfortunately, these numbers are even higher in some regions of the world (Contreras-Urbina et al., 2019). Communities can react differently to dating violence because of long-standing discourses on gender, race, age, socioeconomic status, and culture that often intersect in complex ways. As a black woman, a daughter, a mother, a nurse, and a researcher, this topic is very personal to me. I have experienced violence in a past relationship and have assisted friends, acquaintances, and patients (in my role as a frontline nurse) who have experienced violence.

When reflecting on my own youth, I realized that dating violence was common and, to some degree, normalized. For many teenage girls and boys these experiences created an unhealthy starting point for dating and long-term relationships. I also noticed that adolescents faced greater challenges in accessing help when violence is experienced. Many girls I knew kept silent, and when others did seek help from formal social supports, they were often ignored. To prevent other youth from going through similar experiences I have had or witnessed others go through, I made a conscious decision to contribute to the body of knowledge that supports the elimination of violence against women and girls.

MY RESEARCH CONTRIBUTES TO THE FIELD OF VAWG

As a nurse researcher, I consider the knowledge I produce to be twofold: it informs policy development and implementation to better support violence prevention efforts, and it is also a form of advocacy, as my research is co-created with African Caribbean Black (ACB) communities who are not well represented in literature on dating violence.

Specifically, my doctoral thesis was the first study of its kind completed in Guyana to address adolescent dating violence with adolescents, parents, and teachers. It also contributes to the larger body of knowledge on adolescent dating violence, an area that is lacking within the Caribbean (Rodney, 2017), as it delves into the complexity of how community ideas of gender and respectability are shaped by historical understandings of relationships between men and women. My research also addresses how social systems (i.e., education, family) further complicate adolescent dating violence and provides a greater understanding of why previous violence prevention initiatives may not have been successful in Guyana (Rodney, 2017).

From a social justice perspective, I believe a fundamental component of health research is to ensure that communities who participate in research can access, understand, and utilize the information obtained. This means that my attention is focused not only on academic literature but also on providing accessible information to communities. For example, after recognizing the challenges teachers faced working with limited resources in focus group discussions, I organized a one-day experiential learning retreat in collaboration with the Ministry of Education in Guyana for teachers. The goal of this retreat was to provide teachers with an opportunity to learn new techniques and insights in decreasing negative interactions with students by developing a more positive classroom environment and providing opportunity for self-reflection.

I completed the qualitative component of the first national prevalence study on VAWG in Guyana, supported by UN Women (Contreras-Urbina et al., 2019). This study illustrates the regional difference of VAWG in a Caribbean country. The results highlight the importance of qualitative inquiry as it provides an in-depth understanding of why Guyana's rates of violence may be highest in the region, why women do not always disclose violence or seek help from formal supports, and how communities play a vital role in contributing to risk and protective factors of violence (Rodney & Bobbili, 2019). This research also supports the grassroots work of long-standing women's organizations that have been integral to the VAWG movement in Guyana.

In keeping with my commitment to knowledge translation for diverse communities, my colleague and I conducted three radio interviews about the results of this study with a local radio station, 94.1 BOOM FM. These interviews were recorded on Facebook Live and garnered over 8000 listeners in total. Audience members listened to key findings of the report and were provided with tips on how to continue discussions on VAWG within their own communities. Additionally, we also disseminated findings in a local newspaper to provide greater access to research findings (Bobbili & Rodney, 2019).

Furthermore, given the importance of the topic, the established trust and relationship with the community and the long-term commitment required to effect change for the elimination of VAWG, I continue to maintain a connection with communities in Guyana.

IMPLICATIONS FOR NURSING PRACTICE

Given that most nurses worldwide are women, the implications for nursing practice are significant. First, the global numbers of women's experiences of violence (i.e., one in three women) implies that nurses experience violence in their personal lives, while also supporting patients who experience violence. Acknowledging this reality means ensuring that nurses are provided with access to institutional social supports (Rodney & Bobbili, 2019). Advocating for these resources contributes to the overall safety of the public by supporting the mental and physical health of nurses (Rodney & Bobbili, 2019).

Additionally, identifying gaps in professional policies of VAWG (that may vary based on geographical area and healthcare system) can have greater implications for nursing practice and improve health outcomes for women and their families. Irrespective of country, healthcare plays a critical role in the multidisciplinary effort needed for violence prevention and treatment (UN Women, 2015). The research completed in Guyana illustrates that in the absence of directive policies for care, nurses' interaction with patients offered opportunities to provide violence prevention resources—even if women were not comfortable disclosing violence (Rodney & Bobbili, 2019). This means that nurses are ideally positioned to provide violence prevention resources, offer a safe space for women to disclose if they wish, and continue to contribute to the body of literature that examines VAWG and its elimination.

WHAT ARE MY AREAS OF FUTURE FOCUS?

I am focused on two main areas moving forward. Most of my work to date has been within Guyana. While the understanding of dating violence in Guyana is applicable to a Canadian context given the ACB communities that have migrated from the Caribbean, adolescent dating in a North American context is different. Therefore, I am focusing on how to improve dating relationships within a Canadian context for ACB youth. Secondly, while messages denouncing domestic and dating violence are widespread, communities are not necessarily receiving these messages in the way it was intended (Rodney, 2017; Rodney & Bobbili, 2019). Therefore, I am constantly questioning how to improve knowledge translation with the end goal of changing dominant discourses on violence in Guyana and Canada. ■

REFERENCES

Bobbili, S., & Rodney, R. (2019, November 25). Violence against women in Guyana: How do we make a change? In the Diaspora. Stabroek Newspaper. https://www.stabroeknews.com/2019/11/25/features/in-the-diaspora/violence-against-women-in-guyana-how-do-we-make-a-change/.

Contreras-Urbina, M., Bourassa, A., Myers, R., Ovince, J., Rodney, R., & Bobbili, S. (2019). Guyana Womens Health and Life Experiences Survey Report: UN Women. https://caribbean.unwomen.org/en/materials/publications/2019/11/guyana-womens-health-and-life-experiences-survey-report.

Rodney, R. (2017). Building healthier relationships within communities: The critical examination of guyanese perspectives on adolescent dating violence and its Prevention. ProQuest Dissertations & Theses Global. (1996230249). Retrieved from: http://ezproxy.library.yorku.ca/login?url=https://search-proquest-.

Rodney, R., & Bobbili, S. (2019). Women's health and life experiences: A qualitative research report on violence against women in Guyana: UN Women. https://caribbean.unwomen.org/en/materials/publications/2019/11/womens-health-and-life-experiences-a-qualitative-research-report-on-violence-against-women-in-guyana.

UN, Women (2015). A framework to underpin action to prevent violence against women: UN Women Library. Retrieved from https://www.unwomen.org/en/digital-library/publications/2015/11/prevention-framework.

World Health Organization (WHO). (2013). Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and nonpartner sexual violence: WHO Library. Retrieved from https://www.who.int/reproductivehealth/publications/violence/9789241564625/en/.

CHAPTER 10

Introduction to Quantitative Research

Ramesh Venkatesa Perumal | Mina D. Singh

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Define research design.
- Identify the purpose of the research design.
- Define control as it affects the research design.
- Compare and contrast the elements that affect control.
- Begin to evaluate the degree of control that should be exercised in the design.
- Define internal validity.
- Identify the threats to internal validity.
- Define external validity.
- Identify the conditions that affect external validity.
- Identify the links between study design and evidence-informed practice.
- Evaluate the design by using the critiquing questions.

KEY TERMS

accuracy
attrition
bias
constancy
control
control group
experimental group
external validity
extraneous variable
feasibility
Hawthorne effect
history threat
homogeneity
instrumentation threats
internal validity
maturation
measurement effects
mortality
objectivity
pilot study
randomization
reactivity
selection bias
selection effects
testing effect

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

THE WORD "DESIGN" IMPLIES THE ORGANIZATION of elements into a masterful work of art. In the world of art and fashion, the word conjures up images of processes and techniques that are used to express a total concept. When an individual creates something, process and form are employed. The form, process, and degree of adherence to structure depend on the aims of the creator.

The same can be said of the research process. The research process does not need to be a sterile procedure, but it should be one in which the researcher develops a masterful work within the limits of a problem and the related theoretical basis. The organization plan that the researcher creates is the design. When reading a study, the research consumer should be able to recognize that the research problem, purpose, literature review, theoretical framework, and hypothesis all interrelate with, complement, and assist in the operationalization of the design (Figure 10.1). The degree to which a fit exists between these design elements determines the strength of the study and of the consumer's confidence in the evidence provided by the findings and their potential applicability to practice.

Nursing practice is concerned with a variety of activities that require varying degrees of process and form, such as the provision of quality care, cost-effective patient care, responses of patients to disease, and factors that affect caregivers. When nurses administer patient care, they draw on the nursing process. Previous chapters stressed the importance of theory and knowledge of subject matter to research. How a researcher structures, implements, or designs a study affects the results of a research project.

To grasp the implications and the use of research, you need to understand the central issues in the design of a research project. This chapter provides an overview of the meaning, purpose, and issues related to quantitative research design. Chapters 10 and 11 discuss specific types of quantitative designs.

PURPOSE OF THE RESEARCH DESIGN

The purpose of the research design is to provide the plan for answering research questions. These questions can result in research driven by a researcher's curiosity or interest in a theoretical

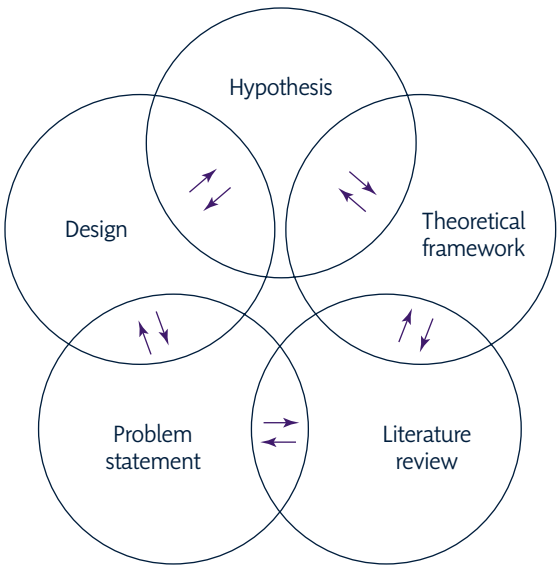


FIG. 10.1 Interrelationships of design, problem statement, literature review, theoretical framework, and hypothesis.

question. This process is called *basic research,* and its motivation is to expand nursing knowledge. In contrast, *applied research* is designed to solve clinical problems rather than to acquire knowledge for knowledge's sake; thus, the goal is to improve the patient's health care condition.

The design in quantitative research then becomes the vehicle for hypothesis testing or answering research questions, whether they are basic or applied. The design involves a *plan,* a *structure,* and a *strategy.* These three design concepts guide a researcher in writing the hypothesis or research questions, conducting the project, and analyzing and evaluating the data. The overall purpose of the research design is twofold: to aid in the solution of research problems and to maintain control (see Practical Application box). All research is an attempt to answer questions. The design, along with the methods and analysis, is the mechanism for finding solutions to research questions. **Control** is defined as the measures that the researcher uses to hold the conditions of the study uniform and avoid possible influence of **bias** (distortion of the results) on the dependent variable or outcome.

Practical Application

A research example that demonstrates how the design can aid in answering a research question and maintain control is the study by Henrique et al. (2018). The main purpose of their study was to test the effect of warm shower therapy and perineal exercises using a ball on pain, anxiety, and stress parameters during childbirth. To maintain control, the researchers had strict sample characteristics. Inclusion criteria were as follows: (1) older than 18 years of age, (2) single live fetus, (3) cephalic presentation, (4) no clinical or obstetric pathology, (5) cervical dilation between 3 and 8 cm, (6) gestational term, and (7) pain score ≥ 5. The exclusion criteria were as follows: (1) caesarean indication on admission, (2) analgesia, (3) smoker, (4) mental disorder, (5) caffeine intake equal to or less than 10 hours, (6) drug user, (7) less than 6 prenatal visits, and (8) without corticosteroids and analgesics in the previous 6 hours. By establishing the specific sample criteria and the participant's eligibility, the researchers were able to maintain control over the study's condition and answer the research question appropriately.

Various considerations, including the type of design, affect the accomplishment of the study. These considerations include **objectivity**—the use of facts without distortion by personal feelings or bias—in the conceptualization of the problem; accuracy; feasibility; control of the experiment; internal validity; and external validity. Statistical principles underlie the many forms of control, but it is more important that the research consumer have a clear conceptual understanding of statistics and how they inform the research questions.

The type of design used in a study also affects its application to practice. Chapters 11 and 12 present a number of experimental, quasiexperimental, and nonexperimental designs. The type of design used in a study is linked to the level of evidence, and, in turn, the contribution of a study's findings is linked to evidence-informed practice. As discussed in Chapter 1, the term *evidence-informed practice* is currently being used instead of *evidence-based practice* because it is more inclusive in that it encompasses many forms of evidence, such as clinical experience and judgement with research utilization. As you critically appraise the design, take into account other aspects of a study's design, which are reviewed in this chapter.

OBJECTIVITY IN THE CONCEPTUALIZATION OF THE PROBLEM

In the conceptualization of the problem, objectivity is derived from a review of the literature and development of a theoretical framework (see Figure 10.1). Using the literature, the researcher assesses the depth and breadth of available knowledge about the problem. The literature review and theoretical framework should show that the researcher reviewed the literature critically and objectively (see Chapters 2 and 5), because this conceptualization of the problem affects the type of design chosen. For example, for a question about the relationship of the length of a breastfeeding education program, either an experimental or a correlational design may be recommended (see

Chapters 10 and 11), whereas for a question regarding the physical changes in a woman's body during pregnancy and the maternal perception of the unborn child, a survey or correlation study may be advised (see Chapter 11). The literature review should reflect the following:

- When the problem was studied
- The aspects of the problem that were studied
- Where the problem was investigated
- By whom the problem was investigated
- The gaps or inconsistencies in the literature

Research Hint

A review that incorporates the aspects presented here allows the research consumer to judge the objectivity of the problem area and therefore whether the design chosen is suitable for investigating the problem.

ACCURACY

Accuracy in determining the appropriate design is also accomplished through the theoretical framework and review of the literature (see Chapters 2 and 5). **Accuracy** means that all aspects of a study systematically and logically follow from the research problem. The beginning researcher is wise to answer a question involving few variables that does not require the use of sophisticated designs. The simplicity of a research project does not render it useless or of a lesser value for practice. Although the project is simple, the researcher should not forgo accuracy. The research consumer should believe that the researcher chose a design that was consistent with the research problem and offered the maximum amount of control.

Many clinical problems have not yet been researched, so a preliminary, or pilot, study is a wise approach to testing the accuracy of a study design before a larger study is undertaken. A **pilot study** is a small, simple study conducted as a prelude to a larger study. The key is the accuracy, validity, and objectivity used by the researcher in attempting to answer the question. Accordingly, you should read various types of research reports and assess whether and how the criteria for each step of the research process were followed. Many nursing journals publish not only sophisticated clinical research projects but also smaller clinical studies whose results can be applied to practice.

FEASIBILITY

When you, as a consumer of research, critique the study design, you must also be aware of the pragmatic consideration of feasibility. Feasibility is the capability of the study to be successfully carried out. Sometimes, the reality of feasibility does not truly sink in until the researcher begins the study. When you review a study, you should consider feasibility, including availability of the participants, timing of the research, time required for the participants to take part in the study, costs, and analysis of the data (Table 10.1). Studies in which researchers are testing feasibility are also called *pilot studies* (see Practical Application box).

An example of a feasibility study is one conducted by Tryphonopoulos and Letourneau (2020). This feasibility study examined the effectiveness of the original protocol for a video feedback interaction guidance intervention designed to improve maternal–infant interaction quality, maternal depressive symptoms, and cortisol patterns of depressed mothers and their infants. The results of this study are used to support the implementation of a large-scale randomized controlled trial.

Practical Application

Santiago et al. (2019) conducted a pilot study to explore the feasibility of using a tablet equipped with a communication app in assisting with communication of patients in the intensive care unit. The results indicated that it was feasible for patients to use the tablet equipped with a communication app as an adjunct to communication.

Before a large experimental study (such as a randomized clinical trial) is conducted, it is helpful to first conduct a pilot study with a small number of participants to determine the feasibility of participant recruitment, the intervention, the

TABLE 10.1

PRAGMATIC CONSIDERATIONS IN DETERMINING THE FEASIBILITY OF A RESEARCH PROBLEM

FACTOR	PRAGMATIC CONSIDERATION
Time	The research problem must be able to be studied within a realistic period of time. All researchers have deadlines for completion of a project. The scope of the problem must be circumscribed enough to provide ample time for the completion of the entire project. Research studies generally take longer than anticipated to complete.
Participant availability	The researcher must determine whether a sufficient number of eligible participants will be available and willing to take part in the study. If a researcher has a "captive" audience (e.g., students in a classroom), it may be relatively easy to enlist their cooperation. When a study involves the participants' independent time and effort, they may be unwilling to participate when they will receive no apparent reward for doing so. Other potential participants may have fears about harm or confidentiality and be suspicious of the research process in general. Participants with unusual characteristics, such as rare diseases, are often difficult to locate. People are generally cooperative about taking part in a study, but a researcher must consider needing a larger participant pool than will actually participate. At times, when reading a research report, the researcher may note how the procedures were liberalized or the number of participants was altered—probably as a result of some unforeseen pragmatic consideration.
Facility and equipment availability	All research projects require some kind of equipment, such as questionnaires, telephones, stationery, stamps, technical equipment, or another apparatus. Most research projects also require the availability of a facility for the work, such as a hospital site for data collection, a laboratory space, or a computer centre for data analysis.
Money	Many research projects require some expenditure of money. Before embarking on a study, the researcher probably itemized the expenses and estimated the total cost of the project. This estimation of cost provides a clear picture of the budgetary needs for items such as books, stationery, postage, printing, technical equipment, telephone and computer charges, and salaries. These expenses can range from about $200 for a small-scale student project to hundreds of thousands of dollars for a large-scale federally funded project.
Researcher experience	The selection of the research problem should be based on the nurse's experience and interest. It is much easier to develop a research study related to a topic that is either theoretically or experientially familiar. Selecting a problem that is of interest to the researcher is essential for maintaining enthusiasm when the inevitable successes and failures occur.
Ethics	Research problems that place unethical demands on participants are not feasible for study. Researchers must take ethical considerations seriously. The consideration of ethics may affect the choice of the design and the methodology.

data-collection protocol, the likelihood that participants will complete the study, the reliability and validity of new measurement tools, and the costs of the study. These pragmatic considerations are not presented as a step in the research process, as are the theoretical framework and methods, but they do affect every step of the process and therefore should be considered when you assess a study. For example, the student researcher may or may not have funding or accessible services. When you critique a study, note the credentials of the author or authors and whether the investigation was part of either a student project or a fully funded grant project. If the project was a student

project, the standards of critiquing are applied more liberally than for projects conducted by an experienced researcher or clinician with a doctoral degree. Finally, the pragmatic issues raised affect the scope and breadth of an investigation and, therefore, its generalizability.

CONTROL

When developing a study, a researcher attempts to use a design to maximize the degree of control over the tested variables. Control involves holding the conditions of the study constant and establishing specific sampling criteria to reduce variability in the sample characteristics that may influence the outcome under investigation, as described by Hall et al. (2019) in a study of the feasibility of implementing a yoga intervention for participants with chronic pain. The intervention consisted of weekly 1-hour session for 10 weeks. The authors ensured homogeneity of the sample by using the inclusion and exclusion criteria. Adult patients (above the age of 18) suffering with chronic pain (with a recommendation from their therapist) and who could speak English were included in the study. Patients who have had some surgery in the past or scheduled in the future, who are pregnant, and who are enrolled in some other yoga course were excluded from the study.

An efficient design can maximize results, decrease errors, and control pre-existing conditions that may affect outcome. To accomplish these tasks, the research design and methods should demonstrate the researcher's efforts at control. For example, in a study to assess the effectiveness of Hernia Repair Education Intervention (HREI) for patients following inguinal hernia repair, the researchers had a specific inclusion and exclusion criteria to ensure homogeneity of the samples (Sawhney et al., 2017).

When research designs are critiqued, the issue of control is always raised, but with varying levels of flexibility. The issues discussed here will become clearer as you review the various types of designs.

Control is accomplished by ruling out extraneous variables that compete with the independent variables as an explanation for a study's outcome. An **extraneous variable** (also called a *mediating variable*) interferes with the operations of the phenomena being studied (e.g., age and gender). Means of controlling extraneous variables include the following:

- Use of a homogeneous sample
- Use of consistent data-collection procedures
- Manipulation of the independent variable
- Randomization

An investigator might be interested in how a new smoking cessation program (independent variable) affects smoking behaviour (dependent variable). The independent variable is assumed to affect the outcome, or dependent variable. An investigator needs to be relatively sure that the decrease in smoking is truly related to the smoking cessation program rather than to another variable, such as motivation.

The following example illustrates and defines these concepts further. In a study to assess the effectiveness of an electronic nursing intervention in improving mood and decreasing stress during first six months postpartum, the authors used a three-arm randomized controlled trial. To rule out the effects of extraneous variables on mood and stress, demographic information was collected, including if mothers had experienced depression or anxiety during or before pregnancy and if they were receiving treatment. This demographic information was then included in the statistical hypothesis testing to check if there were any differences between the groups (McCarter et al., 2019).

Although the design of the research study alone does not inherently provide control, an appropriately designed study with the necessary controls can increase an investigator's ability to answer a research question.

Evidence-Informed Practice Tip

As you read a report, assess whether the study includes a tested intervention and whether the report contains a clear description of the intervention and how

it was controlled. If the details are not clear, the intervention may have been administered differently among the participants, which would affect the interpretation of the results.

Homogeneous Sampling

In the example of smoking cessation, extraneous variables may affect the dependent variable. The characteristics of a study's participants are common extraneous variables. Age, gender, length of time smoked, amount smoked, and even smoking rules may affect the outcome in the smoking cessation example, even though they are extraneous or outside the study's design. As a control for these and other similar problems, the researcher's participants should demonstrate **homogeneity**, or similarity with regard to the extraneous variables relevant to the particular study (see Chapter 12). Extraneous variables are not fixed but must be reviewed, and their inclusion in the analyses is based on the study's purpose and theoretical base. By using a sample of homogeneous participants, the researcher has used a straightforward step of control.

In a randomized controlled trial, Lee et al. (2020) pilot tested an educational intervention to reduce fatigue among HIV clients. In order to have a homogenous sample and to reduce heterogeneity, the authors used a specific sampling criterion. The specific inclusion and exclusion criteria were: (a) diagnosed with HIV/AIDS, (b) at least 45 years of age, (c) unemployed, retired, or on disability, (d) with a phone or email for communicating with research staff, and (e) who stated they had experienced fatigue during the past week. Individuals with cognitive impairment, obesity, and sleep disorders were excluded from the study. In addition to the inclusion and exclusion criteria, the researchers included samples who had the potential confounding factors, such as adults 45 years and older who were unemployed or retired, as varying employment schedules. By this control step, authors limited the generalizability of the outcomes to other populations (see Chapter 17). The results can then be generalized only to a similar population of individuals. Homogeneity could be considered limiting, but not necessarily, because no treatment or program is applicable to all populations and educated consumers of research must take into consideration the differences in populations.

Research Hint

When reviewing studies, remember that it is better to have a "clean" study, whose results can be used to make generalizations about a specific population, than a "messy" study, whose results may be poorly or not at all generalizable.

If the researcher believes that one of the extraneous variables is important, it may be included in the design. In the smoking cessation example, if individuals are working in an area where smoking is not allowed and this condition is considered to be important to the study, the researcher could account for it in the design and set up a control condition for it. This condition can be established by comparing two different work areas: one where smoking is allowed and one where it is not. Of importance is that before the data are collected, the researcher should have identified, planned for, and controlled the important extraneous variables.

Constancy in Data Collection

Another basic but critical component of control is constancy in data-collection procedures. **Constancy** refers to the ability of the data-collection design to hold the conditions of the study to a cookbook-like recipe. In other words, for the purpose of collecting data for the study, each participant is exposed to the same environmental conditions, timing of data collection, data-collection instruments, and data-collection procedures (see Chapter 13).

An example of constancy in data collection is illustrated in the study by Boitor et al. (2019). The objective of this randomized controlled trial was to assess the effect of hand massage on the

pain intensity and pain-related interference with functioning of cardiac surgery patients. The interventions were standardized across patients and provided by one registered nurse trained in massage therapy at the same time of the day to patients. This type of control aided the investigators' ability to draw conclusions, discuss the findings, and cite the need for further research in this area. For the consumer, constancy demonstrates a clear, consistent, and specific means of data collection.

In a study to assess the effectiveness of video feedback intervention on improving the quality of interaction between mothers with postpartum depression and their infants, researchers used an intervention checklist to ensure accurate implementation of the intervention or implementation fidelity (Tryphonopoulos & Letourneau, 2020). All study designs should demonstrate constancy (fidelity) of data collection, but studies that test an intervention require the highest level of intervention fidelity.

Manipulation of the Independent Variable

A third means of control is manipulation of the independent variable. *Manipulation* refers to the administration of a program, treatment, or intervention to only one group within the study but not to the other participants in the study. The first group is known as the **experimental group**, and the other group is known as the **control group**, or comparison group. In a control group, the variables under study are held at a constant or comparison level. For example, in a randomized control trial, authors examined the effect of lanolin on nipple pain among breastfeeding women with damaged nipples. The experimental group received 40 gm tube of Lansinoh® along with a pamphlet. Participants were asked to apply lanolin over the damaged nipples following every feeding for 7 days. The mothers in the control group were not provided with the intervention (Jackson & Dennis, 2017).

Experimental and quasiexperimental designs involve manipulation, whereas in nonexperimental designs, the independent variable is not manipulated. This lack of manipulation does not decrease the usefulness of a nonexperimental design. The use of a control group in an experimental or quasiexperimental design is related to the research question and, again, its theoretical framework.

Blinding is a technique used in experimental and quasiexperimental research in which the participants are not aware of whether they are receiving the intervention. *Double blinding* is a technique in which both the researchers and the participants are not aware of who is receiving the intervention and who is in the control group. For example, Evans et al. (2018) conducted a double blind randomized controlled trial assessing the effectiveness of ginger aromatherapy in relieving chemotherapy-induced nausea in children with cancer, compared with a placebo. The participants were randomly assigned to one of the three groups, namely control group (no intervention), placebo intervention group, and the experimental group with ginger aromatherapy. In this study both the children and the nurse who collected the data were blinded to group assignment.

Research Hint

Be aware that the lack of manipulation of the independent variable does not mean that the study is weaker. The level of the problem, the amount of theoretical work, and the research that has preceded the project affect the researcher's choice of the design. If the problem is amenable to a design in which the independent variable can be manipulated, the power of a researcher to draw conclusions will increase, provided that all of the considerations of control are equally addressed.

Randomization

Researchers may also choose other forms of control, such as randomization. **Randomization** is a participant selection procedure in which each participant in a population has an equal chance of

being assigned to either the experimental group or the control group. Randomization eliminates bias, aids in the attainment of a representative sample, and can be used in various designs. In a randomized controlled trial to assess the effectiveness of Hernia Repair Education Intervention (HREI), the researchers assigned the eligibility participants to either the experimental group or the control group. Randomization occurred in a private office to minimize contamination (Sawhney et al., 2017). Randomization may be especially important in longitudinal studies, in which bias from giving the same instrument to the same participants on a number of occasions can be a problem (see Chapter 12).

QUANTITATIVE CONTROL AND FLEXIBILITY

The same level of control cannot be exercised in all types of designs. At times, when a researcher wants to explore an area in which little or no literature on the concept exists, the researcher will probably use an exploratory design. In this type of study, the researcher is interested in describing or categorizing a phenomenon in a group of individuals. Freeman et al. (2020) used an exploratory cross-sectional design to investigate palliative care nurse attitudes towards medical assistance in dying, as the scholarly literature on this topic was limited. The researchers concluded that the palliative care nurses have moderate attitude overall (based on the aggregated scores) towards MAID. In critiquing this type of study, the issue of control should be applied in a highly flexible manner because of the novelty of the nature of study.

If from a review of a study you determine that the researcher intended to conduct a correlational study (an examination of the relationship between or among the variables), then the issue of control takes on more importance. Control must be exercised as strictly as possible. At this intermediate level of design, it should be clear to the reviewer that the researcher considered the extraneous variables that may affect the outcomes.

All aspects of control are strictly applied to studies that use an experimental design. The reader should be able to locate in the research report how the researcher met these criteria: whether the conditions of the research were constant throughout the study, the assignment of participants was random, and experimental and control groups were used. Because of the control exercised in the study, the reader can determine that all issues related to control were considered and the extraneous variables were addressed.

Evidence-Informed Practice Tip

Remember that establishing evidence for practice is determined by assessing the validity of each step of the study, assessing whether the evidence assists in planning patient care, and assessing whether patients respond to the evidence-informed care.

INTERNAL AND EXTERNAL VALIDITY

Consumers of research must believe that the results of a study are valid, based on precision, and faithful to what the researcher wanted to measure. To form the basis of further research, practice, and theory development, a study must be credible and dependable. The two important criteria for evaluating the credibility and dependability of the results are internal validity and external validity. Threats to validity are listed in Box 10.1, and a discussion of each threat follows.

BOX 10.1

THREATS TO VALIDITY

INTERNAL VALIDITY

History threats
Maturation effects
Testing effects
Instrumentation threats
Mortality (attrition)
Selection bias

EXTERNAL VALIDITY

Selection effects
Reactive effects
Measurement effects

Internal Validity

Internal validity is the degree to which the experimental treatment, not an uncontrolled condition, resulted in the observed effects. To establish internal validity, the researcher rules out other factors or threats as rival explanations of the relationship between the variables. Threats to internal validity may be numerous and are considered by researchers in planning a study and by consumers before implementing the results in practice (Campbell & Stanley, 1996). Research consumers should note that the threats to internal validity are most clearly applicable to experimental designs, but attention to factors that can compromise outcomes should be considered to some degree in all quantitative designs. If these threats are not considered, they could negate the results of the research by affecting the design. Threats to internal validity include history threats, maturation effects, testing effects, instrumentation threats, mortality (attrition), and selection bias. Table 10.2 provides examples of these threats.

History Threats

Not only the independent variable but also another specific concurrent event may affect the dependent variable, either inside or outside the experimental setting. This threat to internal validity is referred to as the **history threat**. For example, in a study on the effects of a breastfeeding education program on the length of time of breastfeeding, government-sponsored breastfeeding promotions on television and in newspapers could affect the length of time of breastfeeding and would be considered a threat of history (see Table 10.2).

Maturation Effects

Maturation refers to the developmental, biological, or psychological processes that operate within an individual as a function of time; these processes are external to the events of the investigation. For example, suppose that a researcher wished to evaluate the effect of a specific teaching method on the achievements of baccalaureate students on a skills test. The investigator would record the students' abilities before and after the teaching method. Between the pretest and the posttest, the students would have grown older and wiser. The growth or change is unrelated to the investigation, and the differences between the two testing periods may be explained by such maturation rather than by the experimental treatment.

Maturation effects could also occur in a study of the relationship between two methods of teaching about children's knowledge of self-care measures. Posttests of student learning must be conducted relatively soon after the teaching sessions are completed. Such a short interval allows the investigator to conclude that the results were the outcome of the design of the study and not maturation in a population of children who are learning new skills rapidly. Maturation is more than change that results from an age-related developmental process; maturation can also be related to physical changes (see Table 10.2).

Testing Effects

Taking the same test repeatedly could influence participants' responses the next time the test is completed. For example, the effect on the participant's posttest score as the result of having taken a pretest is known as a **testing effect**. The effect of taking a pretest may sensitize an individual and improve the score on the posttest. Individuals generally score higher when they take a test a second time regardless of the treatment. The differences between posttest and pretest scores may be a result not of the independent variable but rather of the experience gained through the testing. For example, in one study the researchers used identical pretests and posttests based on a case scenario to assess the effect of simulation on Undergraduate Nursing Students' Knowledge of Nursing Ethics Principles (Donnelly et al., 2017). Whether the increase in scores of posttests regarding the knowledge of nursing ethics principles resulted from the teaching and learning strategies or was the effect

TABLE 10.2

EXAMPLES OF INTERNAL VALIDITY THREATS

THREAT	EXAMPLE
History threat	A study tested a teaching intervention in one hospital and compared outcomes to those of another hospital in which usual care was given. During the final months of data collection, the control hospital implemented a heart failure critical pathway; as a result, data from the control hospital (cohort) was not included in the analysis.
Maturation effect	If the change or difference in the outcome variable is due to the natural development process and not solely due to the effect of an independent variable that is considered as maturation effect.
Testing effect	A researcher wishes to measure acute pain with a repeated-measures design during a lengthy procedure. The researcher must consider the results in view of the possible bias of repeating the pain measurements over a short period of time. The measurements may prime the patients' responses, and the practice of reporting pain repeatedly on the same instrument during a procedure may influence the results. In a randomized controlled trial pilot study to assess the adolescent mothers perceptions of a phone-based peer support intervention designed to prevent postpartum depression, the researchers used a self-reported questionnaire to collect data. The researchers noted that though self-reported questionnaire has good psychometric properties, use of self-report is a possible limitation (Chyzzy et al., 2020).
Instrumentation threat	Boitor et al. (2019) discussed issues that possibly affected instrumentation such as reliance on patients recall of interference in the preceding 24 hours and the ability of patients in the critical units who are at risk of having disturbances in their circadian rhythm to recall the events correctly.
Mortality (attrition)	In a randomized controlled trial to test the effectiveness of psychotherapeutic intervention to relieve anxiety, the researchers noted that they had loss of samples (31% attrition) in the intervention group owing to lack of motivation or poor acceptability of the intervention. The authors also noted that similar attrition rates are found in studies where the length of therapy is specified at the outset (Sampaio et al., 2018).
Selection bias	In a study to assess the self-perceived palliative care competence of nurses and care aides working in settings that do not specialize in palliative care, the researchers reported that the sample may not be a representative sample. They also reported that there could be a self-selection bias among the samples favouring those who were interested or amenable to the integration of palliative care (Sawatzky et al., 2019).

of taking the test more than once was difficult to determine. Table 10.2 provides another example of a testing effect.

Instrumentation Threats

Instrumentation threats are changes in the variables or observational techniques that may account for changes in the obtained measurement. For example, a researcher may wish to study various types of thermometers (e.g., tympanic, digital, electronic, chemical indicator, plastic strip, and mercury) to compare the accuracy of the mercury thermometer with the other temperature-taking methods. To prevent instrumentation threats, the researcher must check the calibration of the thermometers according to the manufacturer's specifications before and after data collection.

Another example concerns techniques of observation or data collection. If a researcher has several raters collecting observational data, they all must be trained in a similar manner. If they are not similarly trained, or even if they are similarly trained but unable to conduct the study as planned, a lack of consistency may occur in their

ratings; therefore, a threat to internal validity will occur.

To avoid instrumentation threats and to maintain consistency in data collection, researchers in a study to validate the functional pain scale for hospitalized adults provided the initial training to the research assistants. After the initial training, the researcher and the research assistant assessed patients concurrently to ensure 90% interrater reliability before the research assistant was ready to collect data independently. The researchers also met with the research assistants once a month to assess if the research assistants needed additional training to maintain consistency in data collection (Arnstein et al., 2019). (For another example, see Table 10.2.) Although the researcher can take steps to prevent problems of instrumentation, the threat of instrumentation may still occur. When a research critiquer finds such a threat, it must be evaluated within the total context of the study.

Mortality/Attrition

Mortality or attrition is the loss of study participants from the first data-collection point (pretest) to the second data-collection point (posttest). If the participants who remain in the study are not similar to those who dropped out, the results could be affected. The loss of participants may be from the sample as a whole or, in a study that has both an experimental group and a control group, more of the participants may drop out from one group than from the other group; this effect is known as *differential loss of participants*. For example, in a study of the ways in which a media campaign affects the incidence of breastfeeding, if most dropouts were non–breastfeeding women, the perception given could be that exposure to the media campaign increased the number of breastfeeding women, whereas the effect of experimental attrition led to the observed results. See Table 10.2 for an example of a study in which mortality (attrition) may have influenced the results.

Selection Bias

If precautions are not used to gain a representative sample, selection bias—the threat to internal validity that arises when pretreatment differences exist between the experimental group and the control group—could result from the way the participants were chosen. Selection effects are a problem in studies in which the individuals themselves decide whether to participate in a study. Suppose an investigator wishes to assess whether a new smoking cessation program contributes to smoking cessation. If the new program is offered to all smokers, chances are that only individuals who are more motivated to stop smoking will take part in the program. Assessment of the effectiveness of the program is problematic because the investigator cannot know for certain whether the new program encouraged smoking cessation behaviours or whether only highly motivated individuals joined the program. To avoid selection bias, the researcher could randomly assign participants to either the new teaching method group or a control group that receives a different type of instruction. Table 10.2 provides another example of selection bias.

Research Hint

The list of threats to internal validity is not exhaustive. More than one threat can be found in a study, depending on the type of study design. Finding a threat to internal validity in a study does not invalidate the results and is usually acknowledged by the investigator in the "Results" or "Discussion" section of the study.

Evidence-Informed Practice Tip

Avoiding threats to internal validity in clinical research can be difficult. However, this reality does not render studies that have threats useless. It is important to take the threats into consideration and weigh the total evidence of a study for not only its statistical meaningfulness but also its clinical meaningfulness.

External Validity

External validity concerns the generalizability of an investigation's findings to additional populations and to other environmental conditions

and hence internal validity must be established prior to establishing external validity. To achieve external validity, variation in the conditions and the types of participants should lead to the same results. The goal of the researcher is to select a design that maximizes both internal and external validity, although attaining this goal is not always possible. If it is not possible, the researcher must attain the minimum criterion of external validity.

The factors that may affect external validity are related to the selection of participants, study conditions, and type of observations. These factors are termed *selection effects, reactive effects,* and *testing effects*. You may notice the similarity in the names of the factors of selection and testing and those of the threats to internal validity. When considering factors as internal threats, the reader assesses them as they relate to the *independent* and *dependent* variables within the study; when assessing them as external threats, the reader considers them in terms of the generalizability, or use outside the study with other populations and settings.

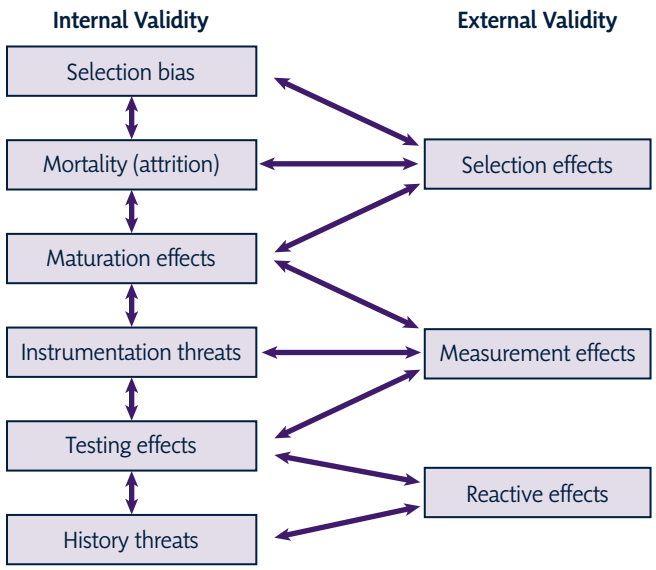
The Critical Thinking Decision Path displays the ways in which threats to internal and external validity can interact. This path is not, however, exhaustive with regard to the type of threats and their interaction. In comparison with problems of internal validity, generalizability issues are typically more difficult to deal with because they mean that the researcher is assuming that other populations are similar to the one being tested.

Selection Effects

Selection concerns the generalizability of the results to other populations. An example of **selection effects** is when the researcher cannot attain the ideal sample population. At times, the number of available participants may be low, or they may not be accessible to the researcher. The researcher may then need to choose a nonprobability method of sampling, not a probability method. Therefore, the type of sampling method used and how participants are assigned to research conditions will affect the generalizability

CRITICAL THINKING DECISION PATH

Potential Threats to a Study's Validity



of findings to other groups, or the external validity. In the following quotations, the authors have noted selection effects:

- "The study was limited from a sample size perspective (n = 12), due to limited resources and difficulties faced in getting the samples" (Tryphonopoulos & Letourneau, 2020)
- "The mothers who participated in this study are highly educated compared to the provincial average of education. This reduces the generalizability" (Letourneau et al., 2019)
- "The modest group size of each staff position limits overall generalizability, calls for replication and extension, and positions the present findings as preliminary" (Ezeobele et al., 2019).
- "The online survey design limited the participation to only those with Internet access, thus not representing potential participants without Internet connectivity. Also, the online survey relied on self-report and is therefore subject to self-report bias. While this study examined previous breastfeeding experience as a potential variable influencing breastfeeding self-efficacy, the survey strictly identified the length of previous breastfeeding experience. More meaningful results could have been obtained had the survey included questions examining the quality of previous breastfeeding experience." (Corby et al., 2019)

These remarks are cautionary, but they also point out the usefulness of the findings for practice and for future research aimed at building the data in these areas.

Reactive Effects

Reactivity is defined as the participants' responses to being studied. Participants may behave in a certain way with the investigator not because of the study procedures but merely as an independent response to being studied. This response is also known as the **Hawthorne effect**, named after Western Electric Corporation's Hawthorne plant, where a study of working conditions was conducted in the 1930s. The researchers developed several different working conditions, such as turning up the lights, piping in music loudly or softly, and changing work hours. They found that no matter what was done, the workers' productivity increased. They concluded that production increased as a result of the workers' knowing that they were being studied rather than because of the experimental conditions.

For example, in a randomized controlled study, Yang et al. (2019) tested the feasibility and acceptability of an online mindfulness intervention for pregnant women as an approach to reduce depressive and anxious symptoms. They also reported that the differences in the scores of the experimental group could also be due to the effect of increased attention and interaction of the experimental group members compared to the control group members with the researchers. The researchers recommend that the further research should use a properly matched control group.

Measurement Effects

Administration of a pretest in a study affects the generalizability of the findings to other populations; the resulting changes are known as **measurement effects**. Just as pretesting affects the posttest results within a study, pretesting affects the posttest results and generalizability outside the study. For example, suppose a researcher wants to conduct a study with the aim of changing attitudes toward acquired immune deficiency syndrome (AIDS). To accomplish this task, an education program on the risk factors for AIDS is incorporated. To test whether the education program changes attitudes toward AIDS, tests are given before and after the teaching intervention. The pretest on attitudes allows the participants to examine their attitudes regarding AIDS. The participants' responses on follow-up testing may differ from those of individuals who were given the education program and did not see the pretest. Therefore, when a study is conducted and a pretest is given, it may prime the participants and affect their subsequent answers, which in turn can affect the generalizability of the findings.

Research Hint

When you review a study, be aware of the internal and external threats to validity. These threats do not render a study useless; instead, they make it more useful to you. Recognition of the threats allows researchers to build on data and allows consumers to think through what part of the study can be applied to practice. Specific threats to validity depend on the type of design and generalizations that the researcher hopes to make.

Other threats to external validity depend on the type of design and methods of sampling used by the researcher but are beyond the scope of this text. Campbell and Stanley (1996) offered detailed coverage of the issues related to internal and external validity.

APPRAISING THE EVIDENCE

Quantitative Research

Critiquing the design of a study requires knowledge of the overall implications of a particular design for the study as a whole (see Critiquing Criteria box). Researchers want to consider the level of evidence provided by the design and how the study can be used to improve or change practice. Minimizing threats to internal and external validity enhances the strength of evidence for any quantitative design. The concept of the research design is all-inclusive and parallels the concept of the theoretical framework. The research design is similar to the theoretical framework in that it deals with a piece of the research study that affects the whole. This chapter has introduced the meaning, purpose, and important factors of design choice, as well as the vocabulary that accompanies these factors. Several criteria for evaluating the design can be drawn from this chapter. Remember that the criteria are applied differently with various designs. Differences in application do not mean that the research consumer will find a haphazard approach to design but rather that each design has particular criteria that allow the consumer to classify the design by type (e.g., experimental or nonexperimental). These criteria must be met and addressed in conducting an experiment. The particulars of specific designs are addressed in Chapters 10 and 11. The following discussion primarily pertains to the overall evaluation of a quantitative research design.

The research outcome should demonstrate that an objective review of the literature and the establishment of a theoretical framework guided the choice of the design. No explicit statement regarding these areas is made in a research article. A consumer can evaluate the design by critiquing the theoretical framework (see Chapter 2) and literature review (see Chapter 5). Is the question new and not extensively researched? Has a great deal of research been conducted on the question, or is the question a new or different way of looking at an old question? Depending on the level of the question, the investigators make certain choices. These choices enable researchers to look for differences in a controlled, comparative manner.

The research consumer should be alert for the methods that investigators use to maintain control (e.g., homogeneity in the sample, consistent data-collection procedures, manipulation of the independent variable, and randomization). As discussed in Chapter 10, all of these criteria must be met for an experimental design. As you begin to understand the types of designs (i.e., experimental, quasiexperimental, and nonexperimental designs, such as survey and relationship designs), you will find that control is applied in varying degrees or—as in the case of a survey study—the independent variable is not manipulated (see Chapter 11). The level of control and its applications presented in Chapters 10 and 11 provide the remaining knowledge for fully critiquing the aspects of a study's design.

Once you have established whether the necessary control or uniformity of conditions has been maintained, you must determine whether the study is believable or valid. You should ask whether the findings are the result of the variables tested—and thus internally valid—or whether another explanation is possible. To assess this aspect, you should review the threats to internal validity. If the investigator's study was systematic, was well grounded in theory, and followed the criteria for each of the processes, you will probably conclude that the study is internally valid.

APPRAISING THE EVIDENCE—*cont'd*

Quantitative Research

In addition, you must know whether a study has external validity or generalizability to other populations or environmental conditions. External validity can be claimed only after internal validity has been established. If the credibility of a study (internal validity) has not been established, a study has no generalizability to other populations (external validity). Determination of external validity is related directly to the sampling method (see Chapter 12). If the study is not representative of any one group or phenomenon of interest, external validity may be limited or not present. The establishment of internal and external validity requires not only knowledge of the threats to internal and external validity but also knowledge of the phenomena being studied, which allows critical judgements to be made about the linkage of theories and variables for testing. You should find that the design follows from the theoretical framework, literature review, research question, and hypotheses. You should believe, on the basis of clinical knowledge and knowledge of the research process, that the investigators are not, as the expression goes, comparing apples with oranges.

CRITIQUING CRITERIA

1. Is the type of study design employed appropriate?
2. Does the researcher use the various concepts of control that are consistent with the type of design chosen?
3. Does the design seem to reflect feasibility?
4. Does the design flow from the proposed research question, theoretical framework, literature review, and hypothesis?
5. What are the threats to internal validity?
6. What are the controls for the threats to internal validity?
7. What are the threats to external validity?
8. What are the controls for the threats to external validity?
9. Is the design appropriately linked to the levels of evidence hierarchy?

CRITICAL THINKING CHALLENGES

- Consider the following statement: "All research attempts to solve problems." How would you support or refute this statement?
- As a consumer of research, you recognize that control is an important concept in the issue of research design. You are critiquing an assigned experimental study as part of your "open-book" midterm examination. From what is written, you cannot determine how the researchers kept the conditions of the study constant. How does this characteristic affect the study's use in an evidence-informed practice model?
- Box 10.1 lists six major threats to the internal validity of an experimental study. Prioritize them and defend the one that you deem the essential, or number one, threat to address in a study.
- You are critiquing the research design of an assigned study as a consumer of research. How does the research design influence the findings of evidence in the study?
- How do threats to external validity contribute to the strength and quality of evidence provided by the findings of a research study?

CRITICAL JUDGEMENT QUESTIONS

1. How would an investigator ensure that the sample is homogenous?
   a. Restrict eligibility criteria to limit extraneous variables relevant to the study
   b. Randomly assign subjects to either the experimental or the control group
   c. Assign one research assistant to collect all data
   d. Collect all data at the same time of day

2. Which situation represents a threat to internal validity in an experimental study measuring the effect of an online post-op education for patients being discharged after coronary artery bypass graft surgery?
   a. Both men and women were included as subjects in the study
   b. Two new surgeons began performing the coronary artery bypass graft surgeries
   c. Patients in the experimental group gave the link to patients in the usual care control group
   d. Data collection took 1 year
3. COVID-19 outbreak in 2020 would represent what type of threat to internal validity in a longitudinal study that started on January 1, 2018 examining mortality rates due to respiratory infection?
   a. Maturation
   b. Instrumentation
   c. Selection bias
   d. History

KEY POINTS

- The purpose of the design is to provide the format of masterful and accurate research.
- Many types of designs exist. No matter which type of design the researcher uses, the purpose remains the same.
- The research consumer should be able to locate within the study a sense of the question that the researcher wished to answer. The question should be proposed with a plan or scheme for the accomplishment of the investigation. Depending on the question, the consumer should be able to recognize the steps taken by the investigator to ensure control.
- The choice of the specific design depends on the nature of the question. To specify the nature of the research question, the design must reflect the investigator's attempts to maintain objectivity, accuracy, pragmatic considerations, and, most important, control.
- Control affects not only the outcome of a study but also its future use. The design should also reflect how the investigator attempted to control threats to both internal and external validity.
- Internal validity must be established before external validity can be established. Both are considered within the sampling structure.
- No matter which design the researcher chooses, it should be evident to the reader that the choice was based on a thorough examination of the research question within a theoretical framework.
- The design, research question, literature review, theoretical framework, and hypothesis should all be interrelated.
- The choice of the design is affected by pragmatic issues. At times, two different designs may be equally valid for the same question.
- The choice of design affects the study's level of evidence.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Arnstein, P., Gentile, D., & Wilson, M. (2019). Validating the functional pain scale for hospitalized adults. *Pain Management Nursing, 20*(5), 418–424. https://doi.org/10.1016/j.pmn.2019.03.006.

Boitor, M., Martorella, G., Maheu, C., Laizner, A. M., & Gélinas, C. (2019). Does hand massage have sustained effects on pain intensity and pain-related interference in the cardiac surgery critically ill? A randomized controlled trial. *Pain Management Nursing, 20*(6), 572–579. https://doi.org/10.1016/j.pmn.2019.02.011.

Campbell, D., & Stanley, J. (1996). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.

Chyzzy, B., Nelson, L. E., Stinson, J., Vigod, S., & Dennis, C.-L. (2020). Adolescent mothers' perceptions of a mobile phone-based peer support intervention. *Canadian Journal of Nursing Research, 52*(2), 129–138. https://doi.org/10.1177/0844562120904591.

Corby, K., Kane, D., & Dayus, D. (2019). Investigating predictors of prenatal breastfeeding self-efficacy.

Canadian Journal of Nursing Research, 53(1), 56–63. https://doi.org/10.1177/0844562119888363.

Donnelly, M. B., Horsley, T. L., Adams, W. H., Gallagher, P., & Zibricky, C. D. (2017). Effect of simulation on undergraduate nursing students' knowledge of nursing ethics principles. *Canadian Journal of Nursing Research, 49*(4), 153–159. https://doi.org/10.1177/0844562117731975.

Evans, A., Malvar, J., Garretson, C., Pedroja Kolovos, E., & Baron Nelson, M. (2018). The use of aromatherapy to reduce chemotherapy-induced nausea in children with cancer: A randomized, double-blind, placebo-controlled trial. *Journal of Pediatric Oncology Nursing, 35*(6), 392–398. https://doi.org/10.1177/1043454218782133.

Ezeobele, I. E., McBride, R., Engstrom, A., & Lane, S. D. (2019). Aggression in acute inpatient psychiatric care: A survey of staff attitudes. *Canadian Journal of Nursing Research, 51*(3), 145–153. https://doi.org/10.1177/0844562118823591.

Freeman, L. A., Pfaff, K. A., Kopchek, L., & Liebman, J. (2020). Investigating palliative care nurse attitudes towards medical assistance in dying: An exploratory cross-sectional study. *Journal of Advanced Nursing, 76*(2), 535–545. https://doi.org/10.1111/jan.14252.

Hall, S. F., Wiering, B. A., Erickson, L. O., & Hanson, L. R. (2019). Feasibility trial of a 10-week adaptive yoga intervention developed for patients with chronic pain. *Pain Management Nursing, 20*(4), 316–322. https://doi.org/10.1016/j.pmn.2019.01.001.

Henrique, A. J., Gabrielloni, M. C., Rodney, P., & Barbieri, M. (2018). Non-pharmacological interventions during childbirth for pain relief, anxiety, and neuroendocrine stress parameters: A randomized controlled trial. *International Journal of Nursing Practice, 24*(3), 1–8. https://doi.org/10.1111/ijn.12642.

Jackson, K. T., & Dennis, C. L. (2017). Lanolin for the treatment of nipple pain in breastfeeding women: a randomized controlled trial. *Maternal and Child Nutrition, 13*(3), 1–10. https://doi.org/10.1111/mcn.12357.

Lee, K. A., Jong, S. S., & Gay, C. L. (2020). Fatigue management for adults living with HIV: A randomized controlled pilot study. *Research in Nursing and Health, 43*(1), 56–67. https://doi.org/10.1002/nur.21987.

Letourneau, N. L., de Koning, A. P. J., Sekhon, B., Ntanda, H. N., Kobor, M., Deane, A. J., Morin, A. M., Dewey, D., Campbell, T. S., & Giesbrecht, G. F. (2019). Parenting interacts with plasticity genes in predicting behavioral outcomes in preschoolers. *Canadian Journal of Nursing Research*. 084456211986361. https://doi.org/10.1177/0844562119863612.

McCarter, D. E., Demidenko, E., Sisco, T. S., & Hegel, M. T. (2019). Technology-assisted nursing for postpartum support: A randomized controlled trial. *Journal of Advanced Nursing, 75*(10), 2223–2235. https://doi.org/10.1111/jan.14114.

Sampaio, F. M. C., Araújo, O., Sequeira, C., Lluch Canut, M. T., & Martins, T. (2018). A randomized controlled trial of a nursing psychotherapeutic intervention for anxiety in adult psychiatric outpatients. *Journal of Advanced Nursing, 74*(5), 1114–1126. https://doi.org/10.1111/jan.13520.

Santiago, C., Roza, D., Porretta, K., & Smith, O. (2019). The use of tablet and communication app for patients with endotracheal or tracheostomy tubes in the medical surgical intensive care unit: A pilot, feasibility study. *Canadian Journal of Critical Care Nursing, 30*(1), 17–23.

Sawatzky, R., Roberts, D., Russell, L., Bitschy, A., Ho, S., Desbiens, J.-F., Chan, E. K. H., Tayler, C., & Stajduhar, K. (2019). Self-perceived competence of nurses and care aides providing a palliative approach in home, hospital, and residential care settings: A cross-sectional survey. *Canadian Journal of Nursing Research, 53*(1), 64–77. https://doi.org/10.1177/0844562119881043.

Sawhney, M., Watt-Watson, J., & McGillion, M. (2017). A pain education intervention for patients undergoing ambulatory inguinal hernia repair: A randomized controlled trial. *Canadian Journal of Nursing, 49*(3), 108–117. https://doi.org/10.1177/0844562117714704.

Tryphonopoulos, P. D., & Letourneau, N. (2020). Promising results From a video-feedback interaction guidance intervention for improving maternal–infant interaction quality of depressed mothers: A feasibility pilot study. *Canadian Journal of Nursing Research, 52*(2), 74–87. https://doi.org/10.1177/0844562119892769.

Yang, M., Jia, G., Sun, S., Ye, C., Zhang, R., & Yu, X. (2019). Effects of an online mindfulness intervention focusing on attention monitoring and acceptance in pregnant women: A randomized controlled trial. *Journal of Midwifery and Women's Health, 64*(1), 68–77. https://doi.org/10.1111/jmwh.12944.

CHAPTER 11

Experimental and Quasiexperimental Designs

Ramesh Venkatesa Perumal | Mina D. Singh

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- List the criteria necessary for inferring cause-and-effect relationships.
- Distinguish the differences between experimental and quasiexperimental designs.
- Define problems with internal validity that are associated with experimental and quasiexperimental designs.
- Describe the use of experimental and quasiexperimental designs for evaluation research.
- Critically evaluate the findings of selected studies in which cause-and-effect relationships were tested.
- Apply levels of evidence to experimental and quasiexperimental designs.
- Differentiate causation from association.

KEY TERMS

a priori
after-only design
after-only nonequivalent control group design
antecedent variable
attrition
control
dependent variable
evaluation research
experiment
experimental design
experimental group
formative evaluation
independent variable
intervening variable
manipulation
mortality
nonequivalent control group design
one-group pretest–posttest design
posttest–only control group design
pre-experimental design
quasiexperiment
quasiexperimental design
randomization
Solomon four-group design
summative evaluation
testing effects
time series design
true experiment

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

CHAPTER 10 PROVIDED AN OVERVIEW OF the meaning, purpose, and issues related to quantitative research design. This chapter provides a discussion of specific types of quantitative designs, inasmuch as choosing the correct design is crucial for hypothesis testing or answering research questions. The design involves a *plan*, a *structure*, and a *strategy*, which guide a researcher in writing the hypothesis or research questions, conducting the project, and analyzing and evaluating the data. Each design has specific characteristics to maintain control: for example, homogeneity in the sample, consistent data-collection procedures, manipulation of the independent variable, and randomization.

One of the fundamental purposes of scientific research in any profession is to determine cause-and-effect relationships. Nurses, for example, are concerned with developing effective approaches to maintaining and restoring wellness. Testing such nursing interventions to determine how well they actually work—that is, evaluating the outcomes in terms of efficacy and cost-effectiveness—is accomplished with the use of experimental and quasiexperimental designs. These designs differ from nonexperimental designs in one important way: The researcher actively seeks to bring about the desired effect and does not passively observe behaviours or actions. In other words, the researcher is interested not merely in observing customary patient care but in making something beneficial happen. Experimental and quasiexperimental studies are also important to consider in relation to evidence-informed practice because they provide level II and level III evidence. The findings of such studies provide the validation of clinical practice and the rationale for changing specific aspects of practice (see Chapter 20).

Experimental designs are particularly suitable for testing cause-and-effect relationships because they help eliminate potential alternative explanations (threats to validity) for the findings. Inferring causality requires that the following three criteria be met:

1. The causal variable and effect variable must be associated with each other.
2. The cause must precede the effect.
3. The relationship must not be explainable by another variable.

When you critique studies in which experimental and quasiexperimental designs were used, the primary focus is on the validity of the conclusion that the experimental treatment, or the **independent variable**, caused the desired effect on the outcome, or **dependent variable**. The validity of the conclusion depends on how well the researcher controlled the other variables that may explain the relationship studied. Thus, the focus of this chapter is to explain how various types of experimental and quasiexperimental designs control extraneous variables.

The purpose of this chapter is to acquaint you with the issues involved in interpreting studies that have an **experimental design** (characterized by three properties: randomization, control, and manipulation) or a **quasiexperimental design** (in which random assignment is not used, but the independent variable is manipulated, and certain mechanisms of control are used). Examples of these designs are listed in Box 11.1. The Critical Thinking Decision Path shows an algorithm that influences a researcher's choice of experimental or quasiexperimental design.

BOX 11.1

SUMMARY OF EXPERIMENTAL AND QUASI-EXPERIMENTAL RESEARCH DESIGNS

EXPERIMENTAL DESIGNS

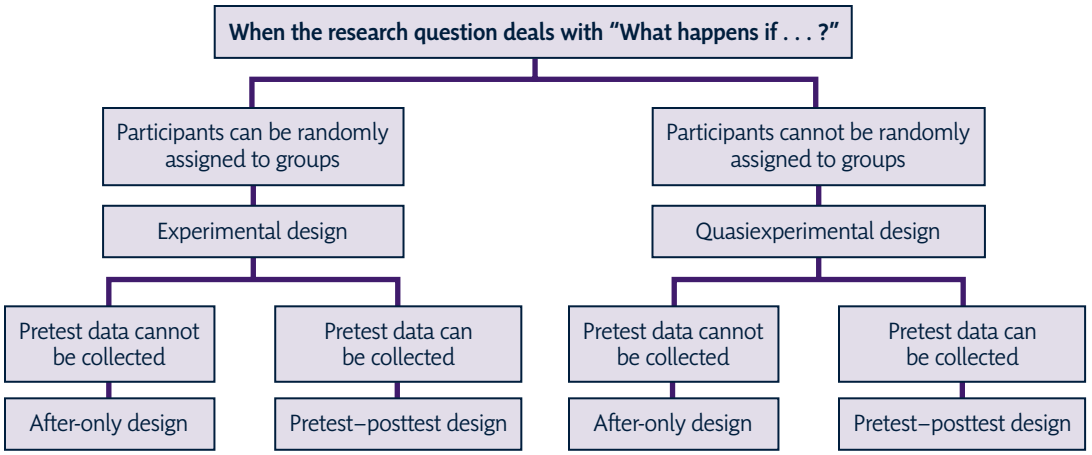
1. True experimental (pretest–posttest control group) design
2. Solomon four-group design
3. After-only design

QUASIEXPERIMENTAL DESIGNS

1. Nonequivalent control group design
2. After-only nonequivalent control group design
3. One-group pretest–posttest design
4. Time series design

CRITICAL THINKING DECISION PATH

Experimental and Quasiexperimental Designs



TRUE EXPERIMENTAL DESIGN

An **experiment** is a scientific investigation that makes observations and collects data according to explicit criteria. A **true experiment**—also known as a *pretest–posttest control group design* or *classic experiment*—has three identifying properties: randomization, control, and manipulation. These properties allow for other explanations of the phenomenon to be ruled out and thereby provide the strength of the design for testing cause-and-effect relationships.

A research study in which an experimental design is used is commonly called a *randomized control trial* (RCT). An RCT or experimental design is considered to be the best research design, "the gold standard," for providing information about cause-and-effect relationships. An individual RCT generates level II evidence because only minimal bias is introduced by this design. The higher level of evidence that a design produces, the more likely the results are to offer an unbiased estimate of the effect of an intervention and the more confident you can be that the intervention will be effective and produce the same results over and over again.

An example of an RCT is a study conducted by Sawhney et al. (2017) to assess the effectiveness of an individualized hernia repair education intervention (HREI) for patients following hernia repair. Patients undergoing ambulatory inguinal hernia repair were randomly assigned preoperatively either to the intervention group (those who received HREI) or to the control group (those received usual care). Authors concluded with positive findings on the effectiveness of HREI in reducing postoperative pain.

Randomization

Randomization, or random assignment to a group, is required for a study to be considered a true experimental design. It involves the assignment of participants to either the experimental or the control group on a purely random basis. In other words, each participant has an equal and known probability of being assigned to any group. Random assignment may be performed individually or by groups (for examples, see Tryphonopoulos & Letourneau, 2020; Anis et al., 2020; Page-cutrara & Turk, 2017). Random assignments to experimental or control groups

allows for the elimination of any systematic bias that may affect the dependent variable being studied. In randomization, it is assumed that any important intervening variable (a condition that occurs during the study that affects the dependent variable) will occur in an equal distribution between the groups. Randomization ensures an equal opportunity for participants to be included in the experimental group and decreases selection bias. Participants are randomly assigned to groups through several procedures, such as a table of random numbers or computer-generated number sequences. Whatever method is used, it is important that the process be truly random, that it be tamper-proof, and that the group assignment is concealed. Note that random assignment to groups is different from the random sampling discussed in Chapter 13.

Control

Control refers to the introduction of one or more constants into the experimental situation. Control is acquired by manipulating the causal or independent variable, randomly assigning participants to a group, carefully preparing experimental protocols, and using comparison groups. In experimental research, the comparison group is the control group, or the group that receives the usual treatment rather than the innovative, experimental treatment.

Manipulation

As discussed previously, experimental designs are characterized by the researcher "doing something" to at least some of the participants. The experimental treatment is administered to some participants in the study but not to others, or different amounts of it are administered to different groups. This difference in how the treatment is provided is the **manipulation** of the independent variable. The independent variable might be a treatment, a teaching plan, or a medication. The effect of this manipulation is measured to determine the result of the experimental treatment.

The concepts of control, randomization, and manipulation and their application to experimental design are sometimes confusing for students. These concepts allow researchers to have confidence in the causal inferences they make by allowing them to rule out other potential explanations.

Consider the use of control, randomization, and manipulation in the following example. In a randomized control trial, the researchers tested the effect of child health parent training programs on parent–child interaction quality and child development. In this pilot study, the researchers randomly assigned 20 families to an experimental or to a control group. The randomization was performed by a random assignment schedule created by an independent research staff (Anis et al., 2020).

The use of random assignment meant that all patients who met the study criteria had an equal and known chance of being assigned to the **control group** or the **experimental group**. The use of random assignment to groups helps ensure that the two study groups are comparable with regard to pre-existing factors that might affect the outcome of interest, such as violence among families, depression prevalence, and the nature of parent–child interaction. Also, the researchers in the above study (Anis et al., 2020) checked statistically whether the procedure of random assignment did, in fact, produce groups that were similar at baseline.

Evidence-Informed Practice Tip

In health care research, the term *randomized control trial (RCT)* often refers to a true experimental design. These designs are being used more frequently in nursing research, which is critical to evidence-informed practice initiatives.

The degree of control exerted over the experimental conditions in Tryphonopoulos and Letourneau's (2020) study is illustrated by its detailed description of the implementation of video feedback intervention. This control helped ensure that all members of the experimental group

received similar treatment and helps readers understand the process of experiment. The control group provided a comparison against which the experimental group could be judged.

In Tryphonopoulos and Letourneau's (2020) study, receiving the video feedback intervention was the manipulated treatment. The study aimed to assess the effectiveness of a video feedback intervention on improving the quality of interaction between depressed mothers and their infants. The results of the pilot feasibility study is positive and hence the authors could claim that video feedback intervention improved the quality of interaction between mothers with postpartum depression and their infants.

The use of the experimental design allows researchers to rule out many of the potential threats to internal validity of the findings, such as selection bias, history, and maturation effects (see Chapter 10). The strength of the true experimental design lies in its ability to help the researcher control the effects of any extraneous variables—alternative events that could explain the findings—that might constitute threats to internal validity. Such extraneous variables can be either *antecedent* or *intervening*.

The **antecedent variable** occurs before the study but may affect the dependent variable and confound the results. Factors such as age, gender, socioeconomic status, and health status might be important antecedent variables in nursing research because they may affect dependent variables, such as recovery time and ability to integrate health care behaviours. Antecedent variables that might have affected the dependent variables in the study by Tryphonopoulos and Letourneau (2020) include being on antidepressant or antipsychotic medication or having other therapies such as counselling. Random assignment to groups ensures that the groups are similar with regard to these variables so that differences in the dependent variable may be attributed to experimental treatment. The researchers also have compared both groups with regard to the above variables and reported in the study.

An **intervening variable** is a condition that occurs during the course of the study and is not part of the study; however, the intervening variable affects the dependent variable and can affect the study outcomes. An example of an intervening variable that might have affected the outcomes of Tryphonopoulos and Letourneau's (2020) study is the effect of antipsychotic medication. This can lead to changes in the primary outcome measured in the study.

Types of Experimental Designs

Several different experimental designs exist (Campbell & Stanley, 1966). Each is based on the classic design called the *true experiment*, diagrammed in Figure 11.1. Above the description diagram, symbolic notations are routinely used:

- R represents random assignment (for both the experimental group and the control group).
- O signifies observation through data collection on the dependent variable.
- O_1 signifies pretest data collection.
- O_2 represents posttest data collection.
- X represents exposure to the intervention.

Therefore, in Figure 11.1, note that the participants were assigned randomly (R) to the experimental or the control group. The experimental treatment (X) was given only to participants in the experimental group, and the pretests (O_1) and posttests (O_2) are the measurements of the dependent variables that were made before and after the experimental treatment was performed. In all true experimental designs, participants are randomly assigned to groups, an experimental treatment is introduced to some of the participants, and the effects of the treatment are observed. The variation in designs primarily concerns the number of observations that are made.

As shown in Figure 11.1, participants are randomly assigned to the two groups, experimental and control, so that antecedent variables are controlled. Next, pretest measurements or observations are made so that the researcher has a baseline for determining the effect of the

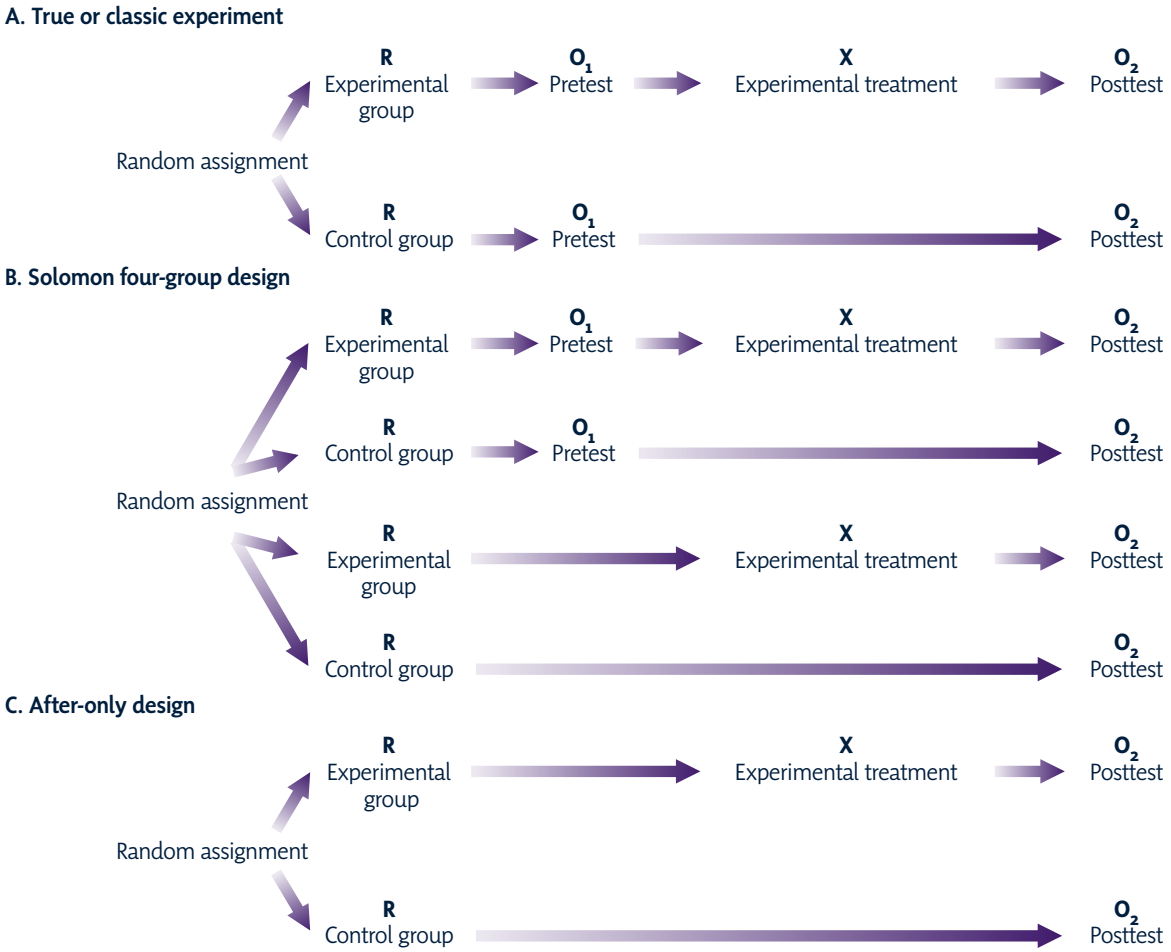


FIG. 11.1 Comparison of experimental designs.

independent variable. The researcher then introduces the experimental variable to one of the groups and measures the dependent variable again to see whether it has changed. The control group receives no experimental treatment, but the dependent variable in that group is also measured later for comparison with the experimental group. The degree of difference between the two groups at the end of the study indicates the confidence the researcher has that a causal link exists between the independent and dependent variables. Because random assignment and the control inherent in this design minimize the effects of many threats to internal validity, the true experimental design is a strong design for testing cause-and-effect relationships.

However, the design is not perfect. Some threats cannot be controlled in true experimental studies (see Chapter 10). People tend to drop out of studies that require their participation over an extended period. The influence over the outcome of an experiment of people dropping out or dying is commonly known as **attrition** or **mortality**. If the number or type of people who drop out of the experimental group differs from that of the control group, a mortality/attrition effect might explain the findings. When you read such a work, examine the sample and the results

carefully to see whether dropouts or deaths occurred.

Testing effects—the effects on the scores of a posttest as the result of having taken a pretest—also can be a problem in these studies because the researcher is usually administering the same test twice, and participants tend to score better the second time just by learning the test. Researchers can circumvent this problem in one of two ways: They might use different forms of the same test for the two measurements, or they might use a more complex experimental design called the Solomon four-group design.

The **Solomon four-group design**, shown in Figure 11.1, consists of two groups that are identical to those used in the classic experimental design plus two additional groups: an experimental after-group and a control after-group. As the diagram shows, all four groups have randomly assigned (*R*) participants, as in all experimental studies. However, the addition of these latter two groups helps rule out testing threats to internal validity that the before- and after-groups may experience. For example, suppose a researcher is interested in the effects of counselling on the self-esteem of patients with chronic illness. Just taking a test of self-esteem (O_1) may influence how the participants report themselves. The items might make the participants think more about how they view themselves so that the next time they fill out the questionnaire (O_2), their self-esteem might appear to have improved. In reality, however, their self-esteem may be the same as it was before; the scores are different only because the participants had previously taken the test. The use of this design with the two groups that do not receive the pretest allows for evaluating the effect of the pretest on the posttest in the first two groups. (See Practical Application box for another example of use of the Solomon four-group design.)

Although this design helps evaluate the effects of testing, the threat of mortality/attrition remains a problem, as in the classic experimental design.

Practical Application

Williams (2019) used the Solomon four-group design to test the effectiveness of high-fidelity simulation (HFS) as a teaching pedagogy for baccalaureate nursing students in a historically black college and university (HBCU). The author hypothesized that the students who received the focused respiratory assessment scenario through HFS would have higher scores in their posttest (measured through HESI). The study participants were randomly assigned to one of four groups.

1. Pretest, HFS intervention and posttest
2. No pretest, HFS intervention and posttest
3. Pretest, no intervention and posttest
4. No pretest, no intervention, only posttest

The study found that although there was an increase in the posttest scores of the group with the pretest, it was not statistically significant.

A less frequently used experimental design is the **after-only design**, shown in Figure 11.1. This design, which is sometimes called the **posttest–only control group design**, is composed of two randomly assigned groups (*R*), but in contrast to the true experimental design, neither group is given a pretest or other measures. Again, the independent variable is introduced to the experimental group (*X*) and not to the control group. The process of randomly assigning the participants to groups is assumed to be sufficient to ensure a lack of bias so that the researcher can still determine whether the treatment (*X*) created significant differences between the two groups (O_1 and O_2). This design is particularly useful when testing effects are expected to be a major problem and the number of available participants is too limited for a Solomon four-group design.

An example of this design would be a study of an intervention on postoperative pain management, inasmuch as pain cannot be measured before surgery and only an after-only design is required.

Research Hint

Remember that mortality/attrition is a problem in most experimental studies because data are usually collected more than once. The researcher should demonstrate that the groups are equivalent both when they enter the study and at the final analysis.

Field and Laboratory Experiments

Experiments also can be classified by setting. Field experiments and laboratory experiments share the properties of control, randomization, and manipulation and involve the same design characteristics but are conducted in different environments. Laboratory experiments take place in an artificial setting created specifically for the purpose of research. In the laboratory, the researcher has almost total control over the features of the environment, such as temperature, humidity, noise level, and participant conditions. Conversely, field experiments are exactly what the name implies: experiments that take place in a real, pre-existing social setting, such as a hospital or clinic, where the phenomenon of interest usually occurs.

Because most experiments in the nursing literature are field experiments and control is such an important element in the conduct of experiments, studies conducted in the field are subject to treatment contamination by factors specific to the setting that the researcher cannot control. However, studies conducted in the laboratory are by nature "artificial" because the setting is created for the purpose of research. Thus, laboratory experiments, although stronger with regard to internal validity questions than field studies are, have more problems with external validity. For example, a participant's behaviour in the laboratory may be quite different from the person's behaviour in the real world; this dichotomy presents problems in generalizing findings from the laboratory to the real world. Therefore, when you read research reports, you need to consider the possible effect of the experiment's setting on the findings of the study.

Consider a hypothetical study on different types of wound treatment gels and creams for the management of pressure ulcers. This study could be performed in a laboratory with animals, which would have allowed complete control over the external environment of the study—a variable that might be important in studying wound healing. However, researchers cannot guarantee that the results found in a study in a laboratory would be applicable to human patients in hospital settings; thus, some external validity would be lost.

Advantages and Disadvantages of the Experimental Design

As previously discussed, experimental designs are the most appropriate design for testing cause-and-effect relationships because the design enables the researcher to control the experimental situation. Therefore, experimental designs offer better corroboration than if the independent variable is manipulated in such a way that certain consequences can be expected. Such studies are important because one of nursing's major research priorities is documenting outcomes to provide a basis for changing or supporting current nursing practice.

Experimental designs are not commonly used in nursing research, for several reasons. First, experimentation is conducted under the assumption that all the relevant variables involved in a phenomenon have been identified. For many areas of nursing research, this is simply not the case, and descriptive studies need to be completed before experimental interventions can be applied. Second, these designs have some significant disadvantages. One problem with an experimental design is that many variables important in predicting outcomes of nursing care are not amenable to experimental manipulation. It is well known that health status varies with age and socioeconomic status. No matter how careful a researcher is, no one can assign participants randomly by age or a certain level of income. In addition, it may be technically possible to manipulate some variables, but their nature may preclude their actually manipulation.

For example, if a researcher tried to randomly assign groups to study the effects of cigarette smoking and asked the experimental group to smoke two packs of cigarettes a day, that researcher's ethics would be seriously questioned. It is also

potentially true that such a study would not work because nonsmokers randomly assigned to the smoking group would be unlikely to comply with the research task. Thus, sometimes even when a researcher plans to conduct a true experiment, participants dropping out of the study or other factors may, in effect, make the study a quasiexperiment.

Quasiexperimental designs are considered when it is not possible to randomly assign participants or when a control group is lacking. For example, Verkuyl et al. (2019) assessed the effects of three different debriefing methods (self-debrief only, self-debrief followed by a small group debriefing session, and self-debrief followed by a large group debrief) on improving the knowledge and debriefing experience after playing a gaming simulation. Randomly assigning students to control and experimental groups was not feasible; therefore, researchers randomly assigned the groups of students (based on lab sections) to the control group or to the intervention groups.

Another problem with experimental designs is that they may be difficult or impractical to perform in field settings. It may be quite difficult to randomly assign patients on a hospital floor to different groups when they might talk to each other about the different treatments. Experimental procedures also may be disruptive to the usual routine of the setting. If several nurses are involved in administering the experimental program, it may be impossible to ensure that the program is administered in the same way to each participant.

Because of these problems in carrying out true experiments, researchers frequently turn to another type of research design to evaluate cause-and-effect relationships. Such designs, because they seem experimental but lack some of the control of the true experimental design, are called *quasiexperiments*.

QUASIEXPERIMENTAL DESIGNS

Quasiexperimental designs are intended to test cause-and-effect relationships; however, in a quasiexperimental design, full experimental control is not possible. A **quasiexperiment** is a research design in which the researcher initiates an experimental treatment, but some characteristic of a true experiment is lacking. Control may not be possible because of the nature of the independent variable or the nature of the available participants. Quasiexperimental designs usually lack the element of true randomization, as described earlier with the Verkuyl et al. (2019) study. In other cases, the control group may be missing. However, like experiments, quasiexperiments involve the introduction of an experimental treatment.

In comparison with the true experimental design, quasiexperimental designs are used similarly. Both types of designs are used when the researcher is interested in testing cause-and-effect relationships. However, the basic problem with the quasiexperimental approach is a weakened confidence in making causal assertions. Because of the lack of some controls in the research situation, quasiexperimental designs are subject to contamination by many, if not all, of the threats to internal validity discussed in Chapter 10.

Types of Quasiexperimental Designs

Many different quasiexperimental designs exist. Only the ones most commonly used in nursing research are discussed in this book. To illustrate, the symbols and notations introduced earlier in the chapter are used. Refer to the true experimental design shown in Figure 11.1 and compare it with the **nonequivalent control group design** shown in Figure 11.2. Note that the latter design looks exactly like that of the true experiment except that participants are not randomly assigned to groups.

For example, suppose a researcher is interested in the effects of a new diabetes education program on the physical and psychosocial outcomes of patients with newly diagnosed diabetes. If the conditions were right, the researcher might be able to randomly assign participants to either the group receiving the new program or the group receiving the usual program, but for any number of reasons,

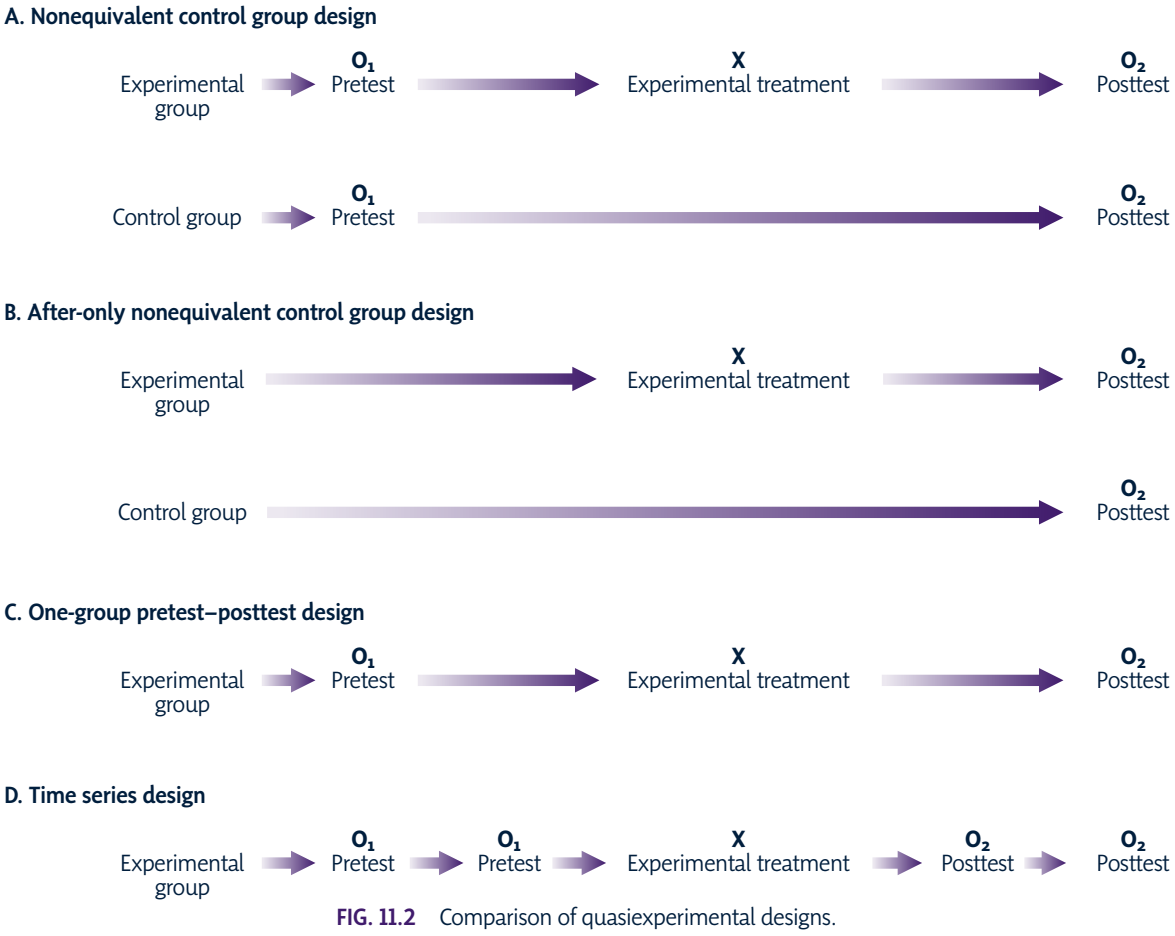


FIG. 11.2 Comparison of quasiexperimental designs.

that design might not be possible (e.g., nurses on the unit where patients are admitted might be so excited about the new program that they cannot help but include the new information for all patients). Thus, the researcher has two choices: to abandon the experiment or to conduct a quasiexperiment. To conduct a quasiexperiment, the researcher might find a similar unit where the new program has not been introduced and study the patients with newly diagnosed diabetes who are admitted to that unit as a comparison group. The study would then involve the quasiexperimental type of design because it lacks randomization.

Studies in which both quantitative and qualitative methods are used are called *mixed-methods studies*.

Research Hint

Remember that researchers often make trade-offs and sometimes use a quasiexperimental design instead of an experimental design because it may be pragmatically impossible to randomly assign participants to groups. The fact that the design is not "pure" does not decrease the value of the study, although the utility of the findings may be decreased.

The nonequivalent control group design is commonly used in nursing research studies conducted in field settings. The basic problem with the design is the weakening of the researcher's confidence in assuming that the experimental and comparison groups are similar at the beginning of the study. Threats to internal validity, such as selection bias, maturation effects, testing effects,

and mortality (attrition), are possible with this design. However, the design is relatively strong because the gathering of the data at the time of the pretest allows the researcher to compare the equivalence of the two groups on important antecedent variables before the independent variable is introduced.

An example of the nonequivalent control group design is that of Sezgin and Esin (2018). The purpose of this project was to evaluate the effects of PRECEDE-PROCEED model of ergonomic risk management program in reducing musculoskeletal symptoms of ICU nurses. The Ergonomic Risk Management Program (ERMP) intervention was the health promotion program based on PRECEDE-PROCEED model that included video training, personal interviews, provision of exercise mats, and regular text messages for six months. A total of 30 ICU nurses in the intervention group and 31 ICU nurses in control group were recruited for the study. The nurses were chosen randomly and not through randomization. As a nonequivalent control group design, the pretest done in the study was a baseline measurement of musculoskeletal symptoms and risks assessments (environment, predisposing factors, and upper rapid upper risk assessment) through surveys. These measurements were then compared with posttest measurements in order to determine the effectiveness of the intervention.

An example of the nonequivalent control group design is that of Kim and Kim (2020). The purpose of this study was to evaluate the effectiveness of electrocardiography training program using team-based learning for early-stage nurses in intensive care units. Thirty-six participants in the intervention group and 29 participants in the control group were recruited for the study. The participants were chosen considering their convenience to participate in the study. The intervention (Team Based Learning) was implemented in three stages over a 5-week period on selected topics whereas the control group received only a lecture on the same topics. As a nonequivalent control group design, the pretest done in the study was assessment of the knowledge of EKG and reading ability of bedside EKG monitoring and 12-lead EKG. These assessments were then compared with the posttest measurements in order to determine the effectiveness of the intervention.

Suppose that the researcher did not measure the participants' responses before the introduction of the new treatment (or the researcher was hired after the new program began) but later decided that data demonstrating the effect of the program would be useful. Perhaps, for example, a third party asks for such data to determine whether it should pay the extra cost of the new teaching program. Sometimes the outcomes simply cannot be measured before the intervention, as with prenatal interventions that are expected to affect birth outcomes. The study that could be conducted would have an **after-only nonequivalent control group design**, illustrated in Figure 11.2. This design is similar to the after-only experimental design, but randomization is not used to assign participants to groups. In this design, the two groups are assumed to be equivalent and comparable before the introduction of the independent variable (X). Thus, the soundness of the design and the confidence that the researchers can have in the findings depend on the soundness of this assumption of preintervention comparability. Often, the assumption that the two nonrandomly assigned groups are comparable at the outset of the study is difficult to assert because the validity of the statement cannot be assessed.

In the example of the teaching program for patients with newly diagnosed diabetes, measuring the participants' motivation after the teaching program would not reveal whether their motivations differed before they received the program, and it is possible that the teaching program would motivate individuals to learn more about their health problem. Therefore, the researcher's conclusion that the teaching program improved physical status and psychosocial outcome would be subject to the alternative conclusion that the results were

an effect of pre-existing motivations (selection effect) in combination with greater learning by participants so motivated (selection–maturation interaction). Nonetheless, this design is frequently used in nursing research because opportunities for data collection are often limited and because this design is particularly useful when testing effects may be problematic.

An approach used by researchers when only one group is available is to study that group over a longer period—that is, to test participants before an intervention and again afterwards. This quasiexperimental design is called a **time series design** and is illustrated in Figure 11.2. Time series designs are useful for determining trends as in a study on the effectiveness of a concussion workshop among the youth of First Nations community, their parents, and coaches in improving their knowledge and attitude (Hunt et al., 2018). The participants completed the same survey at three points (before intervention, immediately post workshop, and six months after the workshop). The survey had questions that measured the knowledge and attitude related to concussion. Sometimes data are collected many times before the introduction of the treatment to establish a baseline point of reference on outcomes. The experimental treatment is then introduced, and data are collected multiple times afterwards to determine a change from baseline. The broad range and number of data-collection points help rule out alternative explanations, such as history effects. However, a testing threat to internal validity is ever present because of multiple data-collection points, and without a control group, the threats of selection bias and maturation effects cannot be ruled out (see Chapter 10).

To rule out some alternative explanations for the findings of a one-group pretest–posttest design, researchers typically measure the phenomenon of interest over a longer period and introduce the experimental treatment sometime during the course of the data-collection period (see Figure 11.2). Even with the absence of a control group, the broader range of data-collection points helps rule out threats to validity such as history threats. Obviously, the earlier example of teaching patients with diabetes does not lend itself to this design because researchers do not have access to the patients before the diagnosis.

Research Hint

One of the reasons replication is so important in nursing research is that many problems cannot be subjected to experimental methods. Therefore, the consistency of findings across many patient populations helps support a cause-and-effect relationship even when an experiment cannot be conducted.

Advantages and Disadvantages of Quasiexperimental Designs

Because of the problems inherent in interpreting the results of studies with quasiexperimental designs, you may wonder why anyone would use them. Quasiexperimental designs are used frequently because they are practical and feasible, and the results are generalizable. These designs are more adaptable to the real-world practice setting than controlled experimental designs. In addition, for some hypotheses, these designs may be the only way to evaluate the effect of the independent variable of interest.

The weaknesses of the quasiexperimental approach involve mainly the inability to establish clear cause-and-effect relationships. However, if the researcher can rule out any plausible alternative explanations for the findings, such studies can lead to increased knowledge about causal relationships. Researchers have several options for ruling out these alternative explanations. They may control extraneous variables **a priori** (before initiating the intervention) by design.

Researchers can also use methods to control extraneous variables statistically. In some cases, common-sense knowledge of the problem and the population can suggest that a particular explanation is not plausible. Nonetheless, replicating such studies is important to support the causal assertions developed through the use of quasiexperimental designs.

The literature on cigarette smoking is an excellent example of how findings from many studies, experimental and quasiexperimental, can be linked to establish a causal relationship. A large number of well-controlled experiments with laboratory animals randomly assigned to smoking and nonsmoking conditions have documented that lung disease does develop in "smoking" animals. Although such evidence is suggestive of a link between smoking and lung disease in humans, it is not directly transferable because animals and humans are different. Because humans cannot be randomly assigned to smoking and nonsmoking groups, for ethical and other reasons, researchers interested in this problem must use quasiexperimental data to test their hypotheses about smoking and lung disease.

Several different quasiexperimental designs have been used to study this problem, and all have yielded similar results: A causal relationship does exist between cigarette smoking and lung disease. Note that the combination of results from both experimental and quasiexperimental studies led to the conclusion that smoking causes lung disease, because the studies together meet the causal criteria of relationship, timing, and lack of an alternative explanation.

The tobacco industry has argued that because the studies on humans are not true experiments, another explanation is possible for the relationships that have been found. For example, these relationships suggest that the tendency to smoke is linked to the tendency for lung disease to develop, and smoking is merely an unimportant intervening variable. The reader needs to review the evidence from studies to determine whether the cause-and-effect relationship postulated is believable.

Evidence-Informed Practice Tip

Findings from studies with experimental designs are considered level II evidence, and those from studies with quasiexperimental designs are considered level III evidence. Quasiexperimental designs are lower on the hierarchy of evidence because of a lack of a research control, which limits the ability to establish confident cause-and-effect statements that influence clinical decision-making.

PRE-EXPERIMENTAL DESIGNS

Pre-experimental designs follow similar experimental steps but do not include a control or comparison group. There is only a single group, with no comparison with an equivalent or nonequivalent nontreatment group. Examples are the one-group pretest–posttest ($O_1 \rightarrow X \rightarrow O_2$) and the one-group posttest-only ($X \rightarrow O_1$) designs, where X is the treatment or intervention and O is the data-collection points.

In the one-group pretest–posttest design, data are collected before and after an experimental treatment on this one group of participants. In this type of design, the participants act as their own controls, and no randomization occurs. Because controls and randomization are important characteristics that enhance the internal validity of the study, the evidence generated by the findings of this type of pre-experimental design needs to be interpreted with careful consideration of the design limitations.

The advantage of these designs is that they can be used to evaluate treatments, ruling out ineffective treatments before large-scale experimental or quasiexperimental studies are initiated. The disadvantage of this design is that without a control or comparison group, it is difficult to make any conclusions as to whether the treatment, (X), really caused the outcomes or changes.

Practical Application

Hickin et al. (2017) conducted a one-group pretest–posttest design study to check if the focused educational intervention for critical care nurses is effective in increasing long-term delirium knowledge and improving perception and rates of routine delirium screening in the ICU. The ICU nurses were provided with PowerPoint presentations on various aspects of delirium and assessments tools during their annual ICU refresher. Each participant acted as his own control by being tested thrice: before the educational intervention and then 3 and 18 months after the interventions. Findings revealed that the educational intervention led to an increase in the knowledge and the frequency of delirium assessment of patients in ICU.

EVALUATION RESEARCH AND EXPERIMENTATION

As the science of nursing expands and the cost of health care rises, nurses and other health care providers have become increasingly concerned with the ability to document the costs and the benefits of nursing care (see Chapter 1). This task is a complex process, but at its heart is the ability to evaluate or measure the outcomes of nursing care to inform health care decision-making. Such studies usually are associated with quality assurance, quality improvement, and evaluation. Studies of evaluation or quality assurance do exactly what the name implies: They are concerned with the determination of the quality of nursing and health care and with the assurance that the public is receiving high-quality care.

Quality assurance and quality improvement in nursing are current and important topics for nursing care. Many early studies of quality assurance documented whether nursing care met predetermined standards. The goal of quality improvement studies is to evaluate the effectiveness of nursing interventions and to provide direction for further improvement in the achievement of quality clinical outcomes and cost-effectiveness.

Evaluation research is the use of scientific research methods and procedures to evaluate a program, treatment, practice, or policy. In evaluation research, analytical means are used to document the worth of an activity such as an intervention, but such research is not a different design. Both experimental and quasiexperimental designs (as well as nonexperimental designs) are used to determine the effect or outcomes of a program. When these designs a reused in evaluating a program, the term *evaluation research* is used. Bigman (1961) listed the following purposes and uses of evaluation research:

1. To discover whether and how well the objectives are being fulfilled
2. To determine the reasons for specific successes and failures
3. To direct the course of the experiment with techniques for its effectiveness
4. To reveal principles that underlie a successful program
5. To base further research on the reasons for the relative success of alternative techniques
6. To redefine the means to be used for attaining objectives and to redefine subgoals in view of research findings

According to Clarke (2001), the following four levels of evaluation research are being highlighted in health care, especially nursing research: evaluation of the effectiveness of clinical interventions; evaluation of the effect of new ways of delivering health care; evaluation of structured programs aimed at specific patient groups; and evaluation of the quality of service. In many evaluation research studies, investigators use mixed methods, with both quantitative and qualitative information.

Evaluation studies may be either formative or summative. In **formative evaluation**, a program is assessed as it is being implemented; usually, the focus is on evaluation of the process of a program rather than the outcomes. In **summative evaluation**, the outcomes of a program are assessed after completion of the initial program.

Fraser et al. (2017) used a summative evaluation to assess the effects of an adult and feedback delivered to care providers on home care client outcomes. In contrast Vanstone et al. (2020) used a formative evaluation to evaluate the implementation of a 3 Wishes Project (3WP) in the ICUs. Knowledge related to summative (outcomes) and formative (process) evaluation of programs is important in translating research into clinical practice.

The use of experimental and quasiexperimental designs in studies of quality improvement and evaluation enables researchers to determine not only whether care is adequate but also which method of care is best under certain conditions. Furthermore, such studies can be used to determine whether a particular type of nursing care or intervention is cost-effective—that is, that the care or intervention does what it is intended to do but at lower or equivalent cost. Cost studies

are usually incorporated into the evaluation of an intervention. For example, Conway et al. (2019) reported that implementation of a thermal care bundle for treating inadvertent perioperative hypothermia reduced cost and improved quality of life of patients undergoing surgery.

In an era of health care reform and cost containment for health expenditures, evaluating the relative costs and benefits of new programs of care has become increasingly important. Relatively few studies in nursing and medicine have been dedicated to such evaluation, but in terms of outcomes, nursing costs and cost savings will be important in future studies.

Research Hint

According to Gaudine and Lamb (2015), the term *quality assurance* in health care has been replaced with the term *quality improvement (QI)*, as quality can be improved, not assured. The purpose of QI is thus to bring about immediate improvement to processes and outcomes by using a systematic, evidence-informed approach and compare organizations' quality to standards or benchmarks. QI projects in health care focus on improving patient, staff, and institutional outcomes. The steps are to identify which outcomes need to be improved, discern how they will be improved, and develop a strategy to implement and evaluate QI outcomes. Thus, QI projects include an evaluation. These projects are site specific, and the results may not be generalizable knowledge to other groups or sites.

APPRAISING THE EVIDENCE

Experimental and Quasiexperimental Designs

As discussed earlier in the chapter, various designs for research studies differ in the amount of control the researcher has over the antecedent and intervening variables that may affect the results of the study. True experimental designs, which yield level II evidence, offer the most possibility for control, whereas nonexperimental designs, which yield level IV, V, or VI evidence, offer the least. Quasiexperimental designs, which yield level III evidence, offer evidence that lies somewhere in between. Research designs must balance the needs for internal validity and external validity in order to produce useful results. In addition, judicious use of design requires that the chosen design be appropriate to the problem, free of bias, and capable of answering the research question.

Questions that you should pose when reading studies that test cause-and-effect relationships are listed in the Critiquing Criteria box. All of these questions should help you judge, with confidence, whether a causal relationship exists.

For studies in which either experimental or quasiexperimental designs are used, first try to determine the type of design that was used. Often, a statement describing the design of the study appears in the abstract and in the "Methods" section of the article. If such a statement is not present, you should examine the study for evidence of the following three characteristics: control, randomization, and manipulation. If all are discussed, the design is probably experimental. Conversely, if the study involves the administration of an experimental treatment but does not involve the random assignment of participants to groups, the design is quasiexperimental. Next, try to identify which of the variations within these two types of designs was used. Determining the answer to these questions gives you a head start because inherent in each design are particular threats to validity, and this step makes it easier to critically evaluate the study. The next question to ask is whether the researcher required a solution to a cause-and-effect problem. If so, the study is suited to these designs. Finally, think about the conducting of the study in the setting. Is it realistic to think that the study could be conducted in a clinical setting without some contamination?

The most important question to ask as you read experimental studies is "What else could have happened to explain the findings?" Thus, the author must provide adequate accounts of how the procedures for randomization, control, and manipulation were carried out. The study should include a description of the procedures for random assignment to such a degree that the reader can determine the likelihood for any one participant to be assigned to a particular group. The description of the independent variable also should be detailed. The inclusion of this information helps the reader decide whether the treatment given to some participants in the experi-

APPRAISING THE EVIDENCE—*cont'd*

Experimental and Quasiexperimental Designs

mental group might differ from what was given to others in the same group. In addition, threats to validity, such as testing effects and mortality (attrition), should be addressed. Otherwise, the conclusions of the study could potentially be erroneous and less believable to the reader.

This question of potential alternative explanations or threats to internal validity for the findings is even more important when you critically evaluate a quasiexperimental study because these study designs cannot possibly control for many plausible alternative explanations. A well-written report of a quasiexperimental study systematically reviews potential threats to the validity of the findings. Then your work as the reader is to decide whether the author's explanations make sense. When critiquing evaluation research, you should look for a careful description of the program, policy, procedure, or treatment being evaluated. In addition, you may need to determine the design used to evaluate the program and assess the appropriateness of the design for the evaluation. Once you have discerned the design, you can assess threats to validity for the appropriate design in determining the appropriateness of the author's conclusions in relation to the outcomes. As with all research, the results of studies with these designs need to be generalizable to a larger population of people than was actually studied. Thus, researchers need to decide whether the experimental protocol eliminated some potential participants and whether this weakness affected not only internal validity but also external validity.

CRITIQUING CRITERIA

1. What design is used in the study?
2. Is the design experimental or quasiexperimental?
3. Is the problem one of a cause-and-effect relationship?
4. Is the method used appropriate to the problem?
5. Is the design suited to the setting of the study?

EXPERIMENTAL DESIGNS

1. What experimental design is used in the study, and is it appropriate?
2. How are randomization, control, and manipulation applied?
3. Are there reasons to believe that alternative explanations exist for the findings?
4. Are all threats to validity, including mortality (attrition), addressed in the report?
5. Whether the experiment was conducted in the laboratory or a clinical setting, are the findings generalizable to the larger population of interest?

QUASIEXPERIMENTAL DESIGNS

1. What quasiexperimental design is used in the study, and is it appropriate?
2. What are the most common threats to the validity of the findings of this design?
3. What are the plausible alternative explanations, and have they been addressed?
4. Are the author's explanations of threats to validity acceptable?
5. What does the author say about the limitations of the study?
6. Do other limitations related to the design exist that are not mentioned?

EVALUATION RESEARCH

1. Do the authors identify a specific problem, practice, policy, or treatment that they will evaluate?
2. Do the authors identify the outcomes to be evaluated?
3. Is the problem analyzed and described?
4. Is the program to be analyzed described and standardized?
5. Do the authors identify the measurement of the degree of change (outcome) that occurs?
6. Do the authors determine whether the observed outcome is related to the activity or to one or more other causes?

CRITICAL THINKING CHALLENGES

- Discuss the barriers to nurse researchers in meeting the three criteria of a true experimental design.
- How is it possible to have a research design that includes an experimental treatment intervention and a control group and yet is not considered a true experimental study? How does this affect the usefulness of the findings in an evidence-informed practice?

- Argue your case for supporting or not supporting the following claim: "The fact that true experimental design is not used does not decrease the value of the study, even though it may decrease the utility of the findings in practice." Include examples with your rationale.
- Respond to the following question: Why are experimental studies considered the best evidence for an evidence-informed practice model? Justify your answer.

CRITICAL JUDGEMENT QUESTIONS

1. What aspect should be the primary consideration in critiquing the research report of an experimental study and determining the validity of the conclusions presented?
   a. How well the researcher controlled for extraneous variables
   b. The direction of the relationship between the dependent and independent variables
   c. The credentials and previous experience of the researcher
   d. The number of persons involved in the data collection process
2. In a study to help people avoid overdosing on opioids, one group of participants received a single supportive phone call 10 days after attending a program on harm reduction approaches. A second group received a weekly supportive phone call for 6 weeks after attending the same program, and a third group received no supportive phone calls after attending the program. What property of experimental research did the researcher employ in this study?
   a. Quasiexperimental research—no control group
   b. Random assignment to research groups
   c. Manipulation of the intervention
   d. Controlling for extraneous variables
3. In a study to assess the effect of a videotape format teaching method on the learning of adolescent males about the warning signs of testicular cancer, why would a nurse researcher select to implement a quasiexperimental study design instead of an experimental study design?
   a. The study is planned to be conducted in a laboratory setting
   b. An experimental treatment is not part of the study
   c. The researcher has not conducted research before
   d. Full experimentational control is not possible

KEY POINTS

- Experimental designs or randomized clinical trials provide the strongest evidence (level II) in terms of whether an intervention or treatment affects patient outcomes.
- Two types of design commonly used in nursing research to test hypotheses about cause-and-effect relationships are experimental and quasiexperimental designs. Both are useful for the development of nursing knowledge because they test the effects of nursing actions and lead to the development of prescriptive theory.
- True experiments are characterized by the ability of the researcher to control extraneous variation, manipulate the independent variable, and randomly assign participants to research groups.
- Experiments conducted in clinical settings or in the laboratory provide the best evidence in support of a causal relationship because the following three criteria can be met: (1) the independent and dependent variables are related to each other; (2) the independent variable chronologically precedes the dependent variable; and (3) the relationship cannot be explained by the presence of a third variable.
- Researchers frequently use quasiexperimental designs to test cause-and-effect relationships because experimental designs are often impractical or unethical.
- Quasiexperiments may lack either randomization or the comparison group, or both, which are characteristics of true experiments. Their

usefulness in studying causal relationships depends on the ability of the researcher to rule out plausible threats to the validity of the findings, such as history threats, selection bias, and maturation and testing effects.
- The level of evidence (level III) provided by quasiexperimental designs weakens confidence that the findings were the result of the intervention rather than extraneous variables.
- The overall purpose of critiquing such studies is to assess the validity of the findings and to determine whether these findings are worth incorporating into the nurse's personal practice.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Anis, L., Letourneau, N. L., Benzies, K., Ewashen, C., & Hart, M. J. (2020). Effect of the attachment and child health parent training program on parent–child interaction quality and child development. *Canadian Journal of Nursing Research, 52*(2), 157–168. https://doi.org/10.1177/0844562119899004.

Bigman, S. K. (1961). Evaluating the effectiveness of religious programs. *Review of Religious Research, 2*, 99–110.

Campbell, D., & Stanley, J. (1996). *Experimental and quasiexperimental designs for research*. Chicago: Rand-McNally.

Clarke, A. (2001). Evaluation research in nursing and health care. *Nurse Researcher, 8*(3), 4–14.

Conway, A., Gow, J., Ralph, N., Duff, J., Edward, K., Alexander, K., Munday, J., & Bräuer, A. (2019). Implementing a thermal care bundle for inadvertent perioperative hypothermia: A cost-effectiveness analysis. *International Journal of Nursing Studies, 97*, 21–27. https://doi.org/10.1016/j.ijnurstu.2019.04.017.

Fraser, K. D., Sales, A. E., Baylon, M. A. B., Schalm, C., & Miklavcic, J. J. (2017). Data for Improvement and Clinical Excellence: A report of an interrupted time series trial of feedback in home care. *Implementation Science, 12*(66), 1–10. https://doi.org/10.1186/s13012-017-0600-1.

Gaudine, A., & Lamb, M. (2015). *Nursing leadership and management: Working in Canadian healthcare organizations*. North York, ON: Pearson.

Hickin, S. L., White, S., & Knopp-sihota, J. (2017). Nurses' knowledge and perception of delirium screening and assessment in the intensive care unit: Long-term effectiveness of an education-based knowledge translation intervention. *Intensive & Critical Care Nursing*, *41*, 43–49. https://doi.org/10.1016/j.iccn.2017.03.010.

Hunt, C., Michalak, A., Lefkimmiatis, C., Johnston, E., Macumber, L., Jocko, T., & Ouchterlony, D. (2018). Exploring concussion awareness in hockey with a First Nations community in Canada. *Public Health Nursing, 35*(3), 202–210. https://doi.org/10.1111/phn.12407.

Kim, S., & Kim, C. (2020). Effects of an electrocardiography training program: Team-based learning for early-stage intensive care unit nurses. *The Journal of Continuing Education in Nursing, 51*(4), 174–180. https://doi.org/10.3928/00220124-20200317-07.

Page-cutrara, K., & Turk, M. (2017). Impact of prebriefing on competency performance, clinical judgment and experience in simulation: An experimental study. *Nurse Education Today, 48*, 78–83. https://doi.org/10.1016/j.nedt.2016.09.012.

Sawhney, M., Watt-Watson, J., & McGillion, M. (2017). A pain education intervention for patients undergoing ambulatory inguinal hernia repair: A randomized controlled trial. *Canadian Journal of Nursing Research, 49*(3), 108–117. https://doi.org/10.1177/0844562117714704.

Sezgin, D., & Esin, M. N. (2018). Effects of a PRECEDE-PROCEED model based ergonomic risk management programme to reduce musculoskeletal symptoms of ICU nurses. *Intensive & Critical Care Nursing*, *47*, 89–97. https://doi.org/10.1016/j.iccn.2018.02.007.

Tryphonopoulos, P. D., & Letourneau, N. (2020). Promising results from a video-feedback interaction guidance intervention for improving maternal–infant interaction quality of depressed mothers: A feasibility pilot study. *Canadian Journal of Nursing Research, 52*(2), 74–87. https://doi.org/10.1177/0844562119892769.

Vanstone, M., Neville, T. H., Clarke, F. J., Swinton, M., Sadik, M., Takaoka, A., Smith, O., Baker, A. J., Leblanc, A., & Foster, D. (2020). Compassionate end-of-life care: Mixed-methods multisite evaluation of the 3 Wishes Project. *Annals of Internal Medicine, 172*, 1–12. https://doi.org/10.7326/M19-2438.

Verkuyl, M., Hughes, M., Atack, L., McCulloch, T., Lapum, J. L., Romaniuk, D., & St-Amant, O. (2019). Comparison of self-debriefing alone or in combination with group debrief. *Clinical Simulation in Nursing, 37*, 32–39. https://doi.org/10.1016/j.ecns.2019.08.005.

Williams, T. (2019). Using high fidelity simulation to prepare baccalaureate nursing students enrolled in a historically Black college and university. *ABNF Journal, 30*(2), 37–43. http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=136798779&site=ehost-live.

Nonexperimental Designs

Mina D. Singh | Ramesh Venkatesa Perumal

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Describe the overall purpose of nonexperimental designs.
- Describe the characteristics of survey and relationship/difference designs.
- Define the differences between survey and relationship/difference designs.
- List the advantages and disadvantages of surveys and each type of relationship/difference design.
- Identify methodological, secondary analysis, and meta-analysis research.
- Identify the purposes of methodological, secondary analysis, and meta-analysis research.
- Discuss relational inferences versus causal inferences as they relate to nonexperimental designs.
- Identify the criteria used to critique nonexperimental research designs.
- Apply the critiquing criteria to the evaluation of nonexperimental research designs as they appear in research reports.
- Apply levels of evidence to nonexperimental designs.

KEY TERMS

cohort
correlational study
cross-sectional study
descriptive/exploratory survey
developmental study
epidemiological study
ex post facto study
hierarchical linear modelling (HLM)
incidence
longitudinal study
meta-analysis
methodological research
nonexperimental research design
prediction study
prevalence
prospective study
psychometrics
relationship/difference study
retrospective data
retrospective study
secondary analysis
survey study

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

MANY PHENOMENA OF INTEREST AND RELEVANCE to nursing do not lend themselves to an experimental design. For example, nurses studying pain may be interested in knowing the amount of pain, variations in the amount of pain, and patients' responses to postoperative pain. The investigator would not design an experimental study that would potentially intensify a patient's pain just to study the pain experience; that would be unethical. Instead, the researcher would use a nonexperimental design to examine the factors that contribute to the variability in a patient's postoperative pain experience. Nonexperimental research designs are used in studies when the researcher wishes to construct a picture of a phenomenon; examine events, people, or situations as they naturally occur; or test relationships and differences among variables. Nonexperimental designs may enable the researcher to understand how a phenomenon occurs at one point or over a period of time.

In experimental research, the independent variable is manipulated; in a nonexperimental research design, the independent variable is not manipulated. In nonexperimental research, the independent variables have occurred naturally, and the investigator cannot directly control them by manipulation. In contrast, in an experimental design, the researcher actively manipulates one or more variables. The researcher in a nonexperimental design explores relationships or differences among the variables. Nonexperimental research requires a clear, concise research problem or hypothesis that is based on a theoretical framework. Even though the researcher does not actively manipulate the variables, the concepts of control (see Chapter 10) should be considered as much as possible.

Researchers do not agree on how to classify nonexperimental studies. A continuum of quantitative research designs is shown in Figure 12.1. This chapter divides nonexperimental designs into *survey studies* and *relationship/difference studies*, as illustrated in Box 12.1. These categories are flexible; nonexperimental studies may be classified in a different way in other sources. Some studies belong exclusively to one of these categories, whereas other studies have the characteristics of more than one category or more than one design label. As you read the research literature, you will often find that researchers who are conducting a nonexperimental study use several design classifications. This chapter introduces the various types of nonexperimental designs, their advantages and disadvantages, the use of nonexperimental research, the issues of causality, and the critiquing process as it relates to nonexperimental research. The Critical Thinking Decision Path outlines the path to the choice of a nonexperimental design.

Evidence-Informed Practice Tip

When you critically appraise nonexperimental studies, be aware of possible sources of bias that can be introduced at any point in the study.

BOX 12.1

SUMMARY OF NONEXPERIMENTAL RESEARCH DESIGNS

SURVEY STUDIES

- Descriptive
- Exploratory
- Comparative

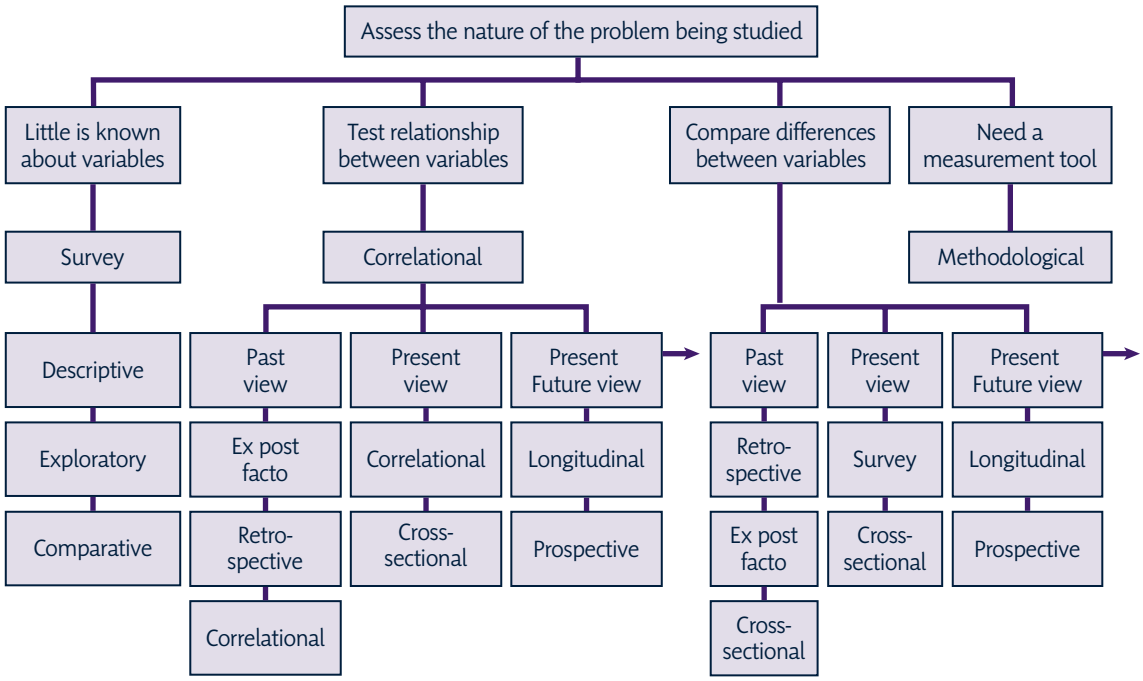
RELATIONSHIP/DIFFERENCE STUDIES

- Correlational
- Developmental
- Cross-sectional
- Longitudinal or prospective
- Retrospective or ex post facto

FIG. 12.1 Continuum of quantitative research designs.

CRITICAL THINKING DECISION PATH

Nonexperimental Design Choices



SURVEY STUDIES

The broadest category of nonexperimental designs is the survey study. In a survey study—further classified as *descriptive, exploratory*, or *comparative*—detailed descriptions of existing variables are collected, and the data are used to justify and assess current conditions and practices or to make more plans for improving health care practices. When you read research, you will find that the terms *exploratory, descriptive, comparative*, and *survey* may be used either alone, interchangeably, or together to describe the design of a study (Table 12.1). For example, investigators may use a descriptive/exploratory survey to search for accurate information about the characteristics of particular participants, groups, institutions, or situations or about the frequency of a phenomenon's occurrence, particularly when little is known about the phenomenon. The data are used to justify or assess current conditions or to make plans for improvement of conditions. Qualitative researchers also use the term *descriptive* in their reports, as in the study by Arnaert et al. (2019), who used a qualitative descriptive design to explore the experiences of midwives and community health workers using mHealth to improve services to pregnant women in rural Burkina Faso. You will be able to determine the difference in study type by checking in the analysis and findings sections, inasmuch as the qualitative descriptive study entails the use of the analyses outlined in Chapter 16, whereas the descriptive correlational or survey studies entail the use of descriptive and inferential statistical analyses.

In survey studies, the types of variables of interest can be classified as opinions, attitudes, or facts. For example, Covell and colleagues (2017) conducted a Pan-Canadian survey to describe

TABLE 12.1

EXAMPLES OF STUDIES WITH MORE THAN ONE DESIGN LABEL

DESIGN TYPE	STUDY'S PURPOSE
Observational, cross-sectional	Della Pelle et al. (2018) conducted a study to assess Italian nurses' knowledge and attitudes towards gay and lesbian sexual orientation and lesbian, gay, bisexual, and transgender (LGBT) patients.
Descriptive, correlational	This descriptive correlational study examines relationship among change fatigue, resilience, and job satisfaction of hospital staff nurses (Brown et al., 2018).
Cross-sectional, descriptive	This study describes the knowledge and attitudes about pain management among nursing students (Hroch et al., 2019).
Explanatory, correlational	This study's aim was to determine the associations between staffing hours and quality of care indicators in long-term care (Boscart et al., 2018).
Retrospective, cross-sectional	The purpose of this investigation was to explore the barriers and facilitators to exercise in individuals with cancer in Ontario (Fernandez et al., 2015).
Prospective, cohort	Adhikari Dahal et al. (2017) followed two cohorts of pregnant women, beginning about 0–24 weeks gestation, until 4 months postpartum. The two cohorts were All OUR Families and Alberta Pregnancy Outcomes and Nutrition datasets. Each study had similar recruitment periods, which were between 2008 and 2012. The researchers were interested in predicting whether neighbourhood socioeconomic status predicted the risk of preterm birth.

a demographic and human capital profile of Internationally Educated Nurses (IENs) and to identify the key human capital characteristics and types of assistance that predict IENs' professional recertification and employment as regulated nurses. The human capital characteristics (fact variables) were nursing education, professional experience, and language proficiency. In another example, Freeman et al. (2020) investigated palliative care nurses' attitudes towards medical assistance in dying using a secure online survey platform.

Often, the terms survey and questionnaire are used interchangeably, but they are not the same. A survey is a data collection tool that includes both a questionnaire and/or an interview, whereas a questionnaire is a written set of questions (see Chapter 14). Surveys are also thought of as an approach to collecting data, being inclusive of both the tool (questionnaire) and the process of collecting, followed by the analysis of those questions. For example, Dahlke et al. (2019) used a survey, a questionnaire only, to obtain student nurses' perceptions of older people. Another example is the study by Schofield, Forchuk, Montgomery, and associates (2016), who used a face-to-face structured interview to compare personal health practices between individuals with mental illness and the general Canadian population.

Survey researchers study either small or large samples of participants recruited from defined populations. The sample can be either broad or narrow and can be made up of people or institutions. For example, if a primary care rehabilitation unit based on a case-management model is to be established in a hospital, researchers might survey prospective applicants' attitudes with regard to case management before the unit staff members are selected. In a broader example, if a hospital is contemplating converting all patient care units to a case-management model, a survey might be conducted to determine the attitudes of a representative sample of nurses in the hospital toward case management. The data might provide the basis for projecting the in-service needs of nursing with regard to case management. The scope and depth of a survey are a function of the nature of the problem.

In surveys, investigators attempt only to relate one variable to another or to assess differences between variables; they do not attempt to determine causation. The two major advantages of surveys are the great deal of information that can be obtained from a large population in a more economical manner than face-to-face interviews and the surprising accuracy of survey research information. If a sample is representative of the population (see Chapter 13), a relatively small number of participants can accurately represent the views of the population.

However, survey studies have several disadvantages. First, the information obtained in a survey tends to be superficial. The breadth rather than the depth of the information is emphasized. Second, conducting a survey requires a great deal of expertise in various research areas. The survey investigator must have skills in sampling techniques, questionnaire construction, interviewing, and data analysis to elicit reliable and valid data. Third, large-scale surveys can be time consuming and costly, although the use of on-site personnel can reduce costs.

Research Hint

Research consumers should recognize that a well-constructed survey can provide a wealth of data about a particular phenomenon of interest, even though causation is not being examined.

Evidence-Informed Practice Tip

Evidence obtained from a survey population may be coupled with clinical expertise and applied to a similar population to develop an educational program to enhance knowledge and skills in a particular clinical area. For example, a survey designed to measure nursing staff's knowledge and attitudes about evidence-informed practice may yield data that are used to develop a staff development course in evidence-informed practice.

RELATIONSHIP/DIFFERENCE STUDIES

Investigators endeavour to trace the relationships or differences between variables that can provide a deeper insight into a phenomenon. This type of study can be classified as a relationship/difference study. The following types of relationship/difference studies are discussed here: *correlational studies* and *developmental studies*.

Correlational Studies

In a correlational study, an investigator examines the relationship between two or more variables. The researcher is not testing whether one variable causes another variable or how different one variable is from another variable. Instead, the researcher is testing whether the variables covary; in other words, as one variable changes, does a related change occur in the other variable? The researcher using this design is interested in quantifying the strength of the relationship between the variables or in testing a hypothesis about a specific relationship. The positive or negative direction of the relationship is also a central concern (see Chapter 17 for an explanation of the correlation between variables).

In their correlational study, Levya et al. (2019) described nursing faculty attitudes and beliefs about caring for people living with HIV/AIDS. They explored the relationships between prejudice, stereotype, and discrimination. These researchers were not testing a cause-and-effect relationship.

Another example of correlational research is the study by Rayan (2019), who explored the relationship between mindfulness, self-efficacy, and stress among final-year nursing students. They found that stress was significantly and negatively associated with mindfulness and self-efficacy.

Correlational studies offer researchers and research consumers the following advantages:

- An increased flexibility when investigating complex relationships among variables
- An efficient and effective method of collecting a large amount of data about a problem
- A potential for practical application in clinical settings
- A potential foundation for future experimental research studies
- A framework for exploring the relationship between variables that cannot be inherently manipulated

The correlational design has a quality of realism and is particularly appealing because it suggests the potential for practical solutions to clinical problems. However, there are disadvantages of correlational studies:

- Inability to manipulate the variables of interest
- No randomization in the sampling procedures because the study deals with pre-existing groups; therefore, generalizability is decreased
- Inability to determine a causal relationship between the variables because of the lack of manipulation, control, and randomization

One of the most common misuses of a correlational design is the researcher's conclusion that a causal relationship exists between the variables. In their correlational study, Winsett et al. (2016) explored the nurse work environment by evaluating the self-report of missed nursing care and the reasons for the missed care. They correctly determined that staffing adequacy was found to have an inverse relationship with communication, material resource, and labor resource. They recommended that "additional staffing may not always be feasible; but rethinking approaches to the nursing care delivery system, using the current staff in more efficient ways is certainly within reach" (p. 132).

Correlational studies may be further labelled *descriptive correlational* or *predictive correlational*. An example of a descriptive correlational study is by Jin et al. (2017) where they described relationships among perceptions of functional deficits, mood states, and empathic responses in family caregivers for patients with strokes.

The inability to draw causal statements should not lead you to conclude that a nonexperimental correlational study has a weak design. In terms of evidence for practice, researchers—on the basis of the literature review and their findings—frame the utility of the results in view of previous research and therefore help establish supportive evidence of the applicability of the results to a specific patient population. A correlational design is very useful for clinical research studies because many of the phenomena of clinical interest are beyond the researcher's ability to manipulate, control, and randomize.

Developmental Studies

Nonexperimental designs in which a time perspective is used can be further subclassified. A developmental study is concerned not only with the existing status and the relationship and differences among phenomena at one point in time but also with changes that occur as a function of time. The following three types of developmental study designs are discussed here: *cross-sectional, longitudinal* or *prospective*, and *retrospective* or *ex post facto*. Remember that in the literature, studies may be designated by more than one design name. This practice is accepted because many studies have elements of several nonexperimental designs. Table 12.1 provides examples of studies classified with more than one design label.

Cross-Sectional Studies

In a cross-sectional study, researchers examine data at one time; in other words, the data are collected on only one occasion with the same participants rather than with the same participants at several times.

An example of a cross-sectional study is provided by Freeman et al. (2020) who investigated palliative care nurse attitudes towards medical assistance in dying. In another example, Rochefort et al. (2016) explored the rationing of nursing care interventions and its associations with nurse-reported outcomes in the neonatal intensive care unit. Wall et al. (2018) conducted a cross-sectional study to explore the relationship between psychosocial health (anxiety, stress, depression) and preterm birth in Tanzanian women.

Evidence-Informed Practice Tip

Replication of significant findings in nonexperimental studies, with similar or different populations, or both, increases your confidence in the conclusions offered by the researcher and the strength of evidence generated by consistent findings from more than one study.

Longitudinal or Prospective Studies

In contrast to the cross-sectional design, the longitudinal study or prospective study (also referred to as *repeated-measures* studies) involves collecting data from the same group at different times. Researchers also use longitudinal studies to explore differences and relationships. For example, the investigator conducting a study with children with diabetes could use a longitudinal design. In that case, the investigator could collect yearly data or monitor the same children over a number of years to compare changes in the variables at different ages. By collecting data from each participant at yearly intervals, the investigator obtains a longitudinal perspective of the diabetic process.

An example of a prospective cohort study is where the researchers examined and compared risk factors for postpartum depression among: (1) recent (less and equal to 5 years) migrant and Canadian-born women, and (2) refugee, asylum-seeking, and non-refugee immigrant women. Women completed questionnaires at 1–2 weeks and at 16 weeks postpartum (Dennis et al., 2017). Castonguay et al. (2017) conducted a prospective study over 6 months to test body-related shame and guilt as predictors of breast cancer survivors' moderate to vigorous intensity physical activity.

Cross-sectional and longitudinal designs have many advantages and disadvantages. When assessing the appropriateness of a cross-sectional study versus a longitudinal study, the research consumer should first assess the researcher's goal in view of the theoretical framework. For example, in a hypothetical study of infant colic, the researchers are investigating a developmental process; therefore, a longitudinal design seems more appropriate. However, the disadvantages inherent in a longitudinal design also must be considered. The period of data collection may be long because of the time the participants take to progress to each data-collection point. In the infant colic study, it might take the researchers between 12 and 18 months to collect the data from the total sample. Threats to internal validity, such as testing and attrition, also are ever-present and unavoidable in a longitudinal study (see Chapter 17). As a result, longitudinal designs are costly in terms of time, effort, and money. Moreover, confounding variables could affect interpretation of the results. Participants in these studies may respond in a socially desirable way that they believe is congruent with the investigators' expectations (see discussion of the Hawthorne effect, in Chapter 10).

Despite the pragmatic constraints imposed by a longitudinal study, the researcher should proceed with this design if the theoretical framework supports a longitudinal developmental perspective. The advantages of a longitudinal study are that participants are monitored separately and thereby serve as their own controls; an increased depth of responses can be obtained; and early trends in the data can be analyzed. The researcher can assess changes in the variables of interest over time and explore both relationships and differences between variables.

Cross-sectional studies, in comparison with longitudinal studies, are less time consuming and less expensive and are thus more manageable for the researcher. Because large amounts of data can be collected at one time, the results are more readily available. In addition, the confounding variable of maturation, which results from the passage of time, is not present. However, the investigator's ability to establish an in-depth developmental assessment of the interrelationships of the phenomena being studied is reduced. Thus, the researcher is unable to determine whether the change that occurred is related to the change that was predicted because the same participants were not monitored over a period of time. In other words, the participants are unable to serve as their own controls (see Chapter 11).

In summary, longitudinal studies begin in the present and end in the future, and cross-sectional studies encompass a broader perspective of a cross-section of the population at one specific time.

Evidence-Informed Practice Tip

The quality of evidence provided by a longitudinal cohort study is stronger than that from other nonexperimental designs because the researcher can determine the incidence of a problem and its possible causes.

Retrospective or Ex Post Facto Studies

A retrospective study is essentially the same as an ex post facto study. Epidemiologists primarily use the term *retrospective*, whereas social scientists prefer the term *ex post facto*. In either case, the dependent variable has already been affected by the independent variable, and the investigator attempts to link current events to past events.

When scientists wish to explain causality or the factors that determine the occurrence of events or conditions, they prefer to use an experimental design. However, they cannot always manipulate the independent variable or use random assignments. In cases in which experimental designs cannot be employed, ex post facto studies may be used. *Ex post facto* literally means "from after the fact." These studies also are known as *causal-comparative* studies or *comparative* studies. As this design is discussed further, you will see that ex post facto research is similar to quasiexperimental research because in both, differences between variables are examined.

In retrospective studies, a researcher hypothesizes, for example, that variable *X* (cigarette smoking) is related to and a determinant of variable *Y* (lung cancer), but *X*, the presumed cause, is not manipulated, and participants are not randomly assigned to groups. Instead, the researcher chooses a group of participants who have experienced *X* (cigarette smoking) in a normal situation and a control group of participants who have not experienced *X*. The behaviours, performances, or conditions (lung tissue) of the two groups are compared in order to determine whether the exposure to *X* had the effect predicted by the hypothesis. Table 12.2 illustrates this example and reveals that although cigarette smoking appears to be a determinant of lung cancer, the researcher is still not able to conclude that a causal relationship exists between the variables because the independent variable has not been manipulated and the participants were not randomly assigned to groups.

Another example of a retrospective study is that of Dall'Ora et al. (2019) who investigated whether working 12-hr shifts is associated with increased sickness absence among registered nurses and health care assistants (HCAs). They conducted a retrospective longitudinal study analyzing data on all shifts scheduled for RNs and HCAs over a 3-year period. They found that 1,689 staff (86%) experienced at least one sickness episode during the 3-year study. The sickness episodes ranged from 1 day to 496 days in length; the most common length of sickness episodes was 2 days (n = 1,221, 15.1%). 2,532 (31.3%) sickness episodes lasted 7 or more days and were classified as long-term sickness episodes, while 5,555 sickness episodes lasting less than 7 days were classified as short-term sickness episodes. Another example is the study by Strazzieri-Pulido et al. (2019) who estimated the incidence of pressure injury and its predictors including nursing workload in critical patients. They studied a retrospective cohort of 766 patients in nine intensive care

TABLE 12.2

PARADIGM FOR THE EX POST FACTO DESIGN

GROUPS (NOT RANDOMLY ASSIGNED)	INDEPENDENT VARIABLE (NOT MANIPULATED BY INVESTIGATOR)	DEPENDENT VARIABLE
Exposed group: cigarette smokers	*X*: cigarette smoking	Y_e: lung cancer
Control group: nonsmokers		Y_c: no lung cancer

units of two university hospitals over a 3-month period. Pressure injury was present in 143 patients totalling an incidence of 18.7% within 766 patients. On average, pressure injuries developed in 6.9 days (SD = 5.9), with a median of 4 days and a variation of 2–30 days (p.303). The significant predictors of the pressure injury incidence were length of hospitalization, mechanical ventilation, palliative care, age, and nursing activities (p. 307).

The advantages of the retrospective design are similar to those of the correlational design. The additional benefit of the retrospective design is that it offers a higher level of control than a correlational study. For example, in a cigarette smoking study, the lung tissue samples from nonsmokers and smokers could be compared. This comparison would enable the researcher to establish the existence of a differential effect of cigarette smoking on lung tissue. However, the researcher would remain unable to draw a causal link between the two variables. This inability is the major disadvantage of the retrospective design.

Another disadvantage of retrospective research is that an alternative hypothesis may be the reason for the documented relationship. If the researcher obtains data from two existing groups of participants, such as one that has been exposed to *X* and one that has not, and the data support the hypothesis that *X* is related to *Y*, the researcher cannot be sure whether *X* or an extraneous variable is the cause of the occurrence of *Y*. Finding naturally occurring groups of participants who are similar in all ways except for their exposure to the variable of interest is very difficult. The possibility always exists that the groups differ in another way (e.g., in exposure to another lung irritant, such as asbestos), which can affect the findings of the study and produce spurious results. Consequently, when you read about such a study, you need to cautiously evaluate the conclusions drawn by the investigator.

Research Hint

When you read research reports, you will find that, at times, researchers classify a study's design with more than one design label. This classification is correct because research studies often reflect aspects of more than one design.

Longitudinal or prospective (cohort) studies are less common than retrospective studies because it can take a long time for the phenomenon of interest to become evident in a prospective study. For example, if researchers were studying pregnant women who regularly consume alcohol, it would take 9 months for the effect of low birth weight in the participants' infants to become evident. The problems inherent in a prospective study are therefore similar to those of a longitudinal study. However, longitudinal or prospective studies are considered stronger than retrospective studies because of the degree of control that can be imposed on extraneous variables that might confound the data.

Research Hint

Remember that nonexperimental designs can test relationships, differences, comparisons, or predictions, depending on the purpose of the study.

PREDICTION AND CAUSALITY IN NONEXPERIMENTAL RESEARCH

Researchers and research consumers are concerned with the issues of prediction and causality in explaining cause-and-effect relationships. Historically, researchers have said that only experimental research can support the concept of causality. For example, nurses are interested in discovering what causes anxiety in many settings. If nurses can uncover the causes, they can perhaps develop interventions that would prevent or decrease the anxiety. Causality makes it necessary to order events chronologically; therefore, if nurses find in a randomized experiment that event 1 (stress) occurs before event 2 (anxiety) and that

participants who experienced stress were anxious, whereas those in the unstressed group were not, then the hypothesis that stress causes anxiety is supported. If these results occurred in a nonexperimental study in which some participants underwent the stress of surgery and were anxious, whereas others did not have surgery and were not anxious, an association or relationship would be said to exist between stress (surgery) and anxiety. The results of a nonexperimental study, however, do not imply that the stress of surgery caused the anxiety.

Many variables (e.g., anxiety) that nurse researchers wish to study to explore causation cannot be manipulated, nor would it be wise to try to manipulate the variables. However, studies that can assert a predictive or causal sequence are needed. In view of this need, many nurse researchers use several analytical techniques that can explain the relationships among variables to establish predictive or causal links. These techniques are called *causal modelling, model testing*, path analysis, and *associated causal analysis*.

Practical Application

Fox et al. (2018) used structural equation modelling to test the direct and indirect relationships between geriatric practice environment, geriatric nursing practice, and overall quality of care for older adults and their families, while controlling for nurse and hospital characteristics. One of their findings is that the geriatric practice environment is directly associated with quality of care.

You will also find the terms path analysis, LISREL, analysis of covariance structures, structural equation modelling (SEM) (see Practical Application box), and hierarchical linear modelling (HLM) (Raudenbush & Bryk, 2002) used to describe the statistical techniques (see Chapter 17) used in these studies. An HLM is a type of regression analysis that allows for analysis of hierarchically structured data simultaneously at all levels (see the following Practical Application box for an example of the use of HLM).

Practical Application

An example of HLM appeared in a study by Pien et al. (2019) who validated the Chinese version Psychosocial Safety Climate Scale (PSC-12) and examined the associations between PSC, workplace violence, and self-rated health (SRH). In the hierarchical linear model, they found that participants from hospitals with the lowest PSC score had twofold risks of having poor SRH (p. 584).

In a prediction study, a model may be tested to assess which independent variables can best explain one or more dependent variables in order to make a forecast or prediction derived from particular phenomena. For example, Stephenson, DeLongis, Steele, and colleagues (2017) examined the role of maternal posttraumatic growth in changes in behavioural problems among siblings of children with complex chronic health conditions. Results from a time-lagged multilevel regression revealed that higher levels of maternal posttraumatic growth predicted subsequent declines in parent-reported internalizing, externalizing, and total behavioural problems among healthy siblings.

Research Hint

Nonexperimental clinical research studies have progressed to the point at which prediction models are used to explore or test relationships between independent variables and dependent variables.

As nurse researchers develop their programs of research in a specific area, more tests of models will be available. The statistics used in model-testing studies are advanced, but the beginning research consumer should be able to read the article, understand the purpose of the study, and determine whether the model generated was logical and developed with a solid basis from the literature and past research.

A full description of the techniques and principles of causal modelling is beyond the scope of this text.

Evidence-Informed Practice Tip

Research studies that entail the use of nonexperimental designs and provide level IV evidence can build the foundation for a program of research that leads to experimental designs in which the effectiveness of nursing interventions can be tested.

ADDITIONAL TYPES OF QUANTITATIVE STUDIES

Other types of quantitative studies complement the science of research. These additional designs provide a means of viewing and interpreting phenomena to provide further breadth and knowledge to nursing science and practice. These types of quantitative studies are methodological research, systematic review, meta-analysis, integrative review, secondary analysis, and epidemiological studies.

Methodological Research

Methodological research is the development and evaluation of data-collection instruments, scales, and techniques. As noted in Chapters 14 and 15, methodology has a strong influence on research. The most significant and important aspect of methodological research addressed in measurement development is psychometrics—the theory and development of measurement instruments (such as questionnaires) and measurement techniques (such as observational techniques) through the research process. Thus, psychometrics is concerned with the measurement of a concept, such as anxiety or interpersonal conflict, with reliable and valid tools. (See Chapter 15 for a discussion of reliability and validity.)

Nurse researchers have used the principles of psychometrics to develop and test measurement instruments that focus on nursing phenomena. Nurse researchers also use instruments developed in other disciplines, such as psychology and sociology, in which tools have been psychometrically tested. Sound measurement tools are critical for the reliability and validity of a study. A study's purpose, problems, and procedures may be clear, and the data analysis may be correct and consistent, but if the measurement tool has inherent psychometric problems, the findings will be rendered questionable or of limited utility.

The main problem for nurse researchers is locating appropriate measurement tools. Many of the phenomena of interest in nursing practice and research are intangible, such as interpersonal conflict, caring, coping, and maternal–fetal attachment. The intangible nature of various phenomena, and yet the need to measure them, places methodological research in an important position in research. Methodological research differs from other designs of research. First, it does not include all of the research process steps discussed in Chapter 3. Second, to implement methodical research techniques, the researcher must have a sound knowledge of psychometrics or must consult with a researcher knowledgeable in psychometric techniques. The methodological researcher is not interested in the relationship of the independent variable to a dependent variable or in the effect of an independent variable on a dependent variable. Instead, the methodological researcher is interested in identifying an intangible construct (concept) and making it tangible with a paper-and-pencil instrument or observation protocol.

A methodological study includes the following steps:

- Defining the construct, or concept, or behaviour to be measured
- Formulating the tool's items
- Developing instructions for users and respondents
- Testing the tool's reliability and validity

A sound, specific, and exhaustive literature review is necessary to identify the theories underlying the steps in this construct. The literature review provides the basis of item formulation. Once the items have been developed, the researcher assesses the tool's reliability and validity (see Chapter 15). Various aspects of these procedures

may differ according to the tool's use, purpose, and stage of development.

In an example of methodological research, Boscart et al. (2018) documented the evaluation of the psychometric properties of the Team Member Perspectives of Person-Centered Care (TM-PCC) Survey. Another example is the study by Vincelette et al. (2018), where the researchers developed and validated the Nurse Cardiopulmonary Resuscitation Survey (NCRS) among intensive care unit nurses.

Common considerations that researchers incorporate into methodological research are outlined in Table 12.3. Psychometric or methodological studies are found primarily in journals that report research. The *Journal of Nursing Measurement* is devoted to the publication of information on instruments, tools, and approaches for measurement of variables.

TABLE 12.3

COMMON CONSIDERATIONS IN THE DEVELOPMENT OF MEASUREMENT TOOLS

CONSIDERATION	COMMENT
The well-constructed scale, test, interview schedule, or other form of index should consist of an objective, standardized measure of samples of a behaviour that has been clearly defined. Observations should be made on a small but carefully chosen sampling of the behaviour of interest, thus creating confidence that the samples are representative.	A new tool should be based on a thorough review of previous theoretical and research literature to ensure validity.
The tool should be standardized; that is, a set of uniform items and response possibilities are uniformly administered and scored.	Without specific criteria and rating procedures, the evaluations of the items would be based on the subjective impressions, which may have varied significantly between observers and conditions.
The items of a measurement tool should be unambiguous; they should be clear-cut, concise, exact statements with only one idea per item. Negative stems or items with negatively phrased response possibilities result in ambiguity in meaning and scoring.	For example, in constructing a tool to measure job satisfaction, a nurse scientist writes the following item: "I never feel that I don't have time to provide good nursing care." The response format consists of "Agree," "Undecided," and "Disagree." A response of "Disagree" will likely not reflect the respondent's true intention because of the confusion that is created by the double-negative phrasing "never . . . don't."
The type of items used in any one test or scale should be restricted to a limited number of variations. Participants who are expected to shift from one kind of item to another may fail to provide a true response as a result of the distraction of making such a change.	Mixing true-or-false items with questions that require a yes-or-no response and items that provide a response format of five possible answers can lead to a high level of measurement error.
Items should not provide irrelevant clues. Unless carefully constructed, an item may furnish an indication of the expected response or answer. Furthermore, the correct answer or expected response to one item should not be given by another item.	An item that provides a clue to the expected answer may contain value words that convey cultural expectations, such as "A good wife enjoys caring for her home and family."
The items of a measurement tool should not be made difficult by requiring unnecessarily complex or exact operations. Furthermore, the difficulty of an item should be appropriate to the level of the participants being assessed. Limiting each item to one concept or idea helps accomplish this objective.	A test constructed to evaluate learning in an introductory course in research methods may contain an item that is inappropriate for the designated group, such as "A nonlinear transformation of data to linear data is a useful procedure before a hypothesis of curvilinearity is tested."

TABLE 12.3

COMMON CONSIDERATIONS IN THE DEVELOPMENT OF MEASUREMENT TOOLS—cont'd

CONSIDERATION	COMMENT
The diagnostic, predictive, or measurement value of a tool depends on the degree to which it serves as an indicator of a relatively broad and significant area of behaviour, known as the *universe of content* for the behaviour. As already emphasized, a behaviour must be clearly defined before it can be measured. The definition is developed from the universe of content: that is, the information and research findings that are available for the behaviour of interest. The items should reflect that definition. The extent to which the test items appear to accomplish this objective is an indication of the validity of the instrument.	Two nurse researchers are studying the construct of quality of life. Each nurse has defined this construct in a different way. Consequently, the measurement tool that each nurse devises will include different questions. The questions on each tool will reflect the universe of content for quality of life as defined by each researcher.
The instrument also should adequately cover the defined behaviour. The primary consideration is whether the number and nature of items in the sample are adequate. If the sample has too few items, the accuracy or reliability of the measure must be questioned. In general, the sample should have a minimum of 10 items for each independent aspect of the behaviour of interest.	For example, few people would be satisfied with an assessment of intelligence if the scale were limited to three items.
The measure must prove its worth empirically through tests of reliability and validity.	The researcher should demonstrate to the reader that the scale is accurate and measures what it purports to measure (see Chapter 15).

The specific procedures of methodological research are beyond the scope of this book, but you are urged to look closely at the tools used in studies.

Systematic Review

A *systematic review* is a summation and assessment of research studies found in the literature based on a clearly focused question that uses systematic and explicit methods to identify, select, critically appraise, and analyze relevant data from the selected studies to summarize the findings in a focused area (see Chapter 3). The strength of evidence provided by systematic reviews is a key component for developing a practice based on evidence. The qualitative counterpart to a systematic review is meta-synthesis, which uses qualitative principles to assess qualitative research and is described in Chapter 5. In a systematic review, statistical methods such as a meta-analysis may or may not be used to analyze the studies reviewed. A systematic review provides the most powerful and useful evidence available to guide practice: level I evidence (see Chapter 3). Systematic reviews that use multiple randomized clinical trials (RCTs) to combine study results offer stronger evidence (level I) in estimating the magnitude of an effect for an intervention.

You will also find reviews of an area of research or theory synthesis termed *integrative reviews*, discussed later in the chapter. Systematic and integrative reviews are not designs per se, but methods for searching and integrating the literature related to a specific clinical issue. These methods take the results of many studies in a specific area, assesses the studies critically for reliability and validity (quality, quantity, and consistency), and synthesize findings to inform practice.

Meta-analysis provides level I evidence—the highest level of evidence, as it involves statistically analyzing and integrating the results of many studies. Systematic reviews and meta-analyses also grade the level of design or evidence of the studies reviewed. Of all the review types, a meta-analysis provides the strongest summary support because it summarizes studies using data analysis. Box 12.2 outlines the path for completing a systematic review.

The components of a systematic review are the same as in a meta-analysis (Box 12.3) except for the analysis of the studies. An example of a systematic review is that by Richards et al. (2018) on the effect of nursing care interventions for nutrition, elimination, mobility, and hygiene, where the authors:

- Synthesized the literature from studies on the effect of each of nutrition, elimination, mobility, and hygiene
- Included a clear clinical question; all of the sections of a systematic review were presented; in addition, there was a statistical meta-analysis (combination of studies data) of the studies as a whole.

Each study in this review was considered individually, not collectively, for its sample size, effect size, and its contribution to knowledge in the area, based on a set of criteria.

Another systematic review was conducted by Jones et al. (2019) to determine the predictors of infant care competence among women with postpartum depression.

BOX 12.2

COMPLETING A SYSTEMATIC REVIEW

A systematic review is a summary of the quantitative research literature that used similar designs based on a focused clinical question. The goal is to bring together all of the studies concerning a focused clinical question and, using rigorous inclusion and exclusion criteria, assess the strength and quality of the evidence provided by the chosen studies in relation to:

- Sampling issues
- Internal validity (bias) threats
- External validity
- Data analysis

The purpose is to report, in a consolidated fashion, the most current and valid research on intervention effectiveness and clinical knowledge, which will ultimately be the basis for evidence-informed decision-making about the applicability of findings to clinical practice.

Once the studies in a systematic review are gathered from a comprehensive literature search (see Chapter 5), they are assessed for quality and synthesized according to quality or focus; then practice recommendations are made and presented in an article. More than one person independently evaluates the studies to be included or excluded in the review. Generally, the articles critically appraised are discussed in the article and presented in a table format within the article, which helps you to easily identify the specific studies gathered for the review and their quality. The most important principle to assess when reading a systematic review is how the author(s) of the review identified the studies to evaluate and how they systematically reviewed and appraised the literature that leads to the reviewers' conclusions.

BOX 12.3

SYSTEMATIC REVIEW COMPONENTS WITH OR WITHOUT META-ANALYSIS

- Introduction
- Review rationale and a clear clinical question (PICOT)
- Methods
- Information sources, databases used, and search strategy identified: how studies were selected and data extracted as well as the variables extracted and defined
- Description of methods used to assess risk of bias, summary measures identified (e.g., risk, ratio); identification of how data are combined, if studies are graded what quality appraisal system was used (see Chapters 1, 18, and 19)
- Results
- Number of studies screened and characteristics, risk of bias within studies, if a meta-analysis, there will be a synthesis of results including confidence intervals, risk of bias for each study, and all outcomes considered
- Discussion
- Summary of findings including the strength, quality, quantity, and consistency of the evidence for each outcome
- Any limitations of the studies, conclusions, and recommendations of findings for practice
- Funding
- Sources of funding for the systematic review

Although systematic reviews are highly useful, they also have to be reviewed for potential bias. Thus the studies in a review need to be carefully critiqued for scientific rigour in each step of the research process.

Meta-Analysis

A meta-analysis is a systematic summary using statistical techniques to assess and combine studies of the same design to obtain a precise estimate of effect (impact of an intervention on the dependent variable/outcomes or association between variables). The terms *meta-analysis* and *systematic review* are often used interchangeably, and this is incorrect. The main difference is that, as noted earlier, a meta-analysis includes a statistical assessment of the studies reviewed. Meta-analysis involves statistically analyzing the data from each of the studies, treating all the studies reviewed as one large data set in order to obtain a precise estimate of the effect (impact) of the results (outcomes) of the studies in the review.

Meta-analysis involves a rigorous process of summary and determining the impact of a number of studies rather than the impact derived from a single study alone (see Chapter 11). After the clinical question is identified and the search of the review of published and unpublished literature is completed, a meta-analysis is conducted in two phases:

Phase I: The data are extracted (i.e., outcome data, sample sizes, and measures of variability from the identified studies).

Phase II: The decision is made as to whether it is appropriate to calculate what is known as a pooled average result (effect) of the studies reviewed.

Effect sizes are calculated using the difference in the average scores between the intervention and control groups from each study. Each study is considered a unit of analysis. A meta-analysis takes the effect size (see Chapter 13) from each of the studies reviewed to obtain an estimate of the population (or the whole) to create a single effect size of all the studies. Thus the effect size is an estimate of how large a difference there is between intervention and control groups in the summarized studies.

In addition to calculating effect sizes, authors of meta-analyses use multiple statistical methods to present and depict the data from studies reviewed (see Chapters 20 and 21). One of these methods is a forest plot, sometimes called a blobbogram. A forest plot graphically depicts the results of analyzing a number of studies. Figure 12.2 is an example of a forest plot from the study by Rose et al. (2016) where they reviewed prenatal maternal anxiety as a risk factor for preterm birth.

Evidence-Informed Practice Tip

Evidence-informed practice methods such as meta-analysis increase your ability to manage the ever-increasing volume of information produced to develop the best evidence-informed practices.

Figure 12.2 displays the studies that compared the type of anxiety (anxiety disorder, pregnancy-specific disorder, state anxiety, general anxiety) and its relationship to having a spontaneous preterm birth. Each study analyzed is listed on the left under Study ID. To the right of the listed study is a horizontal line that identifies the risk ratio of anxiety, which is set at 1. The box on the vertical line represents the weight of each study, while the line is the confidence interval, and the diamond is the significance of the combined studies. The boxes to the left of the 1 line mean that anxiety did not produce a significant risk in preterm birth. The box to the right of the line indicates studies anxiety was significant in preterm birth. The diamond is a more precise estimate of the interventions, as it combines the data from the studies. The exemplar provided is basic, as meta-analysis is a sophisticated methodology.

A well-done meta-analysis assesses for bias in studies and provides clinicians a means of deciding the merit of a body of clinical research. Besides the repository of meta-analyses found in The Cochrane Library, published by The Cochrane Collaboration, meta-analyses can be found published in journals.

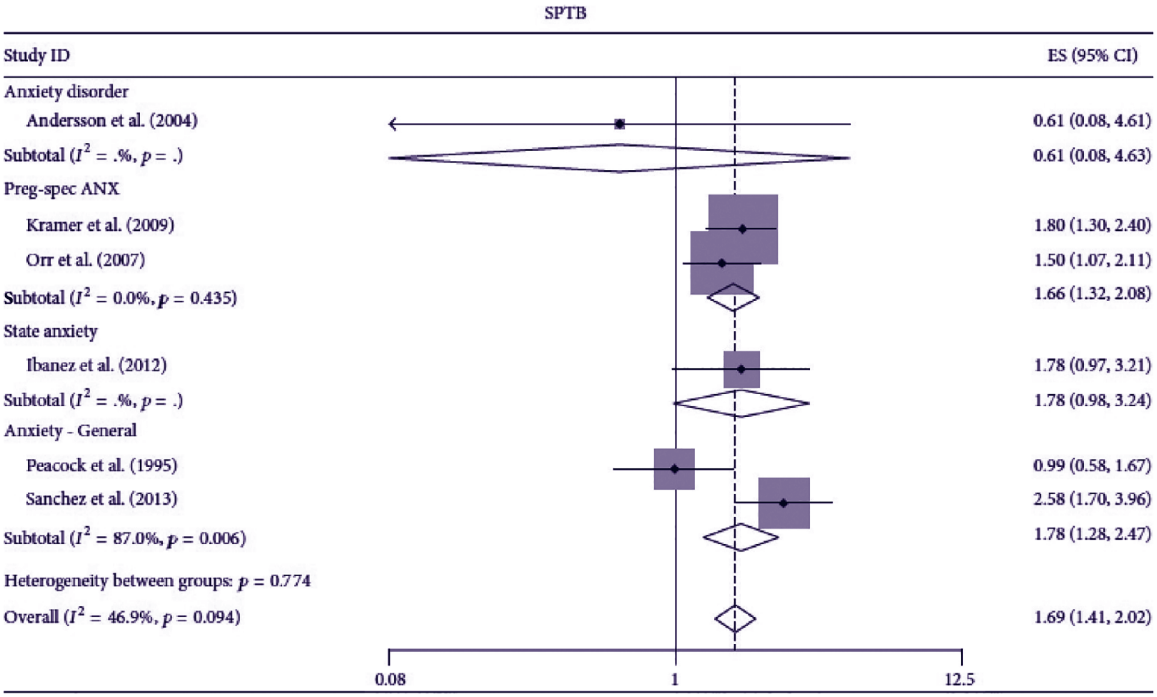


FIG. 12.2 An example of a forest plot. *CI*, confidence interval; *df*, degrees of freedom; *IV*, independent variable; *SD*, standard deviation.

The Cochrane Collaboration

The largest repository of meta-analyses is the Cochrane Collaboration/Review. The Cochrane Collaboration is an international organization that prepares and maintains a body of systematic reviews that focus on health care interventions (Box 12.4). The reviews are found in the Cochrane Database of Systematic Reviews. The Cochrane Collaboration collaborates with a wide range of health care individuals with different skills and backgrounds for developing reviews. These partnerships assist with developing reviews that minimize bias while keeping current with assessment of health care interventions, promoting access to the database, and ensuring the quality of the reviews (Higgins & Green, 2011). The steps of a Cochrane Report mirror those of a standard meta-analysis except for the plain language summary. This useful feature is a straightforward summary of the meta-analysis. The Cochrane Library also publishes several other useful databases (Box 12.5).

BOX 12.4

COCHRANE REVIEW SECTIONS

- Review information: Authors and contact person
- Abstract
- Plain language summary
- The review
- Background of the question
- Objectives of the search
- Methods for selecting studies for review
- Type of studies reviewed
- Types of participants, types of intervention, types of outcomes in the studies
- Search methods for finding studies
- Data collection
- Analysis of the located studies, including effect sizes
- Results including description of studies, risk of bias, intervention effects
- Discussion
- Implications for research and practice
- References and tables to display the data
- Supplementary information (e.g., appendices, data analysis)

BOX 12.5

COCHRANE LIBRARY DATABASES

- Cochrane Database of Systematic Reviews: Full-text Cochrane reviews
- Database of Abstracts of Review of Effects (DARE): Critical assessments and abstracts of other systematic reviews that conform to quality criteria
- Cochrane Central Register of Controlled Trials (CENTRAL): Information of studies published in conference proceedings and other sources not available in other databases
- Cochrane Methodology Register (CMR): Bibliographic information on articles and books on reviewing research and methodological studies

Integrative Review

Critical reviews of an area of research without a statistical analysis or a theory synthesis are termed *integrative reviews*. An integrative review is the broadest category of review. It can include theoretical literature, research literature, or both. An integrative review may include methodology studies, a theory review, or the results of differing research studies with wide-ranging clinical implications. An integrative review can include quantitative or qualitative research, or both. Statistics are not used to summarize and generate conclusions about the studies (see Chapter 3).

Secondary Analysis

Secondary analysis also is not a design but a form of research in which the previously collected and analyzed data from one study are reanalyzed for a secondary purpose. The original study may be either an experimental, a nonexperimental design, or Canadian national surveys; for example, Scott et al. (2016) conducted a study using the 2012 Canadian Community Health Survey – Mental Health data to describe the relationship between sexual orientation and depression in the Canadian population. Another method of obtaining secondary data is from a previous study that the researchers conducted; for example, Sidani et al. (2018) examined the extent to which participant characteristics (age, gender, education, race, employment), treatment type and method, and self-reported outcome factors contributed to satisfaction with the process and outcome attributes of therapies for insomnia. This study consists of a secondary analysis of data obtained from a partially randomized preference trial.

Epidemiological Studies

In an epidemiological study, factors affecting the health and illness of populations are examined in relation to the environment. The purview of public health for many years, epidemiological studies are investigations of the distribution, determinants, and dynamics of health and disease. In these studies, investigators attempt to link effects with cause; however, a clear understanding of the causes is often not possible, especially when the illness or problem has already occurred and the method is to look retrospectively at the evidence.

Some of the questions that epidemiological researchers attempt to answer are "Did exposure to a certain environment affect health?" and "Does staff shortage or do organizational issues affect burnout?" Research cannot answer such questions directly but can establish a statistically significant association between exposure to causative factors and disease or the effects of ill health.

Two frequently conducted types of epidemiological studies are studies of prevalence (the number of people affected by a disease or health problem) and studies of incidence (the number of cases occurring in a particular period).

TOOLS FOR EVALUATING INDIVIDUAL STUDIES

As the importance of practising from a base of evidence has grown, so has the need to have tools or instruments available that can assist practitioners in evaluating studies of various types. When evaluating studies for clinical evidence, it is first important to assess whether the study is valid. At the end of each chapter of this text are critiquing questions that will aid you in assessing whether

studies are valid and whether the results are applicable to your practice. In addition to these questions, there are standardized appraisal tools that can assist with appraising the evidence. The international collaboration Critical Appraisal Skills Programme (CASP), whose focus is on teaching critical appraisal, developed tools known as Critical Appraisal Skills Programme Checklists that provide an evidence-informed approach for assessing the quality, quantity, and consistency of specific study designs (CASP, 2018). These instruments are part of an international network that provides consumers with specific questions to help assess study quality. Each checklist has a number of general questions as well as design-specific questions. The tools centre on assessing a study's methodology, validity, and reliability. The questions focus on the following:

1. Are the study's results valid? Understanding the steps of research methodology, especially threats to internal validity as described in the previous and subsequent chapters, will assist in this process (see Chapters 10 through 17).
2. What are the results? This means, can you rely on the results (analysis) or the study's findings (see Chapters 17 and 18)?
3. Are the findings applicable to your practice? Chapters 20 and 21 are aimed at helping you with this decision.

Each CASP guideline is divided into one of the above three areas in a study. There are eight critical appraisal checklists. The checklist with instructions can be found the at CASP website at https://casp-uk.net/casp-tools-checklists/. The design-specific CASP tools with checklists are available online and include the following:

- Systematic reviews
- Randomized controlled studies
- Cohort studies
- Diagnostic studies
- Case-control studies
- Economic evaluations
- Qualitative studies
- Clinical prediction rule

CLINICAL PRACTICE GUIDELINES

Clinical practice guidelines are systematically developed statements or recommendations that link research and practice and serve as a guide for practitioners. Guidelines have been created to assist in bridging practice and research and are developed by professional organizations, government agencies, institutions, or convened expert panels. Guidelines provide clinicians with an algorithm for clinical management, to assist in decision-making for specific diseases (e.g., colon cancer), or for treatments (e.g., pain management). Not all guidelines are well developed and, like research, must be assessed before implementation (see Chapter 10). Guidelines should present scope and purpose of the practice, detail who the development group included, demonstrate scientific rigour, be clear in their presentation, demonstrate clinical applicability, and demonstrate editorial independence. An example is the National Comprehensive Cancer Network, which is an interdisciplinary consortium of 30 cancer centres across the world. Interdisciplinary groups develop practice guidelines for practitioners and education guidelines for patients. These guidelines are accessible at https://www.nccn.org.

The research findings in a clinical practice guideline need to be evaluated for quality, quantity, and consistency. Practice guidelines can be either expert-based or evidence-informed. Evidence-informed practice guidelines are those developed using a scientific process. This process includes first assembling a multidisciplinary group of experts in a specific field. This group is charged with completing a rigorous search of the literature and completing an evidence table that summarizes the quality and strength of the evidence on which the practice guideline is derived (see Chapters 20 and 21). For various reasons, not all areas of clinical practice have a sufficient research base; therefore, expert-based practice guidelines are developed. Expert-based guidelines depend on having a group of nationally known experts in the field who meet and

solely use opinions of experts along with whatever research evidence is developed to date. If limited research is available for such a guideline, a rationale should be presented for the practice recommendations.

Many national organizations develop clinical practice guidelines. It is important to know which one to apply to your patient population. For example, there are numerous evidence-informed practice guidelines developed for the management of pain. These guidelines are available from organizations such as the Oncology Nurses Society, American Academy of Pediatrics, National Comprehensive Cancer Network, National Cancer Institute, American College of Physicians, and American Academy of Pain Medicine.

The Agency for Healthcare Research and Quality (AHRQ) supports the National Guideline Clearinghouse (NGC). The NGC's mission is to provide health care providers from all disciplines with objective, detailed information on clinical practice guidelines and measures that are disseminated, implemented, and issued. Specific guidelines can be found on the AHRQ Effective Health Care Program website at https://www.ahrq.gov/programs/index.html?search_api_views_fulltext=&field_program_topics=14174.

Evaluating Clinical Practice Guidelines

As the number of evidence-informed practice guidelines increases, it becomes more important that you critique these guidelines with regard to the methods used for guideline formulation and consider how they might be used in practice. Critical areas that should be assessed when critiquing evidence-informed practice guidelines include the following:

- Date of publication or release and authors
- Endorsement of the guideline
- Clear purpose of what the guideline covers and patient groups for which it was designed
- Types of evidence (research, nonresearch) used in guideline formulation
- Types of research included in formulating the guideline (e.g., "We considered only randomized and other prospective controlled trials in determining efficacy of therapeutic interventions")
- Description of the methods used in grading the evidence
- Search terms and retrieval methods used to acquire evidence used in the guideline
- Well-referenced statements regarding practice
- Comprehensive reference list
- Review of the guideline by experts
- Whether the guideline has been used or tested in practice and, if so, with what types of patients and what types of settings

Evidence-informed practice guidelines that are formulated using rigorous methods provide a useful starting point for nurses to understand the evidence base of practice. However, more research may be available since the publication of the guideline, and refinements may be needed. Although information in well-developed, national, evidence-informed practice guidelines are a helpful reference, it is usually necessary to localize the guideline using institution-specific evidence-informed policies, procedures, or standards before application within a specific setting.

There are several tools for appraising the quality of clinical practice guidelines. The Appraisal of Guidelines for Research and Evaluation II (AGREE II) instrument (2013) is one of the most widely used tools to evaluate the applicability of a guideline to practice (Brouwers et al., 2010). The AGREE II was developed to assist in evaluating variability in guideline quality, provide a methodological strategy for guideline development, and inform practitioners about what information should be reported in guidelines and how it should be reported. The AGREE II is available online and replaces the original AGREE tool. The instrument focuses on six domains with a total of 23 questions rated on a 7-point scale and two final assessment items that require the appraiser to make overall judgments of the guideline based on how the 23 items were rated. Along with the

instrument itself, the AGREE Enterprise website offers guidance on tool usage and development. The AGREE II has been tested for reliability and validity. The guideline assesses the following components of a practice guideline:

1. Scope and purpose of the guideline
2. Stakeholder involvement
3. Rigour of guideline development
4. Clarity of presentation of the guideline
5. Applicability of the guideline to practice
6. Demonstrated editorial independence of the developers

APPRAISING THE EVIDENCE

Systematic Reviews and Clinical Practice Guidelines

For each of the review methods described—systematic review, meta-analysis, integrative review, and clinical practice guidelines—think about each method as one that progressively sifts and sorts research studies and the data until the highest quality of evidence is used to arrive at the conclusions. First the researcher combines the results of all the studies that focus on a specific question. The studies considered of lowest quality are then excluded and the data are reanalyzed. This process is repeated sequentially, excluding studies until only the studies of highest quality available are included in the analysis. An alteration in the overall results as an outcome of this sorting and separating process suggests how sensitive the conclusions are to the quality of studies included. No matter which type of review is completed, it is important to understand that the research studies reviewed still must be examined through your evidence-informed practice lens. This means that evidence that you have derived through your critical appraisal and synthesis or through other researchers' review must be integrated with an individual clinician's expertise and patients' wishes. The criteria for critiquing systematic reviews and clinical practice guidelines are presented in the Critiquing Criteria boxes.

CRITIQUING CRITERIA

SYSTEMATIC REVIEWS

1. Does the PICOT (Population, Intervention, Comparison, Outcome, Time) question used as the basis of the review match the studies included in the review?
2. Are the review methods clearly stated and comprehensive?
3. Are the dates of the review's inclusion clear and relevant to the area reviewed?
4. Are the inclusion and exclusion criteria for studies in the review clear and comprehensive?
5. What criteria were used to assess each of the studies in the review for quality and scientific merit?
6. If studies were analyzed individually, were the data clear?
7. Were the methods of study combination clear and appropriate?
8. If the studies were reviewed collectively, how large was the effect?
9. Are the clinical conclusions drawn from the studies relevant and supported by the review?

Clinical practice guidelines, though they are systematically developed and make explicit recommendations for practice, may be formatted differently. Practice guidelines should reflect the components listed. Guidelines can be located on an organization's website, on the Registered Nurses' Association of Ontario website (http://rnao.ca/bpg), on the AHRQ website (https://www.AHRQ.gov), or on MEDLINE (see Chapters 3 and 21). Well-developed guidelines are constructed using the principles of a systematic review.

CRITIQUING CRITERIA

CLINICAL PRACTICE GUIDELINES

1. Is the date of publication or release current?
2. Are the authors of the guideline clear and appropriate to the guideline?
3. Is the clinical problem and purpose clear in terms of what the guideline covers and

patient groups for which it was designed?
4. What types of evidence (research, nonresearch) were used in formulating the guideline, and are they appropriate to the topic?
5. Is there a description of the methods used to grade the evidence?
6. Were the search terms and retrieval methods used to acquire research and nonresearch evidence used in the guideline clear and relevant?
7. Is the guideline well referenced and comprehensive?
8. Are the recommendations in the guideline sourced according to the level of evidence for its basis?
9. Has the guideline been reviewed by experts in the appropriate field of discipline?
10. Who funded the guideline development?

Evidence-informed practice requires that you determine—based on the strength and quality of the evidence provided by the systematic review, coupled with your clinical expertise and patient values—whether or not you would consider a change in practice.

Evidence-Informed Practice Tip

Evidence-informed practice methods, such as systematic reviews, increase a nurse's ability to manage the ever-increasing volume of information produced to develop the best practices that are evidence informed.

Research Hint

As you read the literature, you will find studies with labels such as *outcomes research, needs assessments, evaluation research,* and *quality assurance.* These studies are not designs per se; instead, these studies are conducted with either experimental or nonexperimental designs. Studies with these labels are designed to test the effectiveness of health care techniques, programs, or interventions. When reading such a research study, you should assess which design was used and whether the principles of the design, sampling strategy, and analysis are consistent with the study's purpose.

APPRAISING THE EVIDENCE

Nonexperimental Designs

The criteria for critiquing nonexperimental designs are presented in the Critiquing Criteria box. When you critique nonexperimental research designs, keep in mind that such designs offer the researcher the least amount of control. The first step in critiquing nonexperimental research is to determine which type of design was used in the study. Often, a statement describing the design of the study appears in the abstract and in the "Methods" section of the report. If such a statement is not present, you should closely examine the report for evidence of which type of design was employed. You should be able to discern that either a survey or a relationship design was used, as well as the specific subtype. For example, you would expect an investigation of self-concept development in children from birth to 5 years of age to be a relationship study with a longitudinal design.

Next, you should evaluate the theoretical framework and underpinnings of the study to determine whether a nonexperimental design was the most appropriate approach to the problem. For example, in many of the studies on pain discussed throughout this text, the relationship between pain and any of the independent variables under consideration cannot be manipulated. For such studies, a nonexperimental correlational, longitudinal, or cross-sectional design is appropriate. Investigators use one of these designs to examine the relationship between the variables in naturally occurring groups. Sometimes, you may think that it would have been more appropriate for the investigators to use an experimental or a quasiexperimental design. However, you must recognize that pragmatic or ethical considerations also may have guided the researchers in their choice of design (see Chapters 6 and 11).

You should assess whether the problem is at a level of experimental manipulation. Often, researchers merely wish to examine whether relationships exist between

Continued

APPRAISING THE EVIDENCE—*cont'd*

Nonexperimental Designs

variables. Therefore, when you critique such studies, you should be able to determine the purpose of the study. If the purpose of the study does not include the expectation of a cause-and-effect relationship, you need not look for one. However, be wary when the researcher in a nonexperimental study suggests a cause-and-effect relationship in the findings.

Finally, the factor or factors that influence changes in the dependent variable are often ambiguous in nonexperimental designs. As with all complex phenomena, multiple factors can contribute to variability in the participants' responses. When an experimental design is not used for controlling some of these extraneous variables that can influence results, the researcher must strive to provide as much control of these variables as possible within the context of a nonexperimental design.

When it has not been possible to randomly assign participants to treatment groups as an approach to controlling an independent variable, the researcher may use a strategy of matching participants for identified variables. For example, in a study of birth weight, pregnant women could be matched with regard to variables such as weight, height, smoking habits, drug use, and other factors that might influence the birth weights of their infants. The independent variable of interest, such as the type of prenatal care, would then be the major difference in the groups. You would then feel more confident that the only difference between the two groups was the differential effect of the independent variable because the other factors in the two groups were theoretically the same. However, you should also remember that other influential variables—such as income, education, and diet—might have been present but were not considered in matching. Threats to internal and external validity represent a major influence on the interpretation of a nonexperimental study because they impose limitations on the generalizability of the results.

If you are critiquing one of the additional types of research discussed, you must first identify the type of research used; then you must understand its specific purpose and format. The format and methods of secondary analysis, methodological research, and meta-analysis vary; knowing how they vary allows you to assess whether the process was applied appropriately. Some of the basic principles of these methods were presented in this chapter. The specific criteria for evaluating these designs are beyond the scope of this text; the references provided can assist you in this process. Even though the format and methods vary, all research has a central goal: to answer questions scientifically.

CRITIQUING CRITERIA

1. Which nonexperimental design is used in the study?
2. In accordance with the theoretical framework, is the rationale for the type of design evident?
3. How is the design congruent with the purpose of the study?
4. Is the design appropriate for the research problem?
5. Is the design suited to the data-collection methods?
6. Does the researcher present the findings in a manner congruent with the design used?
7. Does the researcher theorize beyond the relational parameters of the findings and erroneously infer cause-and-effect relationships between the variables?
8. Are alternative explanations for the findings possible?
9. How does the researcher discuss the threats to internal and external validity?
10. How does the researcher deal with the limitations of the study?

CRITICAL THINKING CHALLENGES

- Discuss which type of nonexperimental design might help validate the defining characteristics of a particular nursing diagnosis you use in practice. Do you think it is possible for nurses and patients to serve as the participants in this type of study?
- The midterm group (five-student) assignment for your research class is to critique an assigned quantitative study. To proceed, you must first

decide the study's overall type. You think it is an ex post facto nonexperimental design, whereas the other students think it is an experimental design because the study has several explicit hypotheses. How would you convince the other students that you are correct?

- You are completing your senior practicum on a surgical step-down unit. The nurses completed an evidence-informed practice protocol for patient-controlled analgesics. Some of the nurses want to implement it immediately, whereas others want to implement it with only some patients. You think that it should be implemented as a research study. Could either of the ways the nurses want to implement the protocol be considered in a research study?
- You are part of a journal club at your hospital. Your group has been examining a phenomenon specific to your patient population and noticed that 20 correlational studies on the topic have been published. Your group decides to perform a meta-analysis of the data. What steps need to be considered in performing the meta-analysis? What level of evidence would you expect to obtain with this method? Explain your answer.
- In reviewing a clinical practice guideline, think about what types of evidence (research, nonresearch) were used in formulating the guideline and whether they are appropriate to the topic. Is there a description of the methods used to grade the evidence? Were the search terms and retrieval methods that were used to acquire research and nonresearch evidence in the guideline clear and relevant? Is the guideline well referenced and comprehensive? Are the recommendations in the guideline sourced according to the level of evidence for its basis?

CRITICAL JUDGEMENT QUESTIONS

1. Which study title suggests a cross-sectional design?
   a. Effect of prenatal parenting classes on infant care knowledge in the early post-partum period
   b. Change in self-esteem over time among women participating in a weight loss support group
   c. Relationship between self-esteem and successful breastfeeding at 1 month and 6 months after birth
   d. Women's appraisal of the diagnosis of breast cancer within the first 48 hours after cancer diagnosis
2. Which title is most suggestive of a longitudinal study?
   a. Effect of prenatal parenting classes on maternal–infant bonding in the early postpartum period
   b. Change in self-esteem over time among women participating in a weight loss support group
   c. Relationship between self-esteem and successful breastfeeding at 1 month, 3 months, and 6 months after birth
   d. Women's appraisal of the diagnosis of breast cancer within the first 48 hours after initial cancer diagnosis
3. In a study of psychosocial adjustment to breast cancer, data collection instruments were sent to the same sample of women at six different times during the first year of living with breast cancer. What type of study design does this exemplify?
   a. Cross-sectional
   b. Retrospective
   c. Longitudinal
   d. Correlational

KEY POINTS

- Nonexperimental research designs are used in studies that make an account of events as they naturally occur. The major difference between nonexperimental and experimental research is that in nonexperimental designs, the independent variable is not actively manipulated by the investigator.
- Nonexperimental designs can be classified either as survey studies or as relationship/difference studies.
- Survey studies and relationship/difference studies are both descriptive and exploratory in nature.

- In survey research, the investigator collects detailed descriptions of existing phenomena and uses the data either to justify current conditions and practices or to make more intelligent plans for improving them.
- In relationship/difference studies, researchers endeavour to explore the relationships or differences between variables in order to provide deeper insight into the phenomena of interest.
- In correlational studies, researchers examine relationships.
- Developmental studies are further divided into categories of cross-sectional, longitudinal (prospective), and retrospective (ex post facto) studies.
- Methodological research, secondary analysis, meta-analysis, epidemiological studies, and clinical practice guidelines are examples of other means of adding to the body of nursing research. Both the researcher and the reader must consider the advantages and disadvantages of each design.
- Nonexperimental research designs do not enable the investigator to establish cause-and-effect relationships between the variables. You must be wary of nonexperimental studies in which researchers make causal claims about the findings, unless a causal modelling technique is used.
- Nonexperimental designs offer the researcher the least amount of control. Threats to validity represent a major influence on the interpretation of a nonexperimental study because they impose limitations on the generalizability of the results and, as such, should be fully assessed by the critical reader.
- The critiquing process is directed toward evaluating the appropriateness of the selected nonexperimental design in relation to factors such as the research problem, the theoretical framework, the hypothesis, the methodology, and the data analysis and interpretation.
- Although nonexperimental designs do not provide the highest level of evidence (level I), they do provide a wealth of data that become useful for formulating both level I and level II studies that are aimed at developing and testing nursing interventions.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Arnaert, A., Ponzoni, N., Debe, Z., Meda, M. M., Nana, N. G., & Arnaert, S. (2019). Experiences of women receiving health-supported antenatal care in the village from community health workers in rural Burkina Faso, Africa. *Digital Health, 10*(3), 57–64. https://doi.org/10.1177/2055207619892756.

Boscart, V. M., Davey, M., Ploeg, J., Heckman, G., Dupuis, S., Sheiban, L., Luh Kim, J., Brown, P., & Sidani, S. (2018). Psychometric evaluation of the Team Member Perspectives of Person-Centered Care (TM-PCC) survey for long-term care homes. *Healthcare, 6*(2), 59. https://doi.org/10.3390/healthcare6020059.

Brouwers, M., Kho, M. E., Browman, G. P., Burgers, J. S., Cluzeau, F., Feder, G., Fervers, B., Graham, I. D., Grimshaw, J., Hanna, S., Littlejohns, P., Makarski, J., & Zitzelsberger, L. (2010). AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *Canadian Medical Association Journal, 182*(18), E839–E842. https://doi.org/10.1503/cmaj.090449.

Castonguay, A., Wrosch, C., Pila, E., & Sabiston, C. (2017). Body-related shame and guilt predict physical activity in breast cancer survivors over time. *Oncology Nursing Forum, 44*(4), 465–475. https://doi.org/10.1188/17.ONF.465-475.

Covell, C. L., Primeau, M-D., Kilpatrick, K., & St-Pierre, I. (2017). Internationally educated nurses in Canada: Predictors of workforce integration. *Human Resources for Health, 15*(1), 1–16. https://doi.org/10.1186/s12960-017-0201-8.

Critical Appraisal Skills Programme (CASP). (2018). Critical appraisal skills programme: Making sense of evidence. https://casp-uk.net/casp-tools-checklists/.

Dahlke, S., Davidson, S., Duarte Wisnesky, U., Kalogirou, M. R., Salyers, V., Pollard, C., Fox, M. T., Hunter, K. F., & Baumbusch, J. (2019). Student nurses' perceptions about older people. *International Journal of Nursing Education Scholarship*(1), 16. https://doi.org/10.1515/ijnes-2019-0051.

Dall'Ora, C, Ball, J, Redfern, O, et al. (2019). Are long nursing shifts on hospital wards associated with sickness absence? A longitudinal retrospective

observational study. *Journal of Nursing Management, 27*, 19–26. https://doi-org.ezproxy.library.yorku.ca/10.1111/jonm.12643.

Della Pelle, C., Cerratti, F., Di Giovanni, P., Cipollone, F., &. Cicolini, G. (2018). Attitudes Towards and Knowledge About Lesbian, Gay, Bisexual, and Transgender Patients Among Italian Nurses: An Observational Study. *Journal of Nursing Scholarship, 50*(4), 367–374. https://doi.org/10.1111/jnu.12388.

Dennis, C., Merry, L., & Gagnon, A. (2017). Postpartum depression risk factors among recent refugee, asylum-seeking, non-refugee immigrant, and Canadian-born women: Results from a prospective cohort study. *Social Psychiatry and Psychiatric Epidemiology, 52*(4), 411–422. https://doi.org/10.1007/s00127-017-1353-5.

Fernandez, S., Franklin, J., Amlani, N., DeMilleVille, C., Lawson, D., & Smith, J. (2015). Physical activity and cancer: A cross-sectional study on the barriers and facilitators to exercise during cancer treatment. *Canadian Oncology Nursing Journal = Revue Canadienne de Nursing Oncologique, 25*(1), 37–48. https://doi.org/10.5737/236880762513742.

Fox, M., McCague, H., Sidani, S., & Butler, J. (2018). The relationships between the geriatric practice environment, nursing practice, and the quality of hospitalized older adults' care. *Journal of Nursing Scholarship, 50*(5), 513–521. https://doi.org/10.1111/jnu.12414.

Freeman, L. A., Pfaff, K. A., Kopchek., L, & Liebman, J (2020). Investigating palliative care nurse attitudes towards medical assistance in dying: An exploratory cross-sectional study. *Journal of Advanced Nursing, 76*, 535–545. https://doi.org/10.1111/jan.14252.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0*. Retrieved from http://handbook.cochrane.org.

Jin, C., Lobchuk, M., Chernomas, W., & Pooyania, S. (2017). Examining associations of functional deficits and mood states with empathic responses of stroke family caregivers. *Journal of Neuroscience Nursing, 49*(1), 12–14. https://doi.org/10.1097/JNN.0000000000000250.

Jones, D., Letourneau, N., & Duffett-Leger, L. (2019). Systematic review of predictors of infant care competence among women with postpartum depression. *Journal of Nursing Education and Practice, 9*(5), 118–128. https://doi.org/10.5430/jnep.v9n5p118.

Leyva-Moral, J. M., Dominguez-Cancino, K. A., Guevara-Vasquez, G. M., Edwards, J. E., & Palmieri, P. A. (2019). Faculty attitudes about caring for people living with HIV/AIDS: a comparative study. *Journal of Nursing Education, 58*(12), 712–717.

Pien, L.-C., Cheng, Y., & Cheng, W.-J. (2019). Psychosocial safety climate, workplace violence and self-rated health: A multi-level study among hospital nurses. *Journal of Nursing Management, 27*, 584–591. https://doi-org.ezproxy.library.yorku.ca/10.1111/jonm.12715.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. (Vol. 1). Sage.

Rayan, A. (2019). Mindfulness, self-efficacy, and stress among final-year nursing students. *Journal of Psychosocial Nursing, 57*(4), 49–55. https://doi.org/10.3928/02793695-20181031-01.

Richards, D. A., Hilli, A., Pentecost, C., Goodwin, V. A., & Frost, J. (2018). Fundamental nursing care: A systematic review of the evidence on the effect of nursing care interventions for nutrition, elimination, mobility and hygiene. *Journal of Clinical Nursing, 27*(11-12), 2179–2188. https://doi.org/10.1111/jocn.14150.

Rochefort, C. M., Rathwell, B. A., & Clarke, S. P. (2016). Rationing of nursing care interventions and its association with nurse-reported outcomes in the neonatal intensive care unit: A cross-sectional survey. *BMC Nursing, 15*, 46. https://doi.org/10.1186/s12912-016-0169-z.

Rose, M. S., Pana, G., & Premji, S. (2016). Prenatal maternal anxiety as a risk factor for preterm birth and the effects of heterogeneity on this relationship: A systematic review and meta-analysis. *BioMed Research International, 2016*, Article 8312158. https://doi.org/10.1155/2016/8312158.

Schofield, R., Forchuk, C., Montgomery, P., Rudnick, A., Edwards, B., Meier, A., & Speechley, M. (2016). Comparing personal health practices: Individuals with mental illness and the general Canadian population. *Canadian Nurse, 112*(5), 23–27.

Scott, R., Lasiuk, G., & Norris, C. (2016). Sexual orientation and depression in Canada. *Canadian Journal of Public Health, 107*(6), e545–e549. https://doi.org/10.17269/CJPH.107.5506.

Sidani, S., Epstein, D. R., Fox, M., & Collins, L. (2018). The contribution of participant, treatment, and outcomes factors to treatment satisfaction. *Research in Nursing and Health, 41*(6), 572–582. https://doi.org/10.1002/nur.21909.

Stephenson, E., DeLongis, A., Steele, R., et al. (2017). Siblings of children with a complex chronic health condition: Maternal posttraumatic growth as a

predictor of changes in child behavior problems. *Journal of Pediatric Psychology, 42*(1), 104–113.

Strazzieri-Pulido, K. C., Carol, C. V., Nogueira, P. C., Padilha, K. G., & Vera, V. L. C. (2019). Pressure injuries in critical patients: Incidence, patient-associated factors, and nursing workload. *Journal of Nursing Management, 27*(2), 301–310. https://doi.org/10.1111/jonm.12671.

Vincelette, C., Lavoie, S., Fortin, O., & Quiroz-Martinez, H. (2018). Intensive care unit nurses' knowledge, skills and attitudes regarding three resuscitation procedures: A cross-sectional survey. *Canadian Journal of Critical Care Nursing, 29*(4), 29–35.

Wall, V., Premji, S. S., Letourneau, N., McCaffrey, G., & Nyanza, E. C. (2018). Factors associated with pregnancy-related anxiety in Tanzanian women: A cross sectional study. *BMJ Open, 8*(6), 1–8. https://doi.org/10.1136/bmjopen-2017-020056.

Winsett, R. P., Rottet, K., Schmitt, A., Wathen, E., & Wilson, D. (2016). Medical surgical nurses describe missed nursing care tasks—Evaluating our work environment. *Applied Nursing Research, 32*, 128–13388. https://doi.org/10.1016/j.apnr.2016.06.006.

PART FIVE

Processes Related to Research

RESEARCH **VIGNETTE**

Nursing Workforce

Mélanie Lavoie-Tremblay, BScN, MScN, PhD
Associate Professor
Ingram School of Nursing
McGill University
Montréal, Quebec

Though university nursing programs do their best to provide adequate instruction in theory and practice to their students, it is well documented that newly graduated nurses (NGNs) still feel unprepared for their transition into the work environment, which contributes to an alarming number of new nurses quitting the profession (Duchscher, 2009; Lavoie-Tremblay, O'Brien-Pallas, Gélinas, Desforges, & Marchionni, 2008; Marleau & Lapointe, 2014). Healthcare organizations have been facing this challenge for many years, which has spurred the development of several interventions, such as mentorship programs, to help NGNs acclimate to the clinical environment. However, even with these interventions in place, when NGNs begin to work independently, they continue to experience multifactorial challenges when providing nursing care, be they professional, organizational, or personal in nature (Kelly & Ahern, 2009; Murray, Sundin, & Cope, 2019). As such, it is important to study the experience of new nurses in order to find what can be done to ease the student-to-nurse role transition, both in the educational environment before graduation and in the workplace after graduation.

As a new nurse, I found the working conditions difficult. There was a significant gap between what I learned at school and the reality of the workplace, and I was asking myself if I should stay or quit the nursing profession. I began to realize that I was not alone in thinking this way, as the other new nurses working alongside me also found the start of their careers to be extremely challenging. I was fortunate to have discussed this with a clinical nurse specialist at the hospital where I was working, who suggested that I help them review their orientation program to help new nurses have a better transition at work. This is why I started my Master's in nursing, as part of an effort to modernize their orientation program for new nurses. With the leader of the nursing team, we completed a participatory action research study and were able to improve the orientation program based on new nurses' needs. This work eventually led to the publication of my first manuscript in 2002, entitled "How to facilitate the orientation of new nurses into the workplace."

After finishing my Master's program, I realized that we can increase retention of new nurses in their first years of work by having a healthier workplace. This led me to undertake my PhD with a participatory action research program, in order to optimize the workplace and reduce psychological distress and absenteeism among nurses. This research gave me a better understanding of key concepts linked to being healthy at work, such as rewards, social support, autonomy, and empowerment. In order to foster a healthy work environment, it is also imperative to understand the nature of the constraints in the workplace, and to involve nurses and other healthcare professionals in proposing solutions.

Since I began my career as a professor at McGill University in 2005, I have had the opportunity to collaborate with several health organizations on projects aimed at improving the working environment for nurses. In 2017, as chair of the McGill Nursing Collaborative for Education and Innovation in Patient- and Family-Centred Care, I helped to establish the *Nightingale Fellows Project*, a group mentorship experience aimed at helping graduating students have an easier transition into clinical practice. This program, which is offered to all McGill nursing students in their final year of study, has two main goals: to provide educational and experiential support to students as they prepare to pass from student to nurse, and to provide a safe environment where students can discuss this process with experienced clinicians, thereby increasing their understanding of the knowledge and skills they will need to succeed as graduate nurses working in a clinical unit. Each academic year, students who participate in the Nightingale Fellows Project are placed into small groups of six to eight students, each of which is assigned to one of nine nurse mentors. These mentors, known as "Nightingale Fellows," are all experienced nurses currently working in hospitals in Montreal. Each group of students then meets with their mentors four times during the

spring semester and once more the following fall, and at each meeting they discuss topics related to entering the workforce, such as preparing for their licensure exam, finding a job after graduation, managing workload and stress, and adjusting to organizational culture. Now in its fourth year of existence, this program has helped over 200 students with their transition from student to nurse (Lavoie-Tremblay, Sanzone, Primeau & Lavigne, 2019).

IN SUMMARY

Several initiatives can be implemented to promote attraction and retention of new nurses, to ensure a smooth transition from student to nurse, and to create healthy workplaces for nurses and patients. It is crucial to involve students, nurses, and patients in creating healthy workplaces, so that we can co-construct solutions that are suitable and sustainable in the long-term. ■

REFERENCES

Duchscher, J. E. B. (2009). Transition shock: The initial stage of role adaptation for newly graduated registered nurses. *Journal of Advanced Nursing, 65*(5), 1103–1113. https://doi.org/10.1111/j.1365-2648.2008.04898.x.

Kelly, J., & Ahern, K. (2009). Preparing nurses for practice: A phenomenological study of the new graduate in Australia. *Journal of Clinical Nursing, 18*(6), 910–918.

Lavoie-Tremblay, M., O'Brien-Pallas, L., Gélinas, C., Desforges, N., & Marchionni, C. (2008). Addressing the turnover issue among new nurses from a generational viewpoint. *Journal of Nursing Management, 16*(6), 724–733. https://doi.org/10.1111/j.1365-2934.2007.00828.x.

Lavoie-Tremblay, M., Sanzone, L., Primeau, G., & Lavigne, G. L. (2019). Group mentorship programme for graduating nursing students to facilitate their transition: A pilot study. *Journal of Nursing Management, 27*(1), 66–74. https://doi.org/10.1111/jonm.12649.

Marleau, D., & Lapointe, J. (2014). *Portrait de la relève infirmière: 2013–2014*. (Report No. 293S (13-14)). Quebec: Ordre des infirmières et infirmiers du Quebec.

Murray, M., Sundin, D., & Cope, V. (2019). Benner's model and duchscher's theory: Providing the framework for understanding new graduate nurses' transition to practice. *Nurse Education in Practice, 34*, 199–203. https://doi.org/10.1016/j.nepr.2018.12.003.

CHAPTER 13

Sampling

Mina D. Singh | Ramesh Venkatesa Perumal

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Identify the purpose of sampling.
- Define population, sample, and sampling.
- Compare a population and a sample.
- Discuss the eligibility criteria for sample selection.
- Define nonprobability sampling and probability sampling.
- Identify the types of strategies for both nonprobability and probability sampling.
- Identify the types of qualitative sampling.
- Compare the advantages and disadvantages of specific nonprobability and probability sampling strategies.
- Discuss the contribution of nonprobability and probability sampling strategies to the strength of evidence provided by study findings.
- Discuss the factors that influence determination of sample size.
- Discuss the procedure for drawing a sample.
- Identify the criteria for critiquing a sampling plan.
- Use the critiquing criteria to evaluate the "Sample" section of a research report.

KEY TERMS

accessible population
cluster sampling
convenience sampling
data saturation
delimitations
effect size
element
eligibility criteria
heterogeneity
homogeneous
matching
multistage sampling
network sampling
nonprobability sampling
pilot study
population
probability sampling
purposive sampling
quota sampling
random selection
representative sample
sample
sampling frame
sampling interval
sampling unit
simple random sampling
snowball effect sampling
stratified random sampling
systematic sampling
target population
theoretical sampling

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

SAMPLING IS THE PROCESS OF SELECTING representative units of a population for study in a research investigation. Although sampling is a complex process, it is a familiar one. In their daily lives, people gather knowledge, make decisions, and formulate predictions on the basis of sampling procedures. For example, nursing students may make generalizations about the overall quality of nursing professors as a result of their exposure to a sample of nursing professors during their undergraduate programs. Patients may make generalizations about a hospital's food or quality of nursing care during a 1-week hospital stay. Limited exposure to a limited portion of these phenomena forms the basis of people's conclusions, so much of their knowledge and many of their decisions are based on their experience with samples.

Researchers also derive knowledge from samples. Many questions in scientific and naturalistic research cannot be answered without the use of sampling procedures. For example, when the effectiveness of a new education intervention for diabetic patients is tested, the intervention is administered to a sample of the population. The researcher must come to some conclusions without giving the intervention to the entire population of diabetic patients. To obtain the experiences or outcomes of engaging in this education, the researcher needs to select the appropriate sampling strategy in accordance with the research design and question. This is done to avoid erroneous conclusions or making generalizations from a nonrepresentative sample. Thus, research methodologists have expended considerable effort to develop sampling theories and procedures that produce accurate and meaningful information. Essentially, researchers sample representative segments of the population because sampling the entire population of interest to obtain relevant information is rarely feasible or necessary.

This chapter will familiarize you with the basic concepts of sampling as they pertain to the principles of quantitative and qualitative research designs, nonprobability and probability sampling, sample size, and the related critiquing process.

SAMPLING CONCEPTS

Population

A **population** is a well-defined set that has certain specified properties or characteristics from which data can be gathered and analyzed. A population can be composed of people, animals, objects, or events. For example, if a researcher is studying undergraduate nursing students, the type of educational preparation of the population must be specified. In this example, the population consists of undergraduate students enrolled in a generic baccalaureate nursing program. Examples of other possible populations might be all female patients admitted to a certain hospital for lumpectomies for treatment of breast cancer during 2016, all children with asthma in the province of Alberta, or all men and women with a diagnosis of schizophrenia in North America. These examples illustrate that a population may be broadly defined and potentially involve millions of people, or it may be narrowly specified to include only a few people. The target population is the entire population of interest, while the accessible population is the portion of the target population that is available to the researcher.

When you read a research report, you should consider whether the researcher has identified the population descriptors that form the basis for the inclusion (eligibility) or exclusion criteria

(delimitations) that are used to select the sample from the array of all possible units, whether people, objects, or events. Consider the population previously defined as undergraduate nursing students enrolled in a generic baccalaureate program. Would this population include both part-time and full-time students? Would it include students who had previously attended another nursing program? What about international students? At which level (first year through senior year) would students qualify? As much as possible, the researcher must specifically delineate the exact criteria used to decide whether an individual would be classified as a member of a given population. The population descriptors that provide the basis for inclusion (eligibility) criteria should be evident in the sample; in other words, the characteristics of the population and the sample should be congruent. The degree of congruence is evaluated to assess the representativeness of the sample. For example, if a population is defined as full-time, Canadian-born, senior-level nursing students enrolled in a generic baccalaureate nursing program, the sample would be expected to reflect these characteristics.

Think about the concept of inclusion criteria, or **eligibility criteria** (characteristics of a population that meet requirements for inclusion in a study), applied to a research study in which the participants are patients. For example, in an investigation of the effects of music on dyspnea during exercise in individuals with chronic obstructive pulmonary disease (COPD), the participants had to meet all of the following inclusion (eligibility) criteria:

1. A confirmed medical diagnosis of COPD (i.e., chronic bronchitis, emphysema, or both)
2. Ability to speak and read English
3. Ability to ambulate independently
4. Experiencing dyspnea at least once a week
5. An increase in the level of dyspnea of at least two points on the Borg scale after a 6-minute walk

In their study to better understand the unique narratives of social exclusion for mothers experiencing homelessness, Benbow et al. (2019) had the following inclusion criteria: the participants had to (a) self-identify as mothers (with or without physical custody of their child(ren), (b) self-identify as experiencing homelessness currently or within the last year, and (c) be aged 18 years and older.

Examples of exclusion criteria, or **delimitations** (characteristics that restrict the population to a homogeneous group of participants), include gender, age, marital status, socioeconomic status, religion, ethnicity, level of education, age of children, health status, and diagnosis. In a study examining and comparing risk factors for postpartum depression among recent refugee, asylum-seeking, non-refugee immigrant, and Canadian-born women, the exclusion criteria were if women (1) had a major mental illness (schizophrenia, other psychoses, or profound, previously existing depression) or cognitive impairment that precluded informed consent; (2) were visitors to Canada; (3) had given the infant up for adoption; or (4) had a stillbirth delivery or infant death (p. 413).

As another example, in a randomized control trial, Jackson and Dennis (2017) in Appendix B had strict inclusion and exclusion criteria to study the effects of applying lanolin for the treatment of nipple pain in breastfeeding. Inclusion criteria were breastfeeding women who: (1) had nipple pain and visible nipple damage; (2) delivered a full-term (greater than 37 weeks), singleton infant within the previous 72 hours; and (3) were able to speak English. The exclusion criteria included: (1) infants not expected to be discharged home with their mother; (2) infants with a congenital abnormality that would impair breastfeeding, or ankyloglossis (tongue-tie); (3) maternal allergy to lanolin; (4) maternal health condition(s) that might interfere with breastfeeding; and (5) maternal aversion to, or, strong desire to use lanolin (p. 2).

The **heterogeneity**, or dissimilarities, of a sample group inhibits the researchers' ability to interpret the findings meaningfully and to make

generalizations. It is much wiser to study only one **homogeneous** group—that is, a group with limited variation in attributes or characteristics, or to include specific groups as distinct subsets of the sample and study the groups comparatively, as was the case in Dennis et al. (2017) study on examining risk factors for postpartum depression among recent refugee, asylum-seeking, non-refugee immigrant, and Canadian-born women. They found that recent migrant women had significantly higher rates of depression at 16 weeks postpartum than Canadian-born women (p. 411).

Remember that exclusion criteria or delimitations are not established in a casual or meaningless way but are established to control for extraneous variability or bias. Each exclusion criterion should have a rationale, presumably related to a potential contaminating effect on the dependent variable. Carefully established sample exclusion criteria increase the precision of the study and contribute to accuracy while constraining the generalizability or transferability of the findings (see Chapter 10).

The population criteria establish the **target population**—that is, the entire set of cases about which the researcher would like to make generalizations. A target population might include all undergraduate nursing students enrolled in generic baccalaureate programs in Canada. Because of time, money, and personnel, however, using a target population is often not feasible. An **accessible population**—one that meets the population criteria and is available—is used instead. For example, an accessible population might include all full-time generic baccalaureate students attending school in Manitoba. Pragmatic factors must also be considered in identifying a potential population of interest.

Research Hint

Often, researchers do not clearly identify the population under study, or the population is not clarified until the "Discussion" section, when an effort is made to discuss the group (population) to which the study findings can be generalized.

A population is not restricted to human participants. The population may consist of hospital records; blood, urine, or other specimens taken from patients at a clinic; historical documents; or laboratory animals. For example, a population might consist of all urine specimens collected from patients in the Mount Sinai Hospital antepartum clinic or all patient charts on file at a day surgery centre. A population can be defined in a variety of ways. Of importance is that the basic unit of the population be clearly defined, because the generalizability of the findings is a function of the population criteria.

Evidence-Informed Practice Tip

Consider whether the sample selection was biased, thereby influencing the validity of the evidence provided by the outcomes of the study.

Samples and Sampling

Sampling is a process of selecting a portion or subset of the designated population to represent the entire population. A **sample** is a set of elements that make up the population; an **element** is the most basic unit about which information is collected. A sampling frame is another name for the list of elements from which the sample will be chosen from. The most common element in nursing research is individuals, but other elements (e.g., places or objects) can form the basis of a sample or population. For example, a researcher plans a study to compare the effectiveness of different nursing interventions on reducing falls in older adults in long-term care facilities. Four facilities, each of which having a different treatment protocol, are identified as the sampling units—not the nurses themselves or the treatment alone. A sampling unit can be an organization, a group, or an individual person.

The purpose of sampling is to increase the efficiency of a research study. Examining every element or unit in the population would not be feasible. When sampling is done properly, the

researcher can draw inferences and make generalizations about the population without examining each unit in the population.

In qualitative research, the results can also have good generalizability to the population under study. Sampling procedures that entail the formulation of specific criteria for selection ensure that the characteristics of the phenomena of interest will be, or are likely to be, present in all of the elements being studied. The researcher's efforts to ensure that the sample is representative of the target population provide a stronger position from which to draw conclusions from the sample findings that are generalizable to the population (see Chapter 10).

After reviewing a number of research studies, you will recognize that samples and sampling procedures vary in terms of merit. The foremost criterion in evaluating a sample is its representativeness. A **representative sample** has key characteristics that closely approximate those of the population. For instance, if 70% of the population in a study of child-rearing practices consisted of women and 40% were full-time employees, a representative sample should reflect these characteristics in the same proportions.

The representativeness of a sample cannot be guaranteed without access to a database about the entire population. Because it is difficult and inefficient to assess an entire population, the researcher must employ sampling strategies that minimize or control for sample bias. If an appropriate sampling strategy is used, the sample data will almost always enable a reasonably accurate understanding of the phenomena under investigation.

Evidence-Informed Practice Tip

Determining whether the sample is representative of the population being studied in journal articles will influence both your interpretation of the evidence provided by the findings and your decision-making about the findings' relevance to the your patient population and practice setting.

SAMPLING STRATEGIES USED IN QUANTITATIVE RESEARCH

Sampling strategies are generally grouped into two categories: *nonprobability sampling* and *probability sampling*. In **nonprobability sampling**, elements are chosen through nonrandom methods. The drawback of this strategy is that each element's probability of being included in the samples cannot be estimated. In other words, ensuring that every element has a chance for inclusion in the nonprobability sample is not possible. In **probability sampling**, some form of random selection is used when the sample units are chosen. This type of sample enables the researcher to estimate the probability that each element of the population will be included in the sample. Probability sampling is the more rigorous sampling strategy used in quantitative research and is more likely to result in a representative sample.

The remainder of this section is devoted to a discussion of different types of nonprobability and probability sampling strategies. A summary of sampling strategies appears in Table 13.1. You may refer to this table as the various nonprobability and probability strategies are discussed in the following sections. Note that if there is bias in sampling, it will distort the analysis and the findings of the study.

Research Hint

Research articles are not always explicit about the type of sampling strategy that was used. If the sampling strategy is not specified, assume that in a quantitative study, a convenience sample was used and that in a qualitative study, a purposive sample was used.

Nonprobability Sampling

Because of a lack of random selection, the nonprobability sampling strategy is less generalizable than probability sampling because it tends to produce less representative samples. Such samples are more feasible for the researcher to obtain, however, and most samples—in nursing

TABLE 13.1

SUMMARY OF SAMPLING STRATEGIES

SAMPLING STRATEGY	EASE OF DRAWING A REPRESENTATIVE SAMPLE	RISK OF BIAS	REPRESENTATIVENESS OF THE SAMPLE
NONPROBABILITY			
Convenience	Very easy	Greater than in any other sampling strategy	Because samples tend to be self-selecting, representativeness is questionable
Quota	Relatively easy	Contains an unknown source of bias that affects external validity	Builds in some representativeness by using knowledge about the population of interest
Purposive	Relatively easy	Bias increases with greater heterogeneity of the population; conscious bias is also a danger but is offset with maximal variation	Very limited ability to generalize because the sample is handpicked from a quantitative view, but this approach is necessary for the qualitative researcher to choose participants on the basis of the phenomenon under study
Network	Can be easy if the network is accessible	Minimal if a thorough sampling plan is developed	Represents the event, incident, or experience being studied
Theoretical	Requires a two-stage process; can be prolonged	Minimal if a thorough sampling plan is developed	Typically begins with another type of sampling, such as convenience or criterion sampling aimed at variation in the phenomenon, and thus represents aspects of the theory being constructed
PROBABILITY			
Simple random	Laborious	Low	Maximized; the probability of nonrepresentativeness decreases with increased sample size
Stratified random	Time-consuming	Low	Enhanced
Cluster	Less time-consuming than simple or stratified sampling	Subject to more sampling errors than is simple or stratified sampling	Less representative than simple or stratified sampling
Systematic	More convenient and efficient than is simple, stratified, or cluster sampling	Bias in the form of nonrandomness can be inadvertently introduced	Less representative if bias occurs as a result of coincidental nonrandomness

research and the research of other disciplines—are nonprobability samples. When a nonprobability sample reflects the target population through the careful use of inclusion and exclusion criteria, you can have more confidence in the representativeness of the sample and the external validity of the findings. The major types of nonprobability sampling used in quantitative research are *convenience sampling* and *quota sampling*.

Convenience Sampling

Convenience sampling is the use of the most readily accessible persons or objects as participants in a study. The participants may include volunteers, the first 25 patients admitted to a certain hospital with a particular diagnosis, all of the people who enrolled in a certain program during the month of September, or all of the students enrolled in a certain course at a particular university during 2021. The participants are convenient and accessible to the researcher; hence the term *convenience sample*.

As an example, Dosani et al. (2017) recruited a convenience sample of 122 mothers to study the experiences of mothers breastfeeding the late preterm infant and perceptions of public health nurses. In another study, Chan et al. (2019) used a convenience sample of 69 nursing students to investigate the influence nursing students' self-efficacy (confidence) related to medication administration and medication errors using an electronic administration record in clinical simulation.

The advantage of a convenience sample is that it can be an easy way for the researcher to obtain participants. The researcher may need to be concerned only with obtaining a sufficient number of participants who meet the same criteria. The major disadvantage of a convenience sample is that the risk of bias is greater than in any other type of sample (see Table 13.1). Because convenience samples entail voluntary participation, the probability that researchers will recruit people who feel strongly about the issue being studied is increased, which may favour certain outcomes of the study. The problem of bias is related to the tendency of convenience samples to be self-selecting; in other words, the researcher obtains information only from the people who volunteer to participate. In this case, the following questions must be raised:

- What motivated some of the people to participate and others not to participate?
- What kind of data would have been obtained if nonparticipants had also responded?
- How representative of the population are the people who did participate?

For example, a researcher may stop people on a street corner to ask their opinion on an issue; place advertisements in the newspaper; put signs in local churches, community centres, or supermarkets; or search specific agencies websites to recruit volunteers for a particular study. To study the influence of visible physical signs on caregiver's patient-centred and empathetic behaviours in chronic pain, Paul-Savoie et al. (2018) used a convenience sample of 21 nurses and 21 physicians recruited through advertisements and referrals. To assess the degree to which a convenience sample approximates a random sample, a researcher can compare the convenience sample data with the known demographic information and examine variability around the mean. In this manner, the researcher checks for the representativeness of the convenience sample and the extent to which bias is or is not evident.

Because recruiting research participants is crucial for nurse researchers, innovative recruitment strategies are sometimes used. For example, a researcher may offer to pay the participants for their time. A relatively new method of accessing and recruiting participants is through online computer networks (e.g., disease-specific chat rooms and bulletin boards).

In evaluating a research report, you should recognize that the convenience sample strategy, although the most common, is the weakest form of sampling strategy in quantitative research in terms of generalizability. When a convenience sample is used, researchers should analyze and interpret the data cautiously. When you critique a research study in which this sampling strategy was used, you should be skeptical about the external validity of the findings (see Chapter 10).

Quota Sampling

Quota sampling refers to a form of nonprobability sampling in which knowledge about the population of interest is used to ensure some representativeness about the sample (see Table 13.1). Through quota sampling, the researcher

identifies a particular strata of the population, and the quota sample proportionally represents the strata. For example, the data in Table 13.2 reveal that of the 5,000 nurses in a particular city, 20% are diploma graduates, 40% are post–RN degree graduates, and 40% are baccalaureate graduates. Each of these strata should be proportionately represented in the sample. In this case, the researcher used a proportional quota sampling strategy and decided to include 10% of a population of 5,000 (i.e., 500 nurses). On the basis of the proportion of each stratum in the population, 100 diploma graduates, 200 post-RN graduates, and 200 baccalaureate graduates were the quotas established for the three strata. The researcher recruited participants who met the eligibility criteria of the study until the quota for each stratum was filled. In other words, once the researcher obtained the necessary 100 diploma graduates, 200 post-RN graduates, and 200 baccalaureate graduates, the sample was complete with regard to both the research design and other pragmatic matters, such as economy.

The researcher systematically ensures that proportional segments of the population are included in the sample. For example, in Im et al. (2012) study exploring midlife women's attitudes toward physical activity, the researchers stratified a quota sample of 542 subjects by ethnicity and socioeconomic status. An example of nonproportional quota sampling is in the study by Fox et al. (2010), who examined differences in sleep complaints among adults with varying amounts of bed rest who were residing in extended-care facilities for chronic disease management. The three cohorts (comparative, moderate, and high) reflected different amounts of bed rest that were naturally occurring. To ensure equal representation of the different amounts of bed rest, nonproportional quota sampling was used.

The characteristics chosen to form the strata are selected according to a researcher's judgement on the basis of knowledge of the population and the literature review. The criterion for selection should be a variable that reflects important differences in the independent variables under investigation. Age, gender, religion, ethnicity, medical diagnosis, socioeconomic status, level of completed education, and occupation are among the variables that are likely to be important in stratifying samples in nursing research investigations.

In critiquing a research strategy, you need to determine whether the sample strata appropriately reflect the population under consideration and whether the variables used are homogeneous enough to ensure a meaningful comparison. Even when the researcher has addressed these factors, you must remember that a quota strategy is a nonprobability sample and thus includes an unknown source of bias that affects the external validity. The people who choose to participate may not be typical of the population in terms of the variables being measured, and assessing the possible biases that may be operating is not possible. When the phenomena being investigated are relatively similar within the population, the risk of bias may be minimal; however, in heterogeneous populations, the risk of bias is greater.

TABLE 13.2

NUMBERS AND PERCENTAGES OF STUDENTS IN STRATA OF A QUOTA SAMPLE OF 5,000 GRADUATES OF NURSING PROGRAMS IN A PARTICULAR CITY

CATEGORIES	DIPLOMA GRADUATES	PRACTICAL NURSES	BACCALAUREATE GRADUATES
Strata	1000 (20%)	2000 (40%)	2000 (40%)
Quota sample	100	200	200

Evidence-Informed Practice Tip

When you think about applying study findings to your clinical practice, consider whether the participants in the sample are similar to your own patients.

Probability Sampling

The primary characteristic of probability sampling is the random selection of elements from the population. In **random selection**, each element of the population has an equal and independent chance of being included in the sample. In the hierarchy of evidence, probability sampling represents the strongest type of sampling strategy. That means there is greater confidence that the sample is representative rather than biased and that it more closely reflects the characteristics of the population of interest. Nevertheless, there will always be differences between the sample and the population; this difference is called sampling error. Four commonly used probability sampling strategies are *simple random sampling, stratified random sampling, cluster sampling*, and *systematic sampling*.

Random selection of sample participants should not be confused with random assignment of participants. As discussed in Chapter 11, *randomization* refers to the assignment of participants to either an experimental or a control group on a purely random basis.

Simple Random Sampling

Simple random sampling is a laborious and carefully controlled process. Because the principles of simple random sampling are incorporated in the more complex probability designs, the principles of this strategy are presented.

In **simple random sampling**, the researcher defines the population (a set), lists all units of the population (a **sampling frame**), and selects a sample of units (a subset) from which the sample will be chosen. For example, if Canadian hospitals specializing in the treatment of cancer were the sampling unit, a list of all such hospitals would be the sampling frame. If certified adult nurse practitioners constituted the accessible population, a list of those nurses would be the sampling frame.

Once a list of the population elements has been developed, the best method of selecting a sample is to employ a table of random numbers containing columns of digits, as shown in Figure 13.1. Such tables can be generated by computer programs. For example, Tryphonopoulos and Letourneau (2020) tested a video-feedback interaction guidance intervention designed to improve maternal–infant interaction, depressive symptoms, and cortisol patterns of depressed mothers and their infants. Participants were allocated into either the intervention or control group using an online randomization program (http://www.random.org/sequences/) with the number 1 denoting intervention group placement and the number 2 denoting control group placement. Once pre-test data were collected, the program was used to generate one of these two integers randomly, and subsequent group placement was assigned accordingly (pp. 3–4). The system generated a random blocking table, which ensured even distribution of the control and experimental group in all four sites of the study. After assigning consecutive numbers to units of the population, the researcher starts at any point on the table of random numbers and reads consecutive numbers in any direction (i.e., horizontally, vertically, or diagonally). When a number is read that corresponds with the written unit on a card, that unit is chosen for the sample. The investigator continues to read until a sample of the desired size is drawn. As an example, Henrique et al. (2018) randomly allocated 128 patients into one of the following interventions groups: warm shower hydrotherapy, perineal exercises with a ball, and the combination of interventions of warm shower hydrotherapy with perineal exercises with a ball (p. 2).

The advantages of simple random sampling are as follows:

- The sample selection is not subject to the conscious biases of the researcher.

1000 random integers between 0 and 99

40	23	0	29	10	94	17	58	12	85	13	25	80	84	72	74	54	63	55	31
32	98	59	23	74	97	51	42	21	87	48	64	54	38	84	68	14	17	35	48
84	34	84	14	53	65	67	37	2	45	84	21	71	34	10	80	72	27	11	13
86	37	24	89	23	4	44	40	72	81	44	69	25	44	34	34	34	75	50	50
50	58	85	8	22	24	73	20	63	35	60	87	91	92	96	80	19	22	87	24
1	87	43	82	9	31	40	88	33	28	82	73	18	6	48	64	59	45	34	3
21	19	42	76	84	67	29	68	8	66	93	89	96	28	12	14	38	47	52	65
32	66	33	21	81	97	39	76	67	27	97	22	76	89	41	11	91	29	6	66
16	82	42	75	35	42	92	90	77	24	21	8	36	16	5	54	89	51	57	85
74	32	63	65	93	96	18	36	82	72	39	69	37	97	51	17	36	71	38	30
50	94	4	66	17	37	10	53	8	29	67	74	88	38	11	59	60	91	56	17
71	47	81	18	53	98	7	87	29	37	22	93	13	6	95	7	95	71	14	6
71	93	48	16	33	19	46	21	60	44	52	91	52	58	10	9	41	31	35	18
20	94	13	99	45	6	53	54	1	25	79	28	1	48	36	26	68	37	59	7
75	22	69	56	62	40	64	45	40	99	94	14	98	84	22	38	24	87	43	71
16	87	41	0	88	83	11	37	71	78	22	39	43	37	75	84	84	11	55	58
92	90	80	2	30	37	84	55	56	50	3	71	24	13	62	74	82	44	90	32
96	89	31	32	37	45	70	67	80	55	58	9	55	60	61	55	86	44	27	77
38	29	36	94	65	39	56	29	29	65	88	13	71	38	71	8	81	66	31	44
20	6	61	66	90	13	70	60	92	53	87	49	34	42	14	47	75	33	26	9
63	44	94	21	14	13	41	80	39	72	29	3	25	89	44	88	13	49	18	58
13	32	93	90	31	75	86	95	18	51	61	59	84	95	67	54	40	30	29	63
26	35	48	81	19	24	36	36	76	16	46	5	93	41	97	46	79	54	95	49
89	74	96	95	94	69	31	60	16	69	76	42	28	71	69	34	46	55	20	42
50	39	28	64	20	68	60	33	92	82	61	70	5	68	95	88	12	85	18	94
55	86	5	96	87	69	75	93	54	79	0	57	45	8	86	59	25	21	9	29
75	35	1	2	86	62	70	83	85	13	97	37	13	73	16	38	36	23	54	11
74	50	1	77	87	92	68	87	57	36	17	47	0	97	78	72	72	45	54	51
34	24	35	13	26	42	22	75	47	2	34	87	15	50	65	27	5	72	28	68
73	33	42	65	91	24	44	84	71	55	70	1	27	30	8	61	65	61	18	92
7	55	12	6	61	17	23	95	91	58	60	30	35	61	34	27	75	44	35	64
10	94	18	4	3	19	21	37	28	55	76	25	10	29	80	64	8	81	20	32
20	48	92	87	95	58	57	73	42	1	12	81	94	85	63	97	24	19	93	51
81	10	92	49	70	15	76	4	36	92	62	99	78	32	86	74	43	22	98	46
66	67	82	94	67	75	16	88	84	98	0	52	37	0	43	9	0	51	2	62
84	92	36	11	3	52	44	65	45	67	97	86	92	2	50	5	93	66	73	40
36	29	98	46	88	23	28	44	8	71	69	43	53	16	87	21	56	23	37	24
15	11	82	30	59	94	23	30	40	25	87	26	24	30	44	53	33	65	72	55
89	57	49	79	83	88	42	45	41	93	38	24	15	80	97	18	61	12	13	42
23	36	65	9	64	26	93	37	26	44	42	17	45	68	27	77	74	56	49	34
9	93	90	61	45	40	75	85	64	66	36	89	72	43	99	90	92	10	10	85
53	94	30	31	62	92	82	30	94	56	40	4	50	53	9	74	87	2	36	36
18	69	77	38	89	78	30	68	71	92	22	93	91	74	52	1	97	69	71	42
50	20	76	36	6	20	75	56	36	5	14	70	9	78	23	33	91	33	25	72
30	46	1	10	16	72	69	26	94	39	80	36	36	68	92	74	22	74	41	42
59	47	7	92	77	55	2	12	5	24	0	30	25	62	83	36	92	96	36	75
93	22	3	20	82	44	16	69	98	72	30	57	77	15	90	29	32	38	3	48
9	55	27	41	40	94	77	14	54	10	25	75	1	74	72	15	69	80	33	58
70	8	3	5	46	89	28	86	40	6	25	40	81	26	63	97	87	48	26	41
19	6	89	31	80	60	13	89	17	69	38	93	58	55	54	69	74	33	8	55

FIG. 13.1 A table of random numbers.

- The representativeness of the sample is maximized in relation to the population characteristics.
- The differences in the characteristics of the sample and the population are purely a function of chance.
- The probability of choosing a nonrepresentative sample decreases as the size of the sample increases.

You must remember, however, that although a researcher may use a carefully controlled sampling procedure that minimizes error, no guarantee exists that the sample will be representative. Factors such as sample heterogeneity and participant dropout may jeopardize the representativeness of the sample despite the most stringent random sampling procedure. In examining the relationship between critical care nurses' information-seeking behaviour, perception of personal control, training, and the nonroutineness of tasks, Newman, Doran, and Nagle (2014) drew the sample from a population of critical care nurses working in hospitals in Ontario, Canada, to reduce heterogeneity. A random sample was drawn from the College of Nurses of Ontario database.

The major disadvantage of simple random sampling is that it is a time-consuming and inefficient method of obtaining a random sample. (Consider the task of listing all baccalaureate nursing students in Canada.) With random sampling, it may also be impossible to obtain an accurate or complete listing of every element in the population. Imagine, for example, trying to obtain a list of all completed suicides in Toronto for 2021. Although suicide may have been the cause of death, another cause (e.g., cardiac failure) often appears on the death certificate. It would be difficult to estimate how many elements of the target population would be eliminated from consideration. Bias would definitely be an issue, despite the researcher's best efforts. Thus, the evaluator of a research article must exercise caution in generalizing from reported findings, even when random sampling is the stated strategy, if the target population has been difficult or impossible to list completely.

Stratified Random Sampling

Stratified random sampling requires that the population be divided into strata or subgroups. The subgroups or subsets that the population is divided into are homogeneous. An appropriate number of elements from each subset is randomly selected on the basis of the proportion in the population. The goal of this strategy is to achieve a greater degree of representativeness. Stratified random sampling is similar to the proportional stratified quota sampling strategy discussed earlier in this chapter. The major difference is that stratified random sampling involves a random selection procedure for obtaining sample participants. Figure 13.2 illustrates the use of stratified random sampling.

The population is stratified according to any number of attributes, such as age, gender, ethnicity, religion, socioeconomic status, or level of education completed. The variables selected to make up the strata lead to subgroups that share one or more of the attributes being studied (see Practical Application box). The following questions can be asked in the selection of a stratified sample:

- Does a critical variable or attribute exist that provides a logical basis for stratifying the sample?
- Does the population list contain sufficient information about the attributes that will be used to divide the sample into subsets?
- Is it appropriate for each subset to be equal in size, or is it more appropriate for each subset to be proportionally stratified on the basis of the proportion of each subset in the population?
- If proportional sampling is being used, is the number of participants in each subset sufficient as a base for meaningful comparisons?
- Once the subset comparison has been determined, are random procedures used for selection of the sample?

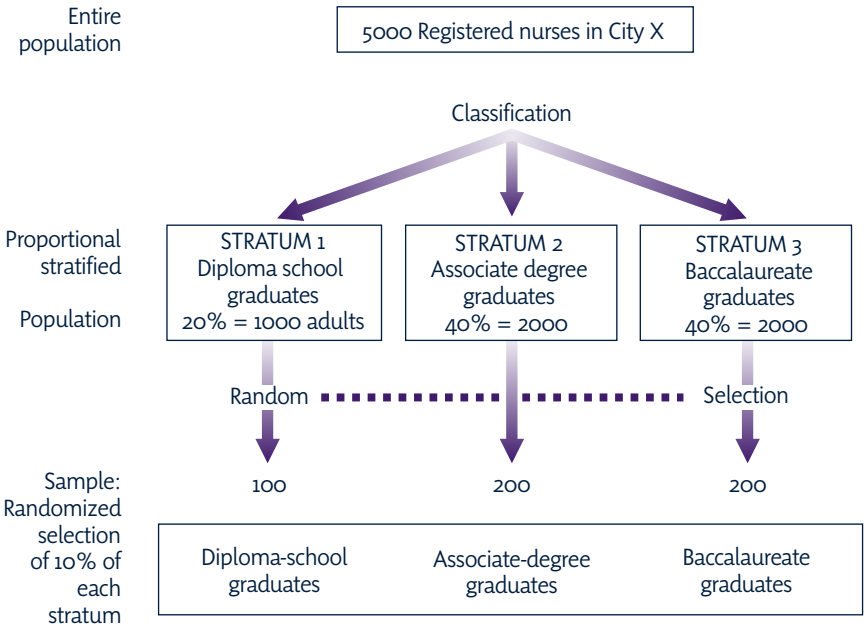


FIG. 13.2 Participant selection through the use of a proportional stratified random sampling strategy.

Practical Application

Koopman et al. (2016) explored the relationship between hope, coping, and quality of life in adults with myasthenia gravis. The researchers stratified a sample of 100 patients to match the proportion of these patients according to the type of myasthenia gravis in the population: 25% ocular myasthenia gravis and 75% generalized myasthenia gravis.

As illustrated in Table 13.1, a stratified random sampling strategy has the following advantages: (1) the representativeness of the sample is enhanced, and (2) the risk of bias is low (i.e., the researcher has a valid basis for making comparisons among subsets if information about the critical variables has been available). A third advantage is that the researcher is able to oversample a disproportionately small stratum to adjust for the researchers' underrepresentation, statistically weigh the data accordingly, and continue to make legitimate comparisons.

The obstacles encountered by a researcher in using this strategy include (1) the difficulty of obtaining a population list containing complete critical variable information; (2) the time-consuming effort of obtaining multiple enumerated lists; (3) the challenge of enrolling proportional strata; and (4) the time and money involved in carrying out a large-scale study with a stratified sampling strategy. In critiquing the study, you must question the appropriateness of this sampling strategy for the problem under investigation.

Havaei et al. (2016) used a proportionate stratified random sample, based on health authorities and employment status, to describe and compare registered nurse (RN) and licensed practical nurse (LPN) emotional exhaustion, intention to leave, and reasons for leaving. It is appropriate for the researcher to strive to represent all strata proportionately in the study sample.

Multistage Sampling (Cluster Sampling)

Multistage sampling, or **cluster sampling**, involves a successive random sampling of units (clusters) that meet sample eligibility criteria; this sampling progresses from large to small. A **sampling unit** is an element or set of elements used for selecting the sample. The first-stage sampling

unit consists of large units or clusters. The second-stage sampling unit consists of smaller units or clusters. Third-stage sampling units are even smaller.

Consider an example in which a sample of nurse practitioners is desired. The first sampling unit is a random sample of hospitals, obtained from a provincial nurses' association list, that meet the eligibility criteria (e.g., size, type). The second-stage sampling unit consists of a list of acute care nurse practitioners (ACNPs) practising at each hospital selected in the first stage (i.e., the list obtained from the vice president for nursing at each hospital). The criteria for inclusion in the list of ACNPs are as follows: (1) participants must be certified ACNPs with at least 2 years' experience as an ACNP; (2) at least 75% of the ACNPs' time must be spent in providing care directly to patients in acute or critical care practices; and (3) the participants must be in full-time employment at the hospital. The second-stage sampling strategy calls for random selection of two ACNPs from each hospital who meet the eligibility criteria.

When multistage sampling is used in relation to large national surveys, provinces are used as the first-stage sampling unit, followed by successively smaller units (such as counties, cities, districts, and blocks) as the second-stage sampling unit and then households as the third-stage sampling unit.

Sampling units or clusters can be selected by simple random or stratified random sampling methods (see Practical Application box). Suppose that the hospitals described in the preceding example are grouped into four strata according to size (i.e., number of beds) as follows: (1) 200 to 299; (2) 300 to 399; (3) 400 to 499; and (4) 500 or more. Stratum 1 comprises 25% of the population; stratum 2 comprises 30% of the population; stratum 3 comprises 20% of the population; and stratum 4 comprises 25% of the population. Thus, either a simple random or a proportional stratified sampling strategy can be used to randomly select hospitals that would proportionately represent the population of hospitals in the provincial nurses' association list. An example of cluster sampling being used is the study by Rostad et al. (2018), where the researchers assessed whether regular pain assessment using a pain assessment tool is associated with changes in (1) pain scores and (2) analgesic use in nursing home residents with severe dementia. A cluster was defined as a single nursing home, and it was used to decrease the risk of contamination effects (see Chapter 17).

The main advantage of cluster sampling, as stated in Table 13.1, is that it is considerably more economical in terms of time and money than other types of probability sampling, particularly when the population is large and geographically dispersed or when a sampling frame of the elements is not available. However, cluster sampling has two major disadvantages: (1) more sampling errors tend to occur than with simple random or stratified random sampling, and (2) the appropriate handling of the statistical data from cluster samples is very complex.

In critiquing a research report, you need to consider whether the use of cluster sampling is justified in light of the research design, as well as other pragmatic matters, such as economy.

Practical Application

Sawatzky et al. (2019) used a two-stage clustered and stratified sampling strategy. Stage 1 involved a random selection of 114 sites from five health authorities in BC, which included both urban and rural jurisdictions. The sampling of sites was clustered by health authority and stratified by type of site (39 acute medical units, 37 home care settings, and 38 residential care facilities), size of site (based on a median split of the sizes of sites within each health authority), and, for acute medical care units, whether the site was specialized or general medical unit.

Laschinger, Read, Wilk, et al. (2014) used a cluster sample of 525 nurses in 49 nursing units in 25 acute care hospitals, across all regions in Ontario, to study the influence of nursing unit empowerment and social capital on unit effectiveness and nurse perceptions of patient care quality.

Systematic Sampling

Systematic sampling is a sampling strategy that involves the selection of every "*k*th" case drawn from a population list at fixed intervals, such as every 10th member listed in the directory of the College and Association of Registered Nurses of Alberta (CARNA). Systematic sampling might be used to recruit every "*k*th" person who enters a hospital lobby or who is hospitalized with a diagnosis with the COVID-19 infection in 2020. When systematic sampling is used, the population must be narrowly defined (e.g., as consisting of all people entering or leaving the hospital lobby) for the sample to be considered a probability sample. If older adults were sampled systematically on entering a hospital lobby, the resulting sample would not be a probability sample because not every older adult would have a chance of being selected. As such, systematic sampling can sometimes represent a nonprobability sampling strategy.

Systematic sampling strategies can be designed, however, to fulfill the requirements of a probability sample. First, the listing of the population (sampling frame) must be random in relation to the variable of interest. For example, suppose that participants were being selected from every 10th hospital room for a study on patient satisfaction with nursing care. In the hospital where the study was being conducted, every 10th room happened to be a private room. Patients in private rooms might respond differently regarding their satisfaction than patients in semiprivate rooms. Because of the nonrandom arrangement of the rooms, bias may be introduced.

Second, the first element or member of the sample must be selected randomly. In this case, the researcher—who has a population list, or sampling frame—first divides the population *(N)* by the size of the desired sample *(n)* to obtain the sampling interval width *(k)*. The **sampling interval** is the standard distance between the elements chosen for the sample. For example, to select a sample of 50 family nurse practitioners from a population of 500 family nurse practitioners, the sampling interval would be as follows:

$$k = \frac{500}{50} = 10$$

Essentially, every 10th case on the family nurse practitioner list would be sampled. Thus, if the starting point was participant 5, the next person chosen would be 15th, then 25th, etc.

Once the sampling interval has been determined, the researcher uses a table of random numbers (see Figure 13.1) to obtain a starting point for the selection of the 50 participants. If the population size is 500 and a sample size of 50 is desired, a number between 1 and 500 is randomly selected as the starting point. In this instance, if the first number is 51, the family nurse practitioners corresponding to numbers 51, 61, 71, and so forth would be included in the sample of 50.

Another procedure recommended in many texts is to randomly select the first element from within the first sampling interval. If the sampling interval is 5, a number between 1 and 5 is selected as the random starting point. For example, the number 3 is randomly chosen. Keeping in mind the sampling interval of 5, the next elements selected would correspond to the numbers 8, 13, 18, and so on, until the sample was obtained. Although this procedure is technically correct, choosing a random starting point from across the total population of elements is more attractive because every element has a chance to be chosen for the sample during the first selection step.

Systematic sampling and simple random sampling are essentially the same type of procedure. The advantage of systematic sampling is that the results are obtained in a more convenient and efficient manner (see Table 13.1). The disadvantage of systematic sampling is that bias in the form of nonrandomness can be inadvertently introduced into the procedure. This problem may occur if the population list is arranged so that a certain type

of element is listed at intervals that coincide with the sampling interval. For example, if every 10th nursing student on a population list of all types of nursing students in Ontario was a baccalaureate student and the sampling interval was 10, baccalaureate students would be overrepresented in the sample.

Cyclical fluctuations are also a factor in systematic sampling. For example, if a list is kept of nursing students using the college library each day to do computer literature searches, a biased sample would probably be obtained if every seventh day, such as Sunday, is chosen as the sampling interval because probably fewer and perhaps different nursing students use the library on Sundays than on weekdays. Therefore, caution must be exercised about departures from randomness because they affect the representativeness of the sample and, as a result, the external validity of the study.

You should note whether a satisfactory random selection procedure was performed. If randomization was not used, the systematic sampling may have become a nonprobability quota sample. You need to be cognizant of this issue because the implications related to interpretation and generalizability are drastically altered when a nonprobability sample is involved.

For example, in their study, Ridout, Aucoin, Browning, and associates (2014) explored the incidence of failure to communicate vital information as patients progressed through the six phases of the perioperative process. One thousand eight hundred fifty-eight eligible surgical cases were identified, and 293 charts needed to be reviewed to achieve a power of 0.8 and determine a difference at the 0.05 level. Every sixth record that met the criteria was used for the study. Because randomization was not used at any phase of this multilevel sampling procedure, you would consider this study to be a nonprobability stratified sample with the external validity limitations of that sampling strategy (see Chapter 10).

Evidence-Informed Practice Tip

The sampling strategy, whether probability or nonprobability, must be appropriate for the study design and evaluated in relation to the level of evidence provided by the design.

Special Sampling Strategies

Several special sampling strategies are used in nonprobability sampling. **Matching** is a special strategy used to construct an equivalent comparison sample group by filling it with participants who are similar to each participant in another sample group in terms of pre-established variables, such as age, gender, level of education, medical diagnosis, or socioeconomic status. Theoretically, any variable other than the independent variable that could affect the dependent variable should be matched. In reality, the more variables matched, the more difficult it is to obtain an adequate sample size.

For example, Graziotti, Hammond, Messinger, and colleagues (2012) examined strategies to maximize retention in longitudinal studies involving high-risk families. The researchers conducted a follow-up study of 1388 children, half of whom were initially identified as cocaine or opiate exposed ($n = 658$), who were then matched to a non-cocaine- or non-opiate-exposed control ($n = 730$), based on gestational age, race, and gender.

SAMPLING STRATEGIES USED IN QUALITATIVE RESEARCH

Because nonprobability sampling is the best method of obtaining individuals who are key informants of a phenomenon, these sampling methods are widely used in qualitative research. As described in Chapter 7, qualitative research methods are conducted to gain both insights into and in-depth meaning about experiences, incidents, or events. In qualitative research, the sampling procedure is governed by the methodology used.

Many sampling strategies are used in qualitative sampling, but the most common approaches are *convenience sampling, network sampling, purposive sampling*, and *theoretical sampling*.

Convenience Sampling

Convenience sampling is also used in qualitative research to access participants of a particular phenomenon. Buck-McFadyen and MacDonnell (2017) used a purposeful convenience sampling strategy to recruit nurses and nursing students who had an interest in political activism and who represented a diversity of practice settings, educational backgrounds, and generational cohorts to obtain their sample of 13 educators and 14 nursing students.

Network Sampling

Network sampling, sometimes referred to as **snowball effect sampling** or *snowballing*, is a strategy used for locating samples that are difficult or impossible to locate in other ways. This sampling strategy takes advantage of social networks and the tendency of friends to share characteristics. When a few participants with the necessary eligibility criteria are found, the researcher asks for their assistance in getting in touch with other people with similar characteristics that meet these criteria.

Network sampling was described by O'Byrne et al. (2014), who studied how gay and bisexual men perceived the criminal prosecution of persons living with HIV and who do not disclose their HIV status. They recruited participants by raising awareness of the project within AIDS service agencies, distributed posters in venues frequented by gay men, and used snowball sampling by giving participants a supply of the research assistant's business cards to pass on to others who might be willing to participate.

In a qualitative study, Pesut et al. (2019) explored the implications of a legislated approach to assisted death for nurses' experiences and nursing practice (see Appendix A). Fifty-nine registered nurses and nurse practitioners were recruited using convenience, purposive, and snowball sampling. The researchers advertised through the Canadian Nurses' Association, through health regions, and through the Canadian Association of MAiD Assessors and Providers (p. 3).

Today, online computer networks, as described in the following section on purposive sampling, can be used to assist researchers in recruiting participants who are otherwise difficult to locate, thereby taking advantage of the networking or snowball effect. The Critical Thinking Decision Path illustrates the relationship between the type of sampling strategy and the appropriate generalizability.

Purposive Sampling

Purposive sampling is an increasingly common strategy in which the researcher's knowledge of the population and its elements is used to handpick the cases to be included in the sample. The researcher usually selects participants who are considered typical of the population.

For example, Currie and Szabo (2020) used purposive sampling to recruit 15 parents (11 mothers and 4 fathers) from medical genetic, endocrine, and neuropsychiatry clinics to study social isolation and exclusion in parents who are caring for children with rare neurodevelopmental disorders.

A purposive sample is also used when a highly unusual group is being studied, such as a population with a rare genetic disease (e.g., Tay-Sachs disease). In this case, the researcher would describe the sample characteristics precisely to ensure that the reader will have an accurate picture of the participants in the sample. This type of sample can also be used to study the differential effect of risk factors in a specific population longitudinally. In another situation, the researcher may wish to interview individuals who reflect a particular characteristic. For example, Guruge

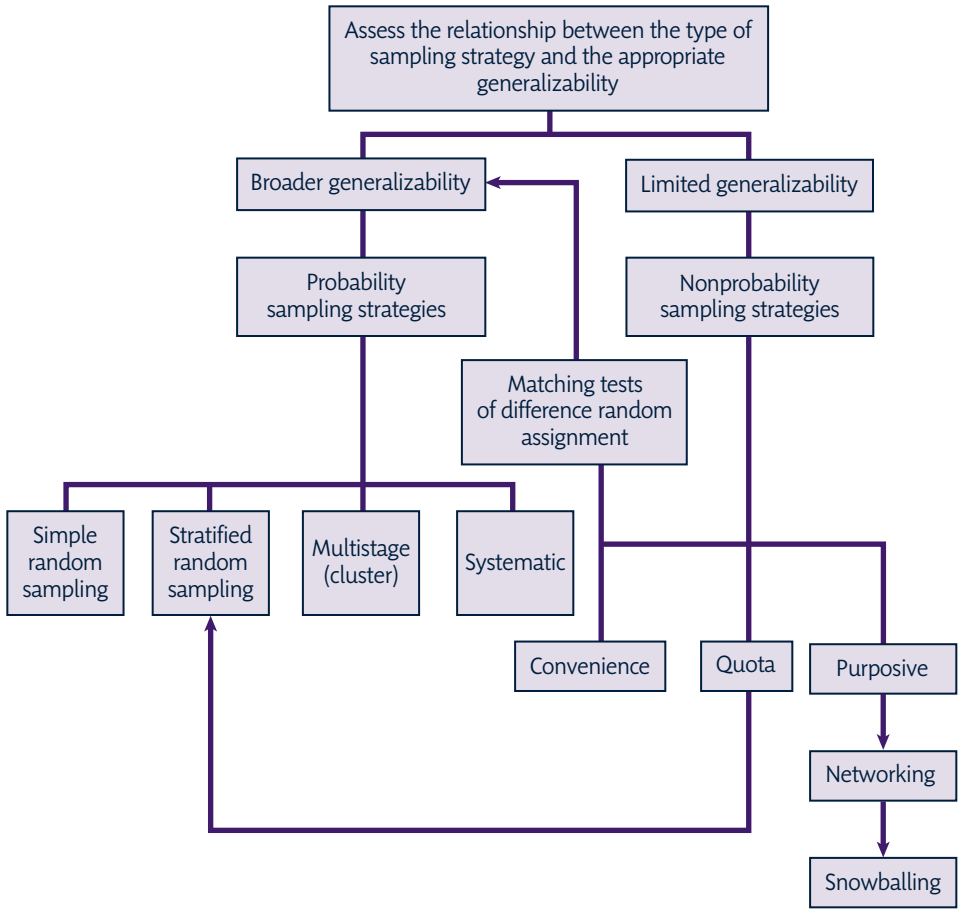
et al. (2019) used purposive sampling to clarify older immigrants' social needs, networks, and support and how they shape their capacity, resilience, and independence in aging well in Ontario.

Ganann et al. (2019) used two types of qualitative sampling strategies in their study to explore provider perspectives on facilitators and barriers to accessible service provision for immigrant women with postpartum depression. A stratified purposeful sampling (i.e., by profession, years in practice, practice settings) was employed to select the initial knowledge users, who were the service providers and administrators, and provided further contacts. Then snowball sampling was used to recruit more knowledge users (p. 192).

Today, computer networks (e.g., online services) can be of great value in helping researchers access and recruit participants for purposive samples. For instance, Balneaves and Alraja (2019) used letters of invitation via e-mails to eligible participants to explore the perspectives of Nurse Practitioner nursing regulatory bodies regarding

CRITICAL THINKING DECISION PATH

Assessing the Relationship between the Type of Sampling Strategy and the Appropriate Generalizability



practice and policy issues related to medical cannabis.

The researcher who uses a purposive sample assumes that errors of judgement in overrepresenting or underrepresenting elements of the population in the sample will tend to balance each other. The validity of this assumption, however, cannot be determined objectively. You must be aware that the more heterogeneous the population, the greater the chance that bias is introduced in the selection of a purposive sample. As indicated in Table 13.1, conscious bias in the selection of participants remains a constant concern. Therefore, the findings from a study involving a purposive sample should be regarded with caution. As with any nonprobability sample, the ability to generalize is very limited. The following are several instances when a purposive sample may be appropriate:

- The effective pretesting of newly developed instruments with a purposive sample of diverse types of people
- The validation of a scale or test with a known-groups technique
- The collection of exploratory data in relation to an unusual or highly specific population, particularly when the total target population remains unknown to the researcher
- The collection of descriptive data (e.g., as in qualitative studies) with which researchers seek to describe the lived experience of a particular phenomenon (e.g., postpartum depression, caring, hope, or surviving childhood sexual abuse)
- The focus of the study population when it is related to a specific diagnosis (e.g., type 1 diabetes, multiple sclerosis), a specific condition (e.g., legal blindness, terminal illness), or a specific demographic characteristic (e.g., same-sex twin pairs)

Many types of purposive sampling exist (Palys, 2008), but the following three types of cases are the most often used:

1. Typical cases: cases that are "normal" or "average" among those being studied
2. Deviant or extreme cases: cases that represent unusual manifestations of the phenomenon of interest
3. Confirming or disconfirming cases: cases that are exceptions, that represent variation, or for which an initial elaborate analysis is necessary

In any type of purposive sampling, sampling is stopped when **data saturation** occurs—that is, when the information being shared with the researcher becomes repetitive.

Criterion sampling is also a form of purposive sampling. The researcher needs to have a set of criteria for a sample, and all cases that meet these criteria are selected. It is important that the criteria are established so that cases that are chosen will yield rich data relevant to the research problem being explored—for example, all patients who were in a smoking cessation program and have resumed smoking. This criterion would enable an understanding of what is needed to support individuals who wish to quit smoking.

Theoretical Sampling

Theoretical sampling is associated with grounded theory research. As you learned in Chapter 8, the goal of grounded research is theory generation; thus, a theoretical sampling strategy is used to fully elaborate and validate variations in the data by finding examples of a theoretical construct (Sandelowski, 1995). In theoretical sampling, the researcher selects experiences that will help test ideas and gather complete information about developing concepts. Sampling is stopped when theory saturation or redundancy occurs.

Convenience and theoretical sampling were used by King-Shier et al. (2019) who examined the process that South Asians undergo when managing their hypertension. The initial sample was from those who volunteered, then theoretical sampling was used where participants were screened based on: 1) additional criteria to ensure adequate reflection of the group being studied (e.g., sex, age, language, and time since immigration); and 2)

emergence of the theory and category. Sampling continued until theoretical saturation occurred, that is, no new data were revealed (p. 322).

Research Hint

Look for a brief discussion of a study's sampling strategy in the "Methods" section of a research article. Some articles have a separate subsection with the heading "Sample," "Participants," or "Study Participants." A statistical description of the characteristics of the actual sample often does not appear until the "Results" section of a research article.

SAMPLE SIZE: QUANTITATIVE

No single rule can be applied to the determination of a sample's size. When researchers estimate sample size, they must consider many factors, such as the following:

- The type of design used
- The type of sampling procedure used
- The type of formula used for estimating the optimal sample size
- The degree of precision required
- The heterogeneity of the attributes under investigation
- The relative frequency at which the phenomenon of interest occurs in the population (i.e., a common versus a rare health problem)
- The projected cost of using a particular sampling strategy
- The homogeneity of the population
- The anticipated response rate of participants
- The attrition rate, especially in longitudinal studies with multiple data collection points

The sample size should be determined before the study is conducted. A general rule is always to use the largest sample possible. The larger the sample, the more likely it is to be representative of the population; smaller samples produce less accurate results.

An exception to the rule about sample size is the **pilot study**, which is a small sample study conducted as a prelude to a larger-scale (parent) study. The pilot study typically is conducted with similar methods and procedures that both yield preliminary data for determining the feasibility of conducting a larger-scale study and establish that sufficient scientific evidence exists to justify subsequent, more extensive research.

Hertzog (2008) summarized methods for justifying sample sizes on the basis of the aim of the pilot study. This author suggests that a sample size as small as 10 to 15 participants per group may be sufficient for the decisions being made but cautions that this is not a simple or straightforward issue as these types of studies are influenced by many factors as noted above. For pilot studies involving group comparisons, 10 to 20 participants per group may be enough. On the other hand, if a researcher is developing or testing an instrument, it is suggested that each group comprise 35 to 40 participants. For example, Santiago et al. (2019) conducted a pilot study to (1) measure the feasibility of implementing a tablet equipped with a communication app (TalkRocketGoTM) for patients with an endotracheal or tracheostomy tube who are unable to communicate using verbal speech, and (2) determine if bedside clinicians find a tablet equipped with a communication app a useful device in clinician-patient communication interactions (p. 18); the sample size was 20 patients. Billingham et al. (2013) concluded in their review of sample sizes for pilot and other feasibility studies that these type of studies do not necessarily require a sample size calculation, but researchers should be able to justify their chosen sample size. There are also other methods of ascertaining sample sizes in a pilot study based on a simple formula with a chosen level of confidence, problems that may arise with a given probability (Viechtbauer et al., 2015).

The principle of "larger is better" holds true for both probability and nonprobability samples. Results based on small samples (fewer than 10 participants) tend to be unstable; the values fluctuate from one sample to the next. Small samples tend to increase the probability of obtaining a markedly nonrepresentative sample. As the

sample size increases, the mean more closely approximates the population values; thus, fewer sampling errors are introduced.

An example of this concept is illustrated by a study in which the average monthly consumption of sleeping pills was investigated for patients on a rehabilitation unit after a cerebrovascular accident. The data in Table 13.3 indicate that the population consisted of 20 patients whose average consumption of sleeping pills was 15.2 per month. The population of 20 patients was divided into sets of two simple random samples with sizes of 2, 4, 6, and 10. Each sample average in the right column represents an estimate of the population average, which is known to be 15.15. In most cases, the population value was unknown to the researchers, but because the population is so small, it could be calculated. In Table 13.3, note that with a sample size of two, the estimate might have been wrong by as much as eight sleeping pills in sample 1B. As the sample size increases, the averages get closer to the population value, and the differences in the estimates between samples A and B also get smaller. Large samples permit the principles of randomization to work effectively (i.e., to counterbalance atypical values in the long run).

The sample size can be estimated with the use of a statistical procedure known as *power analysis* (see Chapter 17). A simple example illustrates this concept. Suppose that a researcher wants to determine the effect of nurse preoperative teaching on patient postoperative anxiety. Patients are randomly assigned to an experimental group or a control group. How many patients should be used in the study? When using power analysis, the researcher must estimate how large a difference will be observed between the groups (i.e., the difference in the mean amount of postoperative anxiety after the experimental preoperative teaching program). This difference is called the **effect size**. If a small difference is expected, the sample must be large (in this case, 196 patients in each group) to ensure that the differences will be revealed in a statistical analysis. If a medium-size difference is expected, the total sample size would be 128 (64 in each group). When expected differences are large, a small sample size can ensure that differences will be revealed through statistical analysis.

An example is illustrated by the study of Penz et al. (2018) who tested a conceptual model of confidence and competence in rural and remote nursing practice. Before data collection, they

TABLE 13.3

COMPARISON OF POPULATION AND SAMPLE VALUES AND AVERAGES IN A STUDY OF SLEEPING PILL CONSUMPTION

NUMBER IN GROUP	GROUP	NUMBER OF SLEEPING PILLS CONSUMED (VALUES EXPRESSED MONTHLY)	AVERAGE
20	Population	1, 3, 4, 5, 6, 7, 9, 11, 13, 15, 16, 17, 19, 21, 22, 23, 25, 27, 29, 30	15.2
2	Sample 1A	6, 9	7.5
2	Sample 1B	21, 25	23.0
4	Sample 2A	1, 7, 15, 25	12.0
4	Sample 2B	5, 13, 23, 29	17.5
6	Sample 3A	3, 4, 11, 15, 21, 25	13.3
6	Sample 3B	5, 7, 11, 19, 27, 30	16.5
10	Sample 4A	3, 4, 7, 9, 11, 13, 17, 21, 23, 30	13.8
10	Sample 4B	1, 4, 6, 11, 15, 17, 19, 23, 25, 27	13.8

conducted a power analysis, with an alpha value set at .05 and the power set at .80; the power analysis indicated that a minimum sample size of 1889 would be required to detect a significant effect (with a small effect size of 0.1). Alpha is the probability of making a type I error (rejecting the null hypothesis when the null hypothesis is true). Another example is provided by Tryphonopoulos and Letourneau (2020), who conducted a feasibility pilot study to test a video-feedback interaction guidance intervention designed to improve maternal–infant interaction, depressive symptoms, and cortisol patterns of depressed mothers and their infants. At the outset of the study, original power calculations (20 per group, alpha .05, power .080) would have been sufficient to detect a moderate effect in the primary outcome of maternal–infant interaction. Instead their recruitment was slow, and they obtained only 12 participants, 6 in each group. Therefore, all analyses were underpowered, leading to possible Type II error, which means that no group differences were found when there could possibly be differences (see Chapter 17).

Power analysis is an advanced statistical technique that is commonly used by researchers and is a requirement for external funding. When power analysis is not used, research studies may be based on samples that are too small, which may lead to a lack of support for the researcher's hypotheses and to a type I error (rejecting a null hypothesis when it should have been accepted); in other words, the researcher finds significant results when none exist (see Chapter 17). A researcher may also commit a type II error (accepting a null hypothesis when it should have been rejected) if the sample is too small; in other words, the sample is too small to detect treatment effects (see Chapter 17).

Despite the principles related to determining sample size that have been identified in this chapter, you should be aware that large samples do not ensure representativeness or accuracy. A large sample cannot compensate for faulty research design. The proportion of the population that is sampled does not provide a guarantee of accurate results. Accurate results can be obtained from only a small fraction of a large population. For example, a 10% probability sample of a population containing 1500 elements will yield more precise results than will a nonprobability .01% sample of a population with 100,000 elements.

You should evaluate the sample size in terms of (1) how representative the sample is of the target population and (2) to which population the researcher wishes to generalize the results of the study. The goal of sampling is to gather a sample as representative as possible with as few sampling errors as possible.

SAMPLE SIZE: QUALITATIVE

In qualitative research, no power analyses are conducted a priori to determine sample size requirements, but there is discussion around thinking of a priori sample size (Sim et al., 2018; Turner-Bowker et al. 2018). According to Sandelowski (1995), sample size is determined by the purpose and type of the sampling and the research method to be used. Morse (1994) recommended about six participants for phenomenological studies and about 30 to 50 cases for ethnographies and grounded theory studies. Creswell (1998) suggests 5 to 25 cases for phenomenology and 20–30 cases for a grounded theory. As you can see, these are suggestions and do not constitute a hard-and-fast rule because a one-person case study may be sufficient for a phenomenological study. When you critique a study, you need to note how the researcher has explained the sampling plan, how data saturation was met, and what limitations have been stated. Participants are added to the sample until data saturation is reached (i.e., new data no longer emerge during the data-collection process). The fittingness of the data is a more important concern than the representativeness of participants (see Chapter 15).

Research Hint

Remember to look for some rationale about the sample size and the strategies that the researcher has used (e.g., matching, test of differences on demographic variables) to ascertain or build in sample representativeness.

Evidence-Informed Practice Tip

Research designs and types of samples are often linked. You would expect to see experimental designs in which probability sampling strategies were used; if a nonprobability purposive sampling strategy is used to recruit participants to such a study, you would expect the participants to then be randomly assigned to intervention and control groups.

SAMPLING PROCEDURES

The criteria for selecting a sample vary according to the sampling strategy. Regardless of which strategy is used, the procedure must be systematically organized. Such organization will eliminate the bias that occurs when sample selection is carried out inconsistently. Bias in sample representativeness and generalizability of findings are important sampling issues that have generated national concern.

For example, many of the landmark adult health studies (e.g., the Framingham Heart Study and the Baltimore Longitudinal Study of Aging) historically excluded women as participants. The findings of these studies were generalized from men to all adults despite the lack of female representation in the samples. Findings based on Euro-American or Euro-Canadian data cannot be generalized to Punjabis, Chinese, West Indians, or any other cultural group. Consequently, careful identification of the target population is a crucial step in the process. For example, Donnelley et al. (2017) conducted a study to determine if participation in an ethics consultation simulation increased nursing students' knowledge of nursing ethics principles compared to students who were taught ethics principles in the traditional didactic format. The researchers noted that the sample consisted of students from three United States Midwest colleges/universities and that these students may not be representative of all United States students (p. 158).

In order to establish conclusions about, for example, psychosocial stressors related to all patients with a first-time myocardial infarction, both men and women must be included in the target population. As another example, to establish conclusions about the incidence of extrapyramidal adverse effects of haloperidol (Haldol) in a psychiatric ward among Chinese patients in comparison with Euro-Canadians, the target population must be diverse. Sometimes, however, the target population must be gender specific, as when breast or prostate cancer or aspects of pregnancy or menopause are studied.

Several general steps (Figure 13.3) ensure the identification of a consistent approach by the researcher. Initially, the target population (i.e., the entire group of people or objects about whom the researcher wants to establish conclusions or make generalizations) must be identified. The target population may consist, for example, of all female patients with a first-time diagnosis of breast cancer, all children with asthma, all pregnant teenagers, or all doctoral nursing students in Canada.

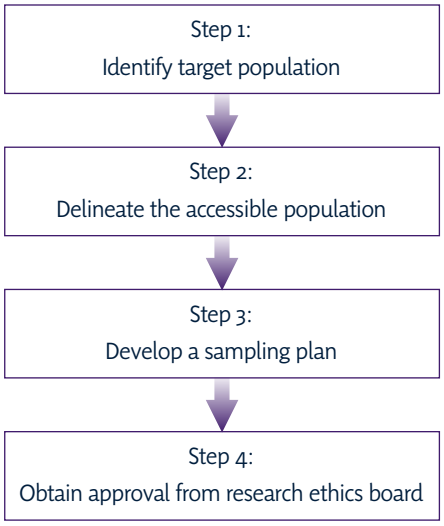


FIG. 13.3 Summary of the general sampling procedures.

Next, the accessible portion of the target population must be delineated. An accessible population might consist of all nurse practitioners in the province of New Brunswick, all older patients with COVID-19 admitted to a certain hospital during 2020, all pregnant teenagers in a specific prenatal clinic, or all children with rheumatoid arthritis under care at a specific hospital specializing in the treatment of autoimmune diseases.

Then a sampling plan or a protocol for actually selecting the sample from the accessible population is formulated. The researcher makes decisions about how participants will be approached, how the study will be explained, and who—the researcher or a research assistant—will select the sample. Regardless of who implements the sampling plan, consistency in how it is done is of paramount importance. In reading a research report, you want to find a description of the sample, as well as the sampling procedure, in the study. On the basis of the appropriateness of what has been reported, you can make judgements about the soundness of the sampling protocol, which of course will affect the interpretations of the findings.

Finally, once the accessible population and sampling plan have been established, permission is obtained from the institution's research board, which is commonly referred to as the *research ethics board*. This permission provides free access to the desired population.

When an appropriate sample size and sampling strategy have been used, the researcher can feel more confident that the sample is representative of the accessible population; however, it is more difficult to feel confident that the accessible population is representative of the target population. Are nurse practitioners in New Brunswick representative of all nurse practitioners in Canada? It is impossible to know for sure. Researchers must exercise judgement when assessing typicality. Unfortunately, no guidelines for making such judgements exist, and critiquers have even less basis on which to make such decisions. The best rule to use when evaluating the representativeness of a sample and its generalizability to the target population is to be realistic and conservative about making sweeping claims in relation to the findings.

Research Hint

Remember to evaluate the appropriateness of the generalizations made about the findings of a quantitative study in view of the target population, the accessible population, the type of sampling strategy, and the sample size. In qualitative research, evaluate the transferability of the findings on the basis of the research design and its sampling strategy and size.

APPRAISING THE EVIDENCE

Sample

The criteria for critiquing the sampling technique of a study are presented in the Critiquing Criteria box. You (the reader) and the researcher approach the "Sample" section of a research report with different perspectives. You need to raise the following two questions:

1. If this study were to be replicated, is enough information available about the nature of the population, the sample, the sampling strategy, and the sample size for another investigator to carry out the study?
2. Are the previously mentioned factors appropriate for the particular research design, and, if not, which factors require modification, especially if the study is to be replicated?

Sampling is considered to be one important aspect of the methodology of a research study. Thus, data pertaining to the sample usually appear in the "Methodology" section of the research report. The sampling content presented should reflect the outcome of a series of decisions based on sampling criteria appropriate to the design of the study, as well as the options and limitations inherent in the context of the investigation. The following discussion highlights several sampling criteria that you should

APPRAISING THE EVIDENCE—*cont'd*

Sample

consider when you evaluate the merit of a sampling strategy in relation to a specific research study.

Initially, the parameters or attributes of the study population should clearly specify to what population the findings may be generalized. In general, the target population of the study is not specifically identified by the researcher, but the nature of it is implied in the description of the accessible population, the sample, or both. For example, if a researcher states that 100 participants were randomly selected from a population of men and women older than 65 and with a diagnosis of COPD who were treated in a respiratory rehabilitation program at a particular hospital during 2011, you can specifically evaluate the parameters of the population. The demographic characteristics of the sample (e.g., age, gender, diagnosis, ethnicity, religion, and marital status) should also be presented in either a tabular or a narrative summary because they provide further explication about the nature of the sample and enable you to evaluate the sampling procedure more accurately. For example, in their study on predictors of workplace integration for internationally educated nurses (IENs) in Canada, Covell et al. (2017) presented detailed data summarizing demographic variables of importance. These data are reproduced as follows:

The participants were mostly middle-aged (M = 47 years, SD = 11), female (85.2%), and self-identified as a visible minority (57.5%). About half of the IENs received their basic nursing education in a developing country, mostly Philippines and India (54.4%). Most respondents were RNs (89.2%), and the remaining were Licensed Practical Nurses (LPNs) (9.6%) or Registered Practical/Psychiatric Nurses (RPNs) (1.1%) (p. 6).

This example illustrates how a detailed description of the sample both provides a frame of reference for the study population and sample and generates questions to be raised. When this demographic sample information is available, you are able to evaluate the sampling strategy and the impact on the findings. Also helpful is the researcher's rationale for having elected to study one type of population versus another. For example, Covell et al. (2017) wrote that in the last decade, there has been a dramatic increase in IENS, and they would like to find out how IENs can be facilitated into the Canadian nursing workforce.

In a research study in which a nonprobability sampling strategy is used, it is particularly important to fully describe the population and the sample in terms of who the study participants were, how they were chosen, and the reason they were chosen. If these criteria are adhered to, the degree of heterogeneity or homogeneity of the sample can be determined. The use of a homogeneous sample minimizes the amount of sampling error introduced, a problem particularly common in nonprobability sampling.

Next, the defined representativeness of the population should be examined. Probability sampling is clearly the ideal sampling procedure for ensuring the representativeness of a study population. Use of random selection procedures (e.g., simple random, stratified random, cluster, or systematic sampling strategies) minimizes the occurrence of conscious and unconscious biases that affect the researcher's ability to generalize about the findings from the sample to the population. You should be able to identify the type of probability strategy used and determine whether the researcher adhered to the criteria for a particular sampling plan. In experimental and quasiexperimental studies, you must also know whether or how the participants were assigned to groups. If the criteria have not been followed, you have a valid reason for being skeptical about the proposed conclusions of the study.

Random selection is the ideal in establishing the representativeness of a study population; more often, however, realistic barriers (e.g., institutional policy, inaccessibility of participants, lack of time or money, and current state of knowledge in the field) necessitate the use of nonprobability sampling strategies. Many important research problems that are of interest to nurses do not lend themselves to experimental design and probability sampling, particularly qualitative research designs. A well-designed, carefully controlled study with a nonprobability sampling strategy can yield accurate and meaningful findings that make a significant contribution to nursing's scientific body of knowledge. As the critiquer, you must ask a philosophical question: "If it is not possible or appropriate to conduct an experimental or quasiexperimental investigation

Continued

APPRAISING THE EVIDENCE—*cont'd*

Sample

with the use of probability sampling, should the study be abandoned?" The answer usually suggests that it is better to perform the investigation and be fully aware of the limitations of the methodology than not to acquire the potential knowledge. The researcher is always able to move on to subsequent studies that either replicate the initial study or entail the use of more stringent design and sampling strategies to refine the knowledge derived from a nonexperimental study.

The greatest difficulty in nonprobability sampling stems from the fact that not every element in the population has an equal chance of being represented in the sample. Therefore, some segment of the population will probably be systematically underrepresented. If the population is homogeneous with regard to critical characteristics, systematic bias will not be an important problem. Few of the attributes that researchers are interested in, however, are sufficiently homogeneous to render sampling bias an irrelevant consideration.

Next, the sampling plan's suitability to the research design should be evaluated. In experimental and quasiexperimental designs, some form of random selection or random assignment of participants to groups is used (see Chapter 11). In critiquing the report, you evaluate whether the researcher adhered to the principles of random selection and assignment. Lack of adherence to such principles compromises the representativeness of the sample and the external validity of the study. The following are questions that you might pose in relation to this issue:

- Has a random selection procedure (e.g., a table of random numbers) been identified?
- Has the appropriate random sampling plan been selected? In other words, has a proportional stratified sampling plan been selected instead of a simple random sampling plan in a study in which three distinct occupational levels appear to be critical variables for stratification?
- Has the particular random sampling plan been carried out appropriately? In other words, if a cluster sampling strategy was used, did the sampling units logically progress from the largest to the smallest?

Random sampling should not be regarded as a perfect method of obtaining a representative sample. Sometimes, bias is inadvertently introduced even when random selection is used. In many nonexperimental designs, nonprobability sampling strategies are used. For such studies, you can ask whether a nonexperimental design and a related nonprobability sampling plan were most appropriate. Sometimes, if the researchers had used another type of design or sampling plan, they could have constructed a stronger study that would have produced findings that were more generalizable and more reliable. In critiquing, however, you are rarely in a position to know what factors entered into the decision to plan one type of study rather than another.

You should then determine whether the sample size is appropriate and its size justifiable. The researcher usually indicates in a research article how the sample size was determined; a similar indication is also seen commonly in doctoral dissertations. The method of arriving at the sample size and the rationale should be briefly mentioned. For example, a researcher may state the following:

A power analysis was performed to calculate the number of participants required in this study exploring mindfulness and anxiety in pregnant women. Previous 8-week mindfulness interventions in pregnant women showed medium-to-large effect sizes for anxiety (Cohen's d, 0.48–0.66) and depression (Cohen's d, 0.42–0.75). Thus, at an effect size of d = 0.58, and allowing for a 25% attrition rate, the total number of participants required was 125 for a power of 0.8 (Yang et al., 2019, p. 69).

Kim and De Gagne (2018) did a power analysis to determine that the required sample size should be 26 participants in each group for an effect size of 0.8 at a significance level of 0.05. (p. 35). Their study compared the effects of two debriefing methods (instructor-led vs. peer-led) on nursing skills, knowledge, self-confidence, and quality of debriefing among undergraduate students (p. 34).

The importance of such examples lies in understanding that this type of statement meets the criteria stated at the beginning of the paragraph and should be evident in the research report. Other considerations with regard to sample size, especially when the sample size appears to be small or inadequate and no rationale is stated for the size, are as follows:

APPRAISING THE EVIDENCE—*cont'd*

Sample

- How will the sample size affect the accuracy of the results?
- Are any subsets or cells of the sample overrepresented or underrepresented?
- Are any of the subsets so small as to limit meaningful comparisons?
- Has the researcher examined the effect of attrition on the results?
- Has the researcher recognized and identified any limitations posed by the size of the sample?

Essentially, these criteria necessitate that you carefully scrutinize several important elements pertaining to sample size that have implications for the generalizability of the findings. Keep in mind that in reports of qualitative studies, neither the predetermining nor the method of determining the sample size will be discussed. Rather, the sample size depends on the methodology used and is a function of data saturation (see Chapter 8).

With qualitative research designs, you apply criteria related to sampling strategies that are relevant for a particular type of qualitative study. In general, sampling strategies are purposive because the study of specific phenomena in their natural setting is emphasized; any participant belonging to a specified group is considered to represent that group. For example, in the qualitative study by Udod and associates (2020), the specified group was nurse managers implementing the lean management system in Saskatchewan. The researchers' goal was to explore the perceptions and experiences of nurse managers involved in implementing the Lean management system in a Western Canadian province.

Finally, the "Sample" section of the research report should provide evidence that the rights of human participants have been protected. You will evaluate whether permission was obtained from an institutional research ethics board that reviewed the study with regard to maintaining ethical research standards (see Chapter 6). For example, the research ethics board examines the research proposal to determine whether the introduction of an experimental procedure may be potentially harmful and therefore undesirable. You also need to examine the report for evidence of the participants' informed consent, as well as protection of their confidentiality or anonymity. Research studies that do not demonstrate evidence of having met these criteria are highly unusual. Nevertheless, you will want to be certain that ethical standards that protect sample participants have been maintained.

Many factors must be considered when you critique the "Sample" section of a research report. The type and appropriateness of the sampling strategy become crucial elements in the analysis and interpretation of data, in the conclusions derived from the findings, and in the generalizability of the findings from the sample to the population. As stated earlier in this chapter, the major purpose of sampling is to increase the efficacy of a research study by representing the particular population so that not every element need be studied, while producing the findings that can be generalized from the sample to the population. You must demonstrate that the sampling strategy used provided a valid basis for the findings and their generalizability.

CRITIQUING CRITERIA

1. Have the sample characteristics been completely described?
2. Can the parameters of the study population be inferred from the description of the sample?
3. To what extent is the sample representative of the population as defined?
4. Are criteria for eligibility in the sample specifically identified?
5. Have sample delimitations been established?
6. Would it be possible to replicate the study population?
7. How was the sample selected? Is the method of sample selection appropriate?
8. What kind of bias, if any, is introduced by this method?
9. Is the sample size appropriate? How is it substantiated?
10. Does the researcher indicate that the rights of participants have been ensured?
11. Does the researcher identify the limitations in generalizability of the findings from the sample to the population? Are those limitations appropriate?
12. Is the sampling strategy appropriate for the design of the study and level of evidence provided by the design?
13. Does the researcher indicate how replication of the study with other samples would provide increased support for the findings?

CRITICAL THINKING CHALLENGES

- A research classmate asks the instructor the following question: "Why isn't it better to study an entire population of patients with lung cancer instead of using the research technique of sampling?" How would you answer this question? Include examples that will help the student see your point of view.
- In the report of a quasiexperimental study, the researchers indicated that they used a convenience sample with random assignment. How is this possible? Would they have used a nonprobability or a probability sample? If you agree that this is a legitimate sampling technique, present both the advantages and the disadvantages; if you disagree, indicate your rationale.
- Your research class is having a debate on probability sampling versus nonprobability sampling with regard to desirability and feasibility. You are assigned to present the advantages of nonprobability sampling in nursing research. What arguments would you use?
- Discuss the principle of "larger is better" and its relationship to network sampling and the sample size of qualitative studies. Include in your discussion the concept of data saturation and the use of computer technology.
- Your research classmate is arguing that a random sample is always better, even if it is small and represents only one site. Another student is arguing that a very large convenience sample representing multiple sites can be very significant. Which classmate would you defend, and why?

CRITICAL JUDGEMENT QUESTIONS

1. Why should a researcher avoid drawing conclusions or making generalizations based on the experience of a small number of participants?
   a. Small samples invalidate hypotheses
   b. The researcher may be unable to eliminate his or her bias
   c. Data obtained from a small number may inadequately represent the phenomenon
   d. Small numbers of participants increase the threat to internal validity influenced by history
2. What is the difference between an "accessible population" and a "target population"?
   a. The accessible population meets the inclusion criteria, and the target population meets the exclusion criteria
   b. The target population meets the inclusion criteria, and the accessible population meets the exclusion criteria
   c. The accessible population represents the entire set of cases the researcher wishes to study, and the target population represents that part of the accessible population that could feasibly be included in the study
   d. The target population represents the entire set of cases the researcher wishes to study, and the accessible population represents that part of the target population that could feasibly be included in the study
3. What is the appropriate sampling interval for drawing a systematic sample of 25 subjects who had breast enhancement surgery from 200 people who had that surgery during 1 year at a specific medical centre?
   a. Every 4th patient
   b. Every 5th patient
   c. Every 8th patient
   d. Every 10th patient

KEY POINTS

- Sampling is a process in which representative units of a population are selected for study. Researchers select representative segments of the population because selecting entire populations of interest to obtain accurate and meaningful information is rarely feasible or necessary.
- Researchers establish eligibility criteria; these are descriptors of the population and provide the basis for inclusion into a sample. Eligibility criteria can include age, gender, socioeconomic status, level of education, religion, and ethnicity.

- The researcher must identify the target population (i.e., the entire set of cases about which the researcher would like to make generalizations). Because of pragmatic constraints, however, the researcher usually uses an accessible population (i.e., one that meets the population criteria and is available).
- A sample is a set of elements that make up the population.
- A sampling unit is the element or set of elements used for selecting the sample. The foremost criterion in evaluating a sample is the representativeness or congruence of characteristics with the population.
- Sampling strategies consist of nonprobability and probability sampling.
- In nonprobability sampling, the elements are chosen by nonrandom methods. Types of nonprobability sampling include convenience, quota, and purposive sampling.
- Probability sampling is characterized by the random selection of elements from the population. In random selection, each element in the population has an equal and independent chance of being included in the sample. Types of probability sampling include simple random, stratified random, cluster, and systematic sampling.
- Sample size is a function of the type of sampling procedure being used, the degree of precision required, the type of sample estimation formula being used, the heterogeneity of the study attributes, the relative frequency of occurrence of the phenomena under consideration, and the cost.
- Criteria for selecting a sample vary according to the sampling strategy. Systematic organization of the sampling procedure minimizes bias. The target population is identified, the accessible portion of the target population is delineated, permission to conduct the research study is obtained, and a sampling plan is formulated.
- In critiquing a research report, you evaluate the sampling plan for its appropriateness in relation to the particular research design.
- The completeness of the sampling plan is examined with regard to the potential replicability of the study. In critiquing, you evaluate whether the sampling strategy is the strongest plan for the particular study under consideration.
- An appropriate systematic sampling plan will maximize the efficiency of a research study. It will increase the accuracy and meaningfulness of the findings and enhance the generalizability of the findings from the sample to the population.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Balneaves, L. G., & Alraja, A. A. (2019). Guarding their practice": A descriptive study of Canadian nursing policies and education related to medical cannabis. *BMC Nursing, 8*, 66. https://doi.org/10.1186/s12912-0190390-7.

Benbow, S., Forchuk, C., Berman, H., Gorlick, C., & Ward-Griffin, C. (2019). Spaces of Exclusion: Safety, Stigma, and Surveillance of Mothers Experiencing Homelessness. *The Canadian Journal of Nursing Research = Revue Canadienne de Recherche En Sciences Infirmieres, 51*(3), 202–213. https://doi.org/10.1177/0844562119859138.

Billingham, S. A., Whitehead, A. L., & Julious, S. A. (2013). An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database. *BMC Medical Research Methodology, 13*, 104. https://doi.org/10.1186/1471-2288-13-104.

Buck-McFadyen, E., & MacDonnell, J. (2017). Contested practice: political activism in nursing and implications for nursing education. *International Journal of Nursing Education and Scholarship, 14*(1). https://doi.org/10.1515/ijnes-2016-0026.

Chan, R., Booth, R., Strudwick, G., & Sinclair, B. (2019). Nursing students' perceived self-efficacy and the generation of medication errors with the use of an electronic medication administration record

(eMAR) in clinical simulation. *International Journal of Nursing Education Scholarship, 16*(1). https://doi.org/10.1515/ijnes-2019-0014.

Covell, C. L., Primeau, M.-D., Kilpatrick, K., & St-Pierre, I. (2017). Internationally educated nurses in Canada: Predictors of workforce integration. *Human Resources for Health, 15*(1), 1–16. https://doi.org/10.1186/s12960-017-0201-8.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publications.

Currie, G., & Szabo, J. (2020). Social isolation and exclusion: the parents' experience of caring for children with rare neurodevelopmental disorders. *International Journal of Qualitative Studies on Health and Well-Being, 15*, (1). https://doi.org/10.1080/17482631.2020.1725362.

Dennis, C., Merry, L., & Gagnon, A. (2017). Postpartum depression risk factors among recent refugee, asylum-seeking, non-refugee immigrant, and Canadian-born women: results from a prospective cohort study. *Social Psychiatry and Psychiatric Epidemiology, 52*(4), 411–422. https://doi.org/10.1007/s00127-017-1353-5.

Donnelly, M. B., Horsley, T. L., Adams, W. H., Gallagher, P., & Zibricky, C. D. (2017). Effect of simulation on undergraduate nursing students' knowledge of nursing ethics principles. *Canadian Journal of Nursing Research, 49*(4), 153–159. https://doi.org/10.1177/0844562117731975.

Dosani, A., Hemraj, J., Premji, S. S., Currie, G., Reilly, S. M., Lodha, A. K., Young, M., & Hall, M. (2017). Breastfeeding the late preterm infant: Experiences of mothers and perceptions of public health nurses. *International Breastfeeding Journal, 12*, 23. https://doi.org/10.1186/s13006-017-0114-0.

Fox, M. T., Sidani, S., & Brooks, D. (2010). Differences in sleep complaints in adults with varying levels of bed days residing in extended care facilities for chronic disease management. *Clinical Nursing Research, 19*(2), 181–202. https://doi.org/10.1177/1054773810365957.

Ganann, R., Sword, W., Newbold, K. B., Thabane, L., Armour, L., & Kint, B. (2019). Provider perspectives on facilitators and barriers to accessible service provision for immigrant women with postpartum depression: A qualitative study. *Canadian Journal of Nursing Research, 51*(3), 191–201. https://doi.org/10.1177/0844562119852868.

Graziotti, A. L., Hammond, J., Messinger, D. S., et al. (2012). Maintaining participation and momentum in longitudinal research involving high-risk families. *Journal of Nursing Scholarship, 44*(2), 120–126.

Guruge, S., Sidani, S., Wang, L., Sethi, B., Spitzer, D., Walton-Roberts, M., & Hyman, I. (2019). Understanding social network and support for older immigrants in Ontario, Canada: Protocol for a mixed-methods study. *JMIR Aging, 2*(1), e12616. https://doi.org/10.2196/12616.

Havaei, F., MacPhee, M., & Susan Dahinten, V. (2016). RNs and LPNs: Emotional exhaustion and intention to leave. *Journal of Nursing Management, 24*, 393–399. https://doi-org.ezproxy.library.yorku.ca/10.1111/jonm.12334.

Henrique, A. J., Gabrielloni, M. C., Rodney, P., & Barbieri, M. (2018). Non-pharmacological interventions during childbirth for pain relief, anxiety, and neuroendocrine stress parameters: A randomized controlled trial. *International Journal of Nursing Practice, 24*(3), e12642. https://doi.org/10.1111/ijn.12642.

Hertzog, M. A. (2008). Consideration in determining sample sizes for pilot studies. *Research in Nursing & Health, 32*, 180–191.

Im, E. O., Chang, S. J., Ko, Y., Chee, W., Stuifbergen, A., & Walker, L. (2012). A national internet survey on midlife women's attitudes toward physical activity. *Nursing Research, 61*(5), 342–352. https://doi.org/10.1097/NNR.0b013e31825da85a.

Jackson, K. T., & Dennis, C. L. (2017). Lanolin for the treatment of nipple pain in breastfeeding women: A randomized controlled trial. *Maternal and Child Nutrition,13*(3), 1–10. https://doi.org/10.1111/mcn.12357.

Kim, S. S., & De Gagne, J. C. (2018). Instructor-led vs. peer-led debriefing in preoperative care simulation using standardized patients. *Nurse Education Today, 71*, 34–39. https://doi.org/10.1016/j.nedt.2018.09.001.

King-Shier, K. M., Dhaliwal, K. K., Puri, R., LeBlanc, P., & Johal, J. (2019). South Asians' experience of managing hypertension: A grounded theory study. *Patient Preference and Adherence, 13*, 321–329. https://doi.org/10.2147/PPA.S196224.

Koopman, W. J., LeBlanc, N., Fowler, S., et al. (2016). Hope, coping, and quality of life in adults with myasthenia gravis. *Canadian Journal of Neuroscience Nursing, 38*(1), 56–64.

Laschinger, H. K. S., Read, E., Wilk, P., et al. (2014). The influence of nursing unit empowerment and social capital on unit effectiveness and nurse perceptions of patient care quality. *Journal of Nursing Administration, 44*(6), 347–352.

Morse, J. M. (1994). Designing funded qualitative research. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 220–235). Thousand Oaks, CA: Sage.

Newman, K., Doran, D., & Nagle, L. M. (2014). The relation of critical care nurses' information-seeking behavior with perception of personal control, training, and non-routineness of the task. *Dynamics, 25*(1), 13–18.

O'Byrne, P., Bryan, A., Hendriks, A., et al. (2014). Social marginalization and internal exclusion: Gay men's understandings and experiences of community. *Canadian Journal of Nursing Research, 46*(2), 57–79.

Palys, T. (2008). Purposive sampling. In L. M. Given (Ed.). *The Sage encyclopedia of qualitative research methods* (Vol. 2, pp. 697–698). Los Angeles: Sage.

Paul-Savoie, E., Bourgault, P., Potvin, S., Gosselin, E., & Lafrenaye, S. (2018). The impact of pain invisibility on patient-centered care and empathetic attitude in chronic pain management. *Pain Research & Management*. https://doi.org/10.1155/2018/6375713.

Penz, K. L., Stewart, N. J., Karunanayake, C. P., Kosteniuk, J. G., & MacLeod, M. L. P. (2018). Competence and confidence in rural and remote nursing practice: A structural equation modelling analysis of national data. *Journal of Clinical Nursing, 28*, 1664–1679. https://doi.org/10.1111/jocn.14772.

Pesut, B., Thorne, S., Stager, M. L., Schiller, C. J., Penney, C., Hoffman, C., Greig, M., & Roussel, J. (2019). Medical assistance in dying: A review of Canadian nursing regulatory documents. *Policy, Politics & Nursing Practice, 20*(3), 113–130. https://doi.org/10.1177/1527154419845407.

Ridout, J., Aucoin, J., Browning, A., et al. (2014). Does perioperative documentation transfer reliability? *Computers, Informatics, Nursing, 32*, 37–42.

Rostad, H. M., Utne, I., Grov, E. K., Småstuen, M. C., Puts, M., & Halvorsrud, L. (2018). The impact of a pain assessment intervention on pain score and analgesic use in older nursing home residents with severe dementia: A cluster randomised controlled trial. *International Journal of Nursing Studies, 84*, 52–60. https://doi.org/10.1016/j.ijnurstu.2018.04.017.

Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing & Health, 18*, 179–183.

Santiago, C., Roza, D., Porretta, K., & Smith, O. (2019). The use of tablet and communication app for patients with endotracheal or tracheostomy tubes in the medical surgical intensive care unit: A pilot, feasibility study. *Canadian Journal of Neuroscience Nursing, 30*(1), 17–23. https://doi.org/10.13140/RG.2.2.19835.57125.

Sawatzky, R., Roberts, D., Russell, L., Bitschy, A., Ho, S., Desbiens, F-F., Chan, E. K. H., Tayler, C., & Stajduhar, K. (2019). Self-perceived competence of nurses and care aides providing a palliative approach in home, hospital, and residential care settings: A cross-sectional survey. *Canadian Journal of Nursing Research, 53*(1), 64–77. https://doi.org/10.1177/0844562119881043.

Sim, J., Saunders, B., Waterfield, J., & Kingstone, T. (2018). Can sample size in qualitative research be determined a priori? *International Journal of Social Research Methodology, 21*(5), 619–634. https://doi.org/10.1080/13645579.2018.1454643.

Tryphonopoulos, P. D., & Letourneau, N. (2020). Promising results from a video-feedback interaction guidance intervention for improving maternal–infant interaction quality of depressed mothers: A feasibility pilot study. *Canadian Journal of Nursing Research, 52*(2), 74–87. https://doi.org/10.1177/0844562119892769.

Turner-Bowker, D., Lamoureux, R. E, Stokes, J., Litcher-Kelly, L., Galipeau, N., Yaworsky, A., Solomon, J., & Shields, A. L. (2018). Informing a priori sample size estimation in qualitative concept elicitation interview studies for clinical outcome assessment instrument development. *Value in Health, 21*(7), 839–842. https://doi.org/10.1016/j.jval.2017.11.014.

Udod, S. A., Duchscher, J. B., Goodridge, D., Rotter, T., McGrath, P., & Hewitt, A. D. (2020). Nurse managers implementing the lean management system: A qualitative study in Western Canada. *Journal of Nursing Management, 28*, 221–228. https://doi-org.ezproxy.library.yorku.ca/10.1111/jonm.12898.

Viechtbauer, W., Smits, L., Kotz, D., Budé, L., Spigt, M., Serroyen, J., & Crutzen, R. (2015). A simple formula for the calculation of sample size in pilot studies. *Journal of Clinical Epidemiology, 68*(11), 1375–1379. https://doi.org/10.1016/j.jclinepi.2015.04.014.

Yang, M., Jia, G., Sun, S., Ye, C., Zhang, R., & Yu, X. (2019). Effects of an online mindfulness intervention focusing on attention monitoring and acceptance in pregnant women: A randomized controlled trial. *Journal of Midwifery & Women's Health, 64*, 68–77. https://doi.org/10.1111/jmwh.12944.

CHAPTER 14

Data-Collection Methods

Mina D. Singh | Ramesh Venkatesa Perumal

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Define the types of data-collection methods used in nursing research.
- List the advantages and disadvantages of each of these methods.
- Compare how specific data-collection methods contribute to the strength of evidence in a research study.
- Critically evaluate the data-collection methods used in published nursing research studies.

KEY TERMS

biological measurement
closed-ended item
concealment
consistency
debriefing
external criticism
internal criticism
intervention
intervention fidelity
interview
Likert-type scale
measurement
objective
open-ended item
operational definition
operationalization
physiological measurement
questionnaire
reactivity
records or available data
scale
scientific observation
social desirability
systematic

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

QUANTITATIVE DATA COLLECTION

NURSES USE ALL OF THEIR SENSES when collecting data from the patients to whom they provide care. Nurse researchers also have many ways to collect information about their research participants. Both the data collected when they perform patient care and the data collected for the purpose of research are objective and systematic. Objective means that the data must not be influenced by the person who collects the information, and systematic means that the data must be collected in the same methodical way by each person involved in the collection procedure. The methods that researchers use to collect information about participants are the identifiable and repeatable operations that define the major variables being studied.

Operationalization is the process of translating the concepts of interest to a researcher into

observable and measurable phenomena. For example, in their study on the nursing students' perceived self-efficacy and generation of medication errors with the use of an electronic medication administration (eMAR) in clinical simulation, Chan et al. (2019) used Bandura's (1997) definition of self-efficacy, in that it refers to an individual's psychological conviction that he or she can successfully execute the behaviour required to produce the desired outcome. They also used the term *confidence* interchangeably with *self-efficacy* in the study.

This purpose of this chapter is to familiarize you with the various ways in which researchers collect information from and about participants. The chapter provides nurse readers with the tools for evaluating the selection, use, and practicality of the various ways to collect data.

MEASURING VARIABLES OF INTEREST

To a large extent, the success of a study depends on the quality of the data-collection methods chosen and employed. Researchers have many types of methods available for collecting information from participants in research studies. *Measurement* is a term used in quantitative research and is the assignment of numbers to objects or events according to rules; determining which measurement to use in a particular investigation may be the most difficult and time-consuming step in the study design. In addition, nurse researchers have an array of quality instruments with adequate reliability and validity (see Chapter 15). This aspect of the research process necessitates painstaking effort from the researcher. Thus, the process of evaluating and selecting the available tools to measure variables of interest is crucial for the potential success of the study. In this section, the selection of measures and the implementation of the data-collection process are discussed. An algorithm that influences a researcher's choice of data-collection methods is diagrammed in the Critical Thinking Decision Path.

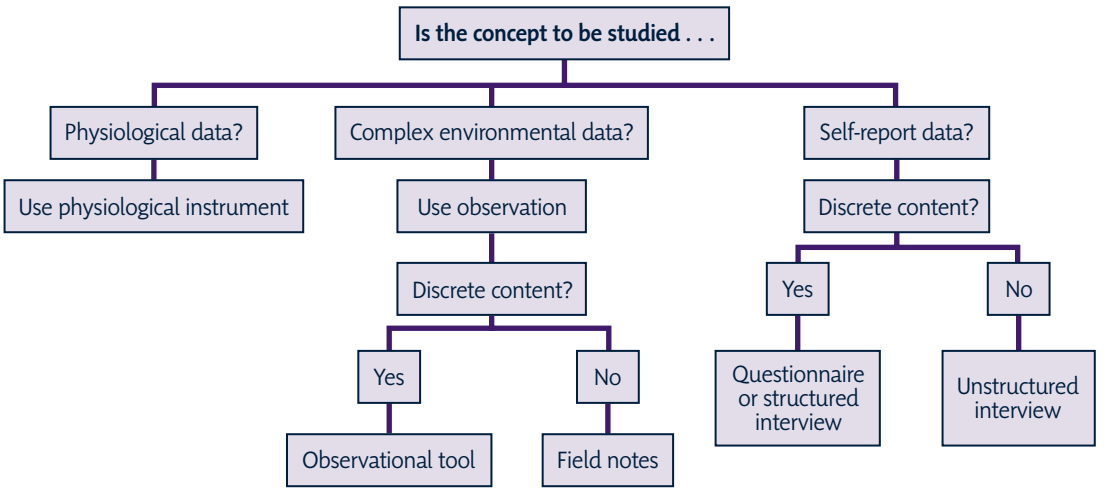
Information about phenomena of interest to nurses can be collected in many different ways. Nurses are interested in the biological and physical indicators of health (e.g., blood pressure and heart rate), but they are also interested in complex psychosocial questions presented by patients. Psychosocial variables, such as anxiety, hope, social support, and self-concept, may be measured by several different techniques, such as observation of behaviour, self-reports of feelings, or self-reports about attitudes in interviews or questionnaires. To study variables of interest, researchers also may use data that have already been collected for another purpose, such as records, diaries, or other media.

Selection of the data-collection method begins during the literature review. As noted in Chapter 5, one purpose of the literature review is to provide clues about instrumentation. As the literature review is conducted, the researcher begins to explore how previous investigators defined and operationalized variables similar to those of interest in the current study. The researcher uses this information to define conceptually the variables to be studied. Once a variable has been defined conceptually, the researcher returns to the literature to define the variable operationally—that is, describe how a concept is measured and what instruments are used to capture the essence of the variable. This operational definition translates the conceptual definition into behaviours or verbalizations that can be measured for the study. In this second literature review, the researcher searches for measurement instruments that might be used "as is" or adapted for use in the study. If instruments are available, the researcher must obtain the author's permission for their use.

The following examples illustrate the relationship of conceptual and operational definitions. Stress research is of interest to researchers from many disciplines, including nursing. Definitions of stressors may be psychological, social, or physiological. If researchers are interested in studying stressors, they must first define

CRITICAL THINKING DECISION PATH

Data-Collection Methods



what they mean by the concept of "stressor," both conceptually and operationally. Quality-of-life or health-related quality-of-life research is popular with researchers from many disciplines, including nursing.

Quality of life or health-related quality of life may also be interpreted in a general way (well-being) or be related specifically to a type of illness. Therefore, if researchers are interested in studying quality of life, they need first to define what they mean by the concept of "quality of life." For example, Oviedo et al. (2019) indicated that health-related quality of life is a subjective, multidimensional, integrative construct that includes individual physical and mental well-being (p. 988). If another researcher disagreed with this definition or was more interested in the quality of life of people with another specific illness or the quality of life of children, a different instrument might be more appropriate.

Sometimes no suitable measuring device exists, and so the researcher must then decide how important the variable is to the study and whether a new device should be constructed. The construction of new instruments for data collection that have reasonable reliability and validity (see Chapter 15) is a difficult task. If no suitable measuring device exists, the researcher may decide not to study a variable, or the researcher may decide to invest time and energy in instrument development. Either decision is acceptable, depending on the goals of the study and the goals of the researcher.

Research Hint

Remember that the researcher may not always present complete information about the way the data were collected, especially when established tools were used. To learn about the tool that was used, the reader may need to consult the original article that described the use or development of the tool.

Whether the researcher uses available methods or creates new ones, once the variables have been operationally defined in a manner consistent with the aims of the study, the population to be studied, and the setting, the researcher decides how the data-collection phase of the study will be implemented. This decision concerns how the

instruments for data collection will be given to the participants. Consistency is the most important issue in this phase.

Consistency in data collection means that the method used to collect data from each participant in the study is exactly the same or as close to the same as possible. Consistency can minimize the bias introduced when more than one person collects the data. Data collectors must be carefully trained and supervised. To ensure consistency in data collection, sometimes referred to as intervention fidelity, researchers must train data collectors in the methods to be used in the study so that each data collector acquires the information in the same way. Information about how to observe, ask questions, and collect data often is included in a kind of "cookbook" protocol or manual for the research project. A researcher needs to spend time developing the protocol and training data collectors to gather data systematically and reliably. Comments about their training and the consistency with which they collected data for the study should be provided by the researcher. An index of agreement called the *kappa statistic* is used to measure interrater agreement, where researchers will indicate the range of agreement expressed as a percentage of agreement among raters or observers or as a coefficient of agreement that considers the element of chance (coefficient kappa).

An example of intervention fidelity is given in the study by Boitor et al. (2019), in which they designed an intervention to evaluate the effects of hand massage on the pain intensity and pain-related interference in cardiac surgery patients. One interventionist delivered all hand massages and hand holdings except for one that was given by the research coordinator. The interventions were standardized across participants. The interventionist was a registered nurse trained in massage therapy through an accredited workshop of 6 hours by a professional therapist (p. 574).

Another example of the importance of training data collectors appears in the study by Arnstein et al. (2019) where they were validating a new pain scale on hospitalized adults. Data collectors completed initial training that included a full discussion of the study and research questions, ethical responsibilities in data collection, and the data-collection protocol. The principle investigator or one of the co-investigators met with the data collectors each month during data collection to determine if retraining is necessary.

Evidence-Informed Practice Tip

It is difficult to place confidence in a study's findings if the data-collection methods are not consistent.

TYPES OF DATA-COLLECTION METHODS

In general, data-collection methods can be divided into the following five types: *physiological measurements, observational methods, interviews* and *questionnaires,* and *records or available data.* Each method has a specific purpose, as well as certain advantages and disadvantages inherent in its use. In the following sections, these data-collection methods are discussed, along with their respective uses and problems.

Physiological or Biological Measurements

In everyday practice, nurses collect physiological data about patients, such as their temperature, pulse rate, blood pressure, blood glucose level, urine specific gravity, and pH of bodily fluids. Such data are frequently useful to nurse researchers. Because physiological variables, such as cardiac output and blood pressure, can be measured in several different ways, researchers need to measure these outcomes at similar intervals and in similar ways for all participants of the study. An example of a study using a physiological variable is that by Wang et al. (2018) who used HbA1c and blood glucose levels in their study on a nurse-led smartphone-based self-management programme for poorly controlled type 2 diabetes.

Physiological measurement and biological measurement involve the use of specialized

equipment to determine the physical and biological status of participants. Frequently, such measurements also require specialized training. These measurements can be *physical*, such as weight or temperature; *chemical*, such as blood glucose level; *microbiological*, as with cultures; or *anatomical*, as in radiological examinations. What distinguishes these measurements from others used in research is that special equipment is needed to make the observation. A researcher can say, "This participant feels warm," but to determine how warm the participant is requires the use of a sensitive instrument: a thermometer.

The advantages of using physiological data-collection methods include their objectivity, precision, and sensitivity. Such methods are generally considered to yield objective findings because unless a technical malfunction occurs, two readings of the same instrument taken at the same time by two different nurses are likely to yield the same result. Because such instruments are intended to measure the variable being studied, they offer the advantage of being precise and sensitive enough to pick up subtle variations in the variable of interest. Also, the deliberate distortion of physiological information by a participant in a study is highly unlikely to occur.

Physiological measurements are not without inherent disadvantages, however. Some instruments, if not available through a hospital, may be quite expensive to obtain and use. In addition, the accurate use of such instruments often necessitates specialized knowledge and training. Another problem with physiological measurements is that simply by using them, the variable of interest may be changed. Although some researchers think of these instruments as being nonintrusive, the presence of some types of devices might change the measurement. For example, the presence of a heart rate monitoring device might make some patients anxious and thereby increase their heart rate. In addition, nearly all types of measuring devices are affected in some way by the environment. Even a simple thermometer can be affected by the participant's drinking something hot or cold immediately before the temperature is taken. Thus, when assessing studies that use physiological measurements you need to consider whether the researcher controlled such environmental variables in the study. Finally, a physiological way to measure the variable of interest may not exist. On occasion, researchers try to force a physiological parameter into a study in an effort to increase the precision of measurement. If the device does not measure the variable of interest, however, the validity of the device's use is suspect.

Observational Methods

Although observing the environment is a normal part of living, scientific observation places a great deal of emphasis on the objective and systematic nature of the observation. The researcher is not merely watching what is happening but is watching with a trained eye for certain specific events. Scientific observation fulfills the following four conditions:

1. The observations undertaken are consistent with the study's specific objectives.
2. A standardized and systematic plan exists for the observation and the recording of data.
3. All of the observations are checked and controlled.
4. The observations are related to scientific concepts and theories.

Observation is particularly suitable as a data-collection method in complex research situations that are best viewed as total entities and that are difficult to measure in parts, such as studies dealing with the nursing process, parent–child interactions, or group processes (see Practical Application box for an example). In addition, observational methods can be the best way to operationalize some variables of interest in nursing research studies, particularly individual characteristics and conditions, such as traits and symptoms; verbal and nonverbal communication behaviours, activities, and skill attainment; and environmental characteristics.

Practical Application

Coker et al. (2017) conducted an observational study related to oral care interventions by nurses to hospitalized older people. Nurses were shadowed (one per evening) during their evening care encounters with their assigned patients. In addition to being observed, nurses were engaged in conversation during the observations. The research observer did not use a structured observation checklist but rather described the interventions that were provided.

Observational methods can also be distinguished by the role of the observer. This role is determined by the amount of interaction between the observer and the people being observed. Each of the following four basic types of observational roles is distinguishable by the amount of concealment or intervention implemented by the observer:

1. Concealment without intervention
2. Concealment with intervention
3. No concealment without intervention
4. No concealment with intervention

These methods are illustrated in Figure 14.1; examples are given later. Concealment refers to a study method in which participants do not know that they are being observed; through intervention, the observer provokes actions from those who are being observed.

Observational studies commonly involve no concealment and no intervention. In this case, the researcher obtains informed consent from the participant to be observed and then simply observes the participant's behaviour.

When a researcher is concerned that the participants' behaviour will change as a result of being observed (reactivity), the type of observation most commonly employed is that of concealment without intervention. In this case, the researcher watches the participants without their knowledge of the observation and does not provoke them into action. Often, such concealed observations involve the use of hidden television cameras, audio recordings, or one-way mirrors. Concealment without intervention is often used in observational studies of children. You may be familiar with rooms with one-way mirrors through which a researcher can observe the behaviour of the occupants of the room without being observed by them. Such studies allow the observation of children's natural behaviour and are often used in developmental research. Observing participants without their knowledge may violate assumptions of informed consent; therefore, researchers face ethical problems with this type of approach. However, researchers sometimes have no other way to collect such data, and the data collected are unlikely to have negative consequences for the participant. In these cases, the disadvantages of the study are outweighed by the advantages. Furthermore, the problem of consent is often handled by informing participants after the observation and allowing them the opportunity to refuse to have their data included in the study and to discuss any questions they might have. This process is called debriefing.

		Concealment: Yes	Concealment: No
Intervention	Yes	Researcher hidden An intervention	Researcher open An intervention
	No	Researcher hidden No intervention	Researcher open No intervention

FIG. 14.1 Types of observational roles in research.

When the observer is neither concealed nor intervening, the ethical question is not a problem. Here, the observer makes no attempt to change the participants' behaviour and informs them that they are to be observed. Because the observer is present, this type of observation allows a greater depth of material to be studied than if the observer is separated from the participants by an artificial barrier, such as a one-way mirror. In a commonly used observational technique, the researcher functions as part of a social group to observe the participants. For example, in their study, Coker et al. (2017) used unconcealed observation, with the nurses and patients giving full consent for participation in the study. The problem with this type of observation, however, is reactivity (also referred to as the Hawthorne effect; see Chapter 10), or the distortion created when the participants change behaviour because they know they are being observed.

No concealment with intervention is used when the researcher is observing the effects of an intervention introduced for scientific purposes. Because the participants know they are participating in a research study, few problems with ethical concerns occur, but reactivity is a problem with this type of study.

Concealed observation with intervention involves staging a situation and observing the behaviours that are evoked in the participants as a result of the intervention. Because the participants are unaware of their participation in a research study, this type of observation has fallen into disfavour and is rarely used in nursing research.

Observational methods may be structured or unstructured. Unstructured observational methods are not characterized by a total absence of structure but rather usually involve collecting descriptive information about the topic of interest. In unstructured observations, the observer keeps field notes that record the activities, as well as the observer's interpretations of these activities. Field notes are usually not restricted to any particular type of action or behaviour; rather, they are intended to depict a social situation in a more general sense.

Another type of unstructured observation is the use of stories or anecdotes, which usually focus on the behaviours of interest and frequently add to the richness of research reports by illustrating a particular point.

The use of structured observations without a standardized tool involves specifying in advance what behaviours or events are to be observed and preparing forms for record keeping, such as categorization systems, checklists, and rating scales. Whichever system is employed, the observer watches the participant and then marks on the recording form what was seen. In both cases, the observations must be similar among the observers (see the earlier discussion and Chapter 15 for an explanation of interrater reliability). Thus, observers need to be trained to be consistent in their observations and ratings of behaviour.

Evidence-Informed Practice Tip

When you read a research report that uses observation as a data-collection method, you will want to note evidence of consistency across data collectors through use of internal consistency reliability data in quantitative research and credibility in qualitative research. When that evidence is present, you can have greater confidence in the results.

Scientific observation has several advantages as a data-collection method. The main advantage is that observation may be the only way for the researcher to study the variable of interest. For example, what people say they do is often not what they really do. Therefore, if the study is designed to obtain substantive findings about human behaviour, observation may be the only way to ensure the validity of the findings. In addition, no other data-collection method can match the depth and variety of information that can be collected with the techniques of scientific observation. Such techniques are also flexible in that they may be used in both experimental and nonexperimental designs and in laboratory and field studies.

Research Hint

Sometimes researchers carefully train observers or data collectors, but the research report does not address this training. The limitations on length of research reports often prevent the inclusion of certain information. Readers can often assume that if reliability data are provided, then appropriate training occurred.

As with all data-collection methods, observation also has its disadvantages. Earlier in this chapter, the problems of reactivity and ethical concerns were mentioned with regard to concealment and intervention. In addition to these problems, data obtained by observational techniques are vulnerable to the bias of the observer. Emotions, prejudices, and values can influence the way that behaviours and events are observed. In general, the more the observer needs to make inferences and judgements about what is being observed, the more likely it is that distortion will occur. Thus, in judging the adequacy of observational methods, you will need to consider how observational tools were constructed and how observers were trained and evaluated.

Interviews, Surveys, and Questionnaires

Participants in a research study often have information that is important to the study and that can be obtained only by asking the participants. Such questions may be asked through the use of interviews, surveys, and questionnaires. For both, the purpose is to ask participants to report data for themselves, but each method has unique advantages and disadvantages. The interview is a method of data collection in which a data collector questions a participant verbally. Interviews may be face to face or performed over the telephone, e-mail, video-conferencing, or other electronic means and may consist of open-ended or closed-ended questions. In contrast, both surveys and questionnaires are instruments designed to gather data from individuals about knowledge, attitudes, beliefs, and feelings. Surveys and questionnaires are used interchangeably, but there is a difference. A survey can include both a questionnaire and/or an interview, whereas a questionnaire is a written set of questions within the survey aimed at getting specific information about individuals. Surveys are also thought of being inclusive of both the questionnaire and the process of collecting, and analyzing and forecasting of a problem, attitude, or opinion. Survey research relies almost entirely on questioning participants with either interviews or questionnaires, but these methods of data collection can also be used in other types of research (see also Chapter 12).

No matter what type of study is conducted, the purpose of questioning participants is to seek information. This information may be of either direct interest, such as the participant's age, or indirect interest, such as when the researcher uses a combination of items to estimate the degree to which the respondent has a particular trait or characteristic. An intelligence test is an example of how individual items are combined with several others to develop an overall scale of intelligence. When items of indirect interest on a survey or questionnaire are combined to obtain an overall score, the measurement tool is called a scale.

The investigator determines the content of an interview or questionnaire from the literature review (see Chapter 5). When evaluating interviews and questionnaires, you should consider the content of the scale, the individual items, and the order of the items. The basic standard for evaluating the individual items in an interview or questionnaire is that the item must be clearly written so that the intention of the question and the nature of the information sought are clear to the respondent. The only way to know whether the questions are understandable to the respondents is to pilot test them in a similar population. It is also critical not to rely on only the instrument developer's reports of reliability and validity (see Chapter 15). A pilot test allows researchers to test the reliability and validity for their unique sample rather than relying only on previously reported results.

Although each questionnaire item must consist of only one question or concept, be free of suggestions, and be worded with correct grammar, such items may be either open-ended or closed-ended. An open-ended item is used when the researcher wants the participants to respond in their own words or when the researcher does not know all of the possible alternative responses. A closed-ended item is a question that the respondent may answer with only one of a fixed number of alternative responses. Many scales use a fixed-response format called a Likert-type scale. A Likert-type scale is a list of statements for which responses are varying degrees of agreement or opinion—for example, whether respondents "strongly agree," "agree," "disagree," or "strongly disagree." Sometimes finer distinctions are given, or a neutral category (e.g., "no opinion") may be provided. The use of the neutral category, however, sometimes creates problems because it is often the most frequent response and is difficult to interpret. Fixed-response items also can be used for questions requiring a "yes" or "no" response or when the interview or questionnaire has categories, as with income.

Evidence-Informed Practice Tip

Scales used in nursing research should have evidence of adequate reliability and validity so that readers feel confident that the findings reflect what the researcher intended to measure (see Chapter 15).

Figure 14.2 shows a few items from a fictional survey of pediatric nurse practitioners. The first items are taken from a list of similar items, and they are both closed-ended and of a Likert-type format. Note that respondents are asked to choose how strongly they agree with each item. In using these questions in the survey, respondents are forced to choose from only these answers because it is thought that these will be the only responses. The only possible alternative response is to skip the item, leaving it blank.

Sometimes researchers have no idea or only a limited idea of what the respondent will say, or researchers want the answer in the respondent's own words, as with the second (open-ended) set of items. In this situation, respondents may also leave the item blank but are not forced to make a particular response.

Interviews and questionnaires are commonly used in nursing research. Both are strong approaches to gathering information for research because they enable the researcher to approach the task directly. In addition, both can elicit certain kinds of information, such as the participants' attitudes and beliefs, that would be difficult to obtain without asking the participant directly.

All methods that involve verbal reports, however, share a problem with accuracy. Often, it is impossible to know whether what the researcher is told is indeed true. For example, people are known to respond to questions in a way that makes a favourable impression. This response style is known as social desirability, which can be regarded as resulting from two factors: self-deception and other-deception.

Neyerhof (2006) has discussed the two main modes of coping with social desirability bias. The first mode is aimed at the detection and measurement of social desirability bias and is represented by two methods: the use of social desirability scales and the rating of item desirability. The second mode is aimed at preventing or reducing social desirability bias and is represented by the following methods: forced-choice items, the randomized response technique, the bogus pipeline, self-administration of the questionnaire, the selection of interviewers, and the use of proxy participants. Neyerhof found that no one method excelled completely and suggested that a combination of prevention and detection methods is the best strategy to reduce social desirability bias. Recent trends have led to the use of social desirability scales, as variables, within studies to determine the extent of social desirability. The most popular of these scales used to be the Marlowe-Crowne Social Desirability scale (MCSDS) but now the current gold standard is the Balanced Inventory of Desirable Responding (BIDR),

Closed-Ended (Likert-Type Scale)

A. How satisfied are you with your current position?

1. Very satisfied
2. Moderately satisfied
3. Undecided
4. Moderately dissatisfied
5. Very dissatisfied

B. To what extent do the following factors contribute to your current level of positive satisfaction?

	Not at all	*Very little*	*Somewhat*	*Moderate amount*	*A great deal*
1. % of time in patient care	1	2	3	4	5
2. Type of patients	1	2	3	4	5
3. % of time in educational activity	1	2	3	4	5
4. % of time in administration	1	2	3	4	5

Closed-Ended

A. On average, how many patients do you see in one day?

1. 1 to 3
2. 4 to 6
3. 7 to 9
4. 10 to 12
5. 13 to 15
6. 16 to 18
7. 19 to 20
8. More than 20

B. How would you characterize your practice?

1. Too slow
2. Slow
3. About right
4. Busy
5. Too busy

Open-Ended

A. Are there incentives that the Canadian Nurses Association ought to provide for members that are not currently being provided?

_____

_____

_____

_____

_____

FIG. 14.2 Examples of closed-ended and open-ended questions.

which is considered more sensitive because its development incorporated newer theoretical and empirical of Social Desirability Responding (SDR) and more sophisticated multivariate techniques (Lambert et al., 2016).

Questionnaires and interviews also have some specific purposes, advantages, and disadvantages. Questionnaires are useful tools when the purpose is to collect information. If questionnaires are too long, however, respondents are not likely to complete them. Questionnaires are most useful when the set of questions to be asked is finite and the researcher can be assured of the clarity and specificity of the items. Face-to-face techniques or interviews are most appropriate when the researcher may need to clarify the task for the respondent or is interested in obtaining more personal information from the respondent. Telephone interviews allow the researcher to reach more respondents than face-to-face interviews and provide more clarity than questionnaires.

Research Hint

Remember that sometimes researchers make trade-offs when determining the measures to be used. For example, if a researcher wants to learn about an individual's attitudes regarding practice, and practicalities preclude using an interview, a questionnaire may be used instead.

Pauly and colleagues (2015) used both unconcealed observations and in-depth interviews to understand what constitutes culturally safe care for people who use illicit drugs. Both the nurses and patients were asked questions about their experiences in giving and receiving care, their understanding of comfort and safety, and their experiences of health care settings and any barriers and enablers in providing and receiving care (p. 124). Two researchers conducted 275 hours of unconcealed observations to gain an understanding of the various contexts under which nurses provide care to persons who use illicit drugs. This use of multiple measures provides a more complete picture than the use of just one measure.

When determining whether to use interviews or questionnaires, researchers often face difficult choices. The final decision is based on the instruments available and their relative costs and benefits.

Both face-to-face and telephone interviews have some advantages over questionnaires. The rate of response to interviews is almost always better than that to questionnaires, which helps eliminate bias in the sample (see Chapter 13). Respondents seem to be less likely to hang up the telephone or to close the door in an interviewer's face than to throw away a questionnaire. Another advantage of the interview is that some people—such as young children, people with visual impairments, and people who are illiterate—cannot fill out a questionnaire but can participate in an interview. With an interview, the data collector knows who is giving the answers. When questionnaires are mailed, for example, anyone in the household could be the person who supplies the answers.

Interviews also allow for some safeguards to be built into the interview situation. Interviewers can clarify misunderstood questions and observe the level of the respondent's understanding and cooperativeness. In addition, the researcher has strict control over the order of the questions. With questionnaires, the respondent can answer questions in any order. Changing the order of the questions can sometimes change the response.

Finally, interviews allow for richer and more complex data to be collected. The interview questioning can be open-ended or closed-ended; in either case, interviewers can probe to understand why a respondent answered in a particular way. In the qualitative study by Woodgate et al. (2020), open-ended, face-to-face interviews were conducted to gather the experiences of youth living with anxiety. Two interview sessions were carried out. The first interview session opened with the question, "Can you please tell me a little bit about yourself?" This was followed up with questions to get at the youth's story (p. 4). Semi-structured interviews were used by Currie and Szabo (2020) in their study to explore the parents' experience

of caring for medical and social care needs for children with rare neurodevelopmental disorders.

Interviews can also be conducted in a group setting, sometimes called a *focus group approach or focus group interview*, which may include about six to eight participants. Bottorff, Haines-Saah, Oliffe, and colleagues (2014) used a semi-structured interview schedule to guide focus group discussions to engage youth in discussing the merits and limitations of a variety of sample messages related to smoking and breast cancer and in generating ideas to guide youth-friendly message development and delivery media. Eight semi-structured focus groups were held outside school hours in community locations over a period of four months. Two focus groups were held with each of the following groups: First Nations and Métis girls, non-Indigenous girls, First Nations and Métis boys, and non-Indigenous boys. The aim was to capture diversity of opinion within each subgroup and to meet target participant numbers. These small-group interviews allowed the participants to freely explain and share information individually and collectively. Agreement and disagreement among participants may be elicited, which allows the researchers to obtain specific information from a number of participants efficiently and simultaneously.

Questionnaires are much less expensive to administer than interviews because interviews may require the hiring and training of interviewers. Thus, if a researcher has a fixed amount of time and money, a larger and more diverse sample can be obtained with questionnaires. Questionnaires also provide complete anonymity, which may be important if the study deals with sensitive issues. Finally, the fact that no interviewer is present assures the researcher and the reader that no interviewer bias will occur. Interviewer bias occurs when the interviewer unwittingly leads the respondent to answer in a certain way. This problem is especially pronounced in studies with unstructured interview formats. A subtle nod of the head, for example, could lead a respondent to change an answer to correspond with what he or she perceives that the researcher wants to hear.

For instance, McDonald et al. (2018) converted their questionnaire, Student Adaptation to College Questionnaire, to an online format for easier access for both the participants for completion and the researchers for data analysis.

In another study, Covell et al. (2017) mailed an invitation letter to participate in their study with the questionnaire to 13,748 internationally educated nurses. There were three reminders to increase response rate. Participants could complete the questionnaire anonymously online or by postal mail (p. 4).

Records or Available Data

All of the data-collection methods discussed thus far concern the ways that nurse researchers gather new data to study phenomena of interest. Not all studies, however, require a researcher to acquire new information. Existing information can sometimes be examined in a new way to study a problem. The use of records and available data is sometimes considered to be primarily the concern of historical research, but hospital records, care plans, and existing data sources (e.g., the census) are frequently used for collecting information. What sets these studies apart from a literature review is that these available data are examined in a new way and not merely summarized; they also answer specific research questions.

Records or available data, then, are forms of information that are collected from existing materials, such as hospital records, historical documents, or audio or video recordings, and are used to answer research questions in a new manner. For example, Kaasalainen et al. (2019) conducted a chart audit to examine current rates of resident deaths, emergency department use within the last year of life, and hospital deaths for long-term care residents. Because the data-collection step of the research process is often the most difficult and time consuming, the use of available records often produces a significant

saving of time. If the records have been kept in a similar manner over time, analysis of these records allows examination of trends over time. In addition, the use of available data decreases problems of reactivity and response set bias. The researcher also does not have to ask individuals to participate in the study.

However, institutions are sometimes reluctant to allow researchers access to their records. If the records are kept so that an individual cannot be identified, access for research purposes is usually not a problem. Also, the Privacy Act, a federal law, protects the rights of individuals who may be identified in records, which would be a violation of anonymity.

One problem that affects the quality of available data concerns survival of records. If the records available are not representative of all of the possible records, the researcher may have a problem with bias. Often, because researchers have no way to tell whether the records have been saved in a biased manner, they need to make an intelligent guess as to their accuracy. For example, a researcher might be interested in studying socioeconomic factors associated with the suicide rate. These data frequently are underreported because of the stigma attached to suicide, and so the records would be biased. Recent interest in computerization of health records has led to an increase in the discussion about the desirability of access to such records for research. At this time, how much of such data will continue to be readily available for research without consent is unclear.

Another problem is related to the authenticity of the records. The distinction of primary and secondary sources is as relevant in this discussion as it was in the discussion of the literature review to determine the source of the work (see Chapter 5). A book, for example, may have been ghostwritten, but all credit was accorded to the known author. The researcher may have a difficult time ferreting out these subtle types of biases.

Lastly, existing records may be missing a significant amount of data. For example, years of education may be recorded on only a portion of the sample records. Nonetheless, records and available data constitute a rich source of data for study.

ONLINE AND COMPUTERIZED METHODS OF DATA COLLECTION

With the fast-paced progression of the Internet and computer technology, many researchers are using online data collection. The information obtained can be quantitative or qualitative, closed-ended or open-ended. This method of data collection can take the form of Web-based surveys or data input directly into microcomputers. For example, Corby et al. (2021) used an online survey hosted by Fluid Surveys in their study to investigate the predictors of prenatal breastfeeding self-efficacy.

Many online survey tools, such as SurveyMonkey or QuestionPro or Fluid Surveys, are available; a survey can be downloaded quickly and the results obtained for a small fee. The advantages of this method are that it is anonymous and inexpensive; respondents can fill out the survey in their own time; a large number of participants can be accessed; respondent time is reduced; data-collection time is reduced; duplicate responses can be identified; and, for the researcher, implementation is time efficient. The disadvantages are that not everyone has access to a computer or is computer literate, the response rates may be low, and a large amount of data may be missing. In addition, the researcher has to ensure that the cloud where the data is collected and stored is in Canada and follow the guidelines for Personal Health Information Protection Act PHIPA) (Information and Privacy Commissioner of Ontario, 2015).

Computerized data collection can be accomplished through the use of laptop computers or electronic tablets or smartphones. Researchers can input their data directly into these handheld microcomputers. The data can then be transferred to a larger computer for analysis.

Evidence-Informed Practice Tip

A critical evaluation of any data-collection method includes evaluating the appropriateness, objectivity, consistency, and credibility of the method employed.

CONSTRUCTION OF NEW INSTRUMENTS

As already mentioned in this chapter, researchers sometimes cannot locate an existing instrument or method with acceptable reliability and validity to measure the variable of interest. This situation is often the case when part of a nursing theory is tested or when the effect of a clinical intervention is evaluated. For example, Vincelette et al. (2019) developed, validated, and assessed the psychometric properties of a survey related to cardiopulmonary resuscitation among intensive care nurses (see Chapter 15).

Instrument development is complex and time consuming, however. It consists of the following steps:

- Defining the construct to be measured
- Formulating the items (questions)
- Assessing the items for content validity
- Developing instructions for respondents and users
- Pretesting and pilot testing the items
- Estimating reliability and validity

Defining the construct (concepts at a higher level of abstraction) to be measured requires that the researcher develop an expertise in the construct, which necessitates an extensive review of the literature and of all tests and measurements that deal with related constructs. The researcher uses all this information to synthesize the available knowledge so that the construct can be defined.

Once the construct is defined, the individual items for measuring the construct can be developed. The researcher will develop many more items than are needed to address each aspect of the construct or subconstruct. A panel of experts in the field evaluates the items so that the researcher is assured that the items measure what they are intended to measure (content validity; see Chapter 15). Eventually, the number of items is decreased because some items will not elicit the intended information and will be dropped. In this phase, the researcher needs to ensure consistency both among the items and in testing and scoring procedures.

Finally, the researcher administers or pilot tests the new instrument by applying it to a group of people who are similar to those who will be studied in the larger investigation. The purpose of this analysis is to determine the quality of the instrument as a whole (reliability and validity) and the ability of each item to discriminate individual respondents (variance in item response). The researcher also may administer a related instrument to see whether the new instrument is sufficiently different from the older one.

It is important that researchers who invest significant time in tool development publish their results. For example, Hart et al. (2019) were interested in understanding the surgical neonatal nursing workload. From their literature review, they determined that there was no tool that adequately describes the nursing workload associated with neonates. They decided to modify a validated NICU nursing tool to better meet the needs of the surgical NICU patients. This type of research serves not only to introduce other researchers to the tool but also to ultimately enhance the field, inasmuch as the ability to conduct meaningful research is limited only by the ability to measure important phenomena.

Research Hint

Determine whether a newly developed survey or questionnaire was pilot tested to obtain preliminary evidence of reliability and validity.

QUALITATIVE DATA COLLECTION

In qualitative research, data collection is more flexible and may evolve over the course of the study. Some of the data-collection methods outlined previously are also used in qualitative research, such as observations and semi-structured interviews. For example, Pesut et al. (2020) used

semi-structured interviews in the qualitative study to better understand nursing practice within the legislative approaches to medically assisted dying. In addition, other methods, such as focus groups and photovoice, are used, or textual data is gathered from media or policy documents.

Focus Groups

A focus group is a type of interview of about five to eight people on the topic of interest. The interviewer has predetermined questions with probes, in the event that the group is not forthcoming with information. The setting for this interview is usually a neutral one. Most qualitative researchers use voice recorders so that they can be sure that they have captured what the participant says. This reduces the need to write things down and frees up the researchers to listen fully. Interview recordings are usually transcribed verbatim and then listened to for accuracy. In a research report, investigators describe their procedures for collecting the data, such as obtaining informed consent, all the steps from initial contact to the end of the study visit, and how long each interview or focus group lasted or how much time the researcher spent "in the field" collecting data. For example, in the qualitative portion of their mixed method study, Verkuyl and Hughes (2019) used focus groups to explore students' experiences and outcomes using virtual gaming simulation as well as to describe and expand on the findings from the quantitative data (p. 11).

Photovoice

Photography has been used in research since the 1950s, as photographs provide a permanent record of events and activities. In the early 1990s, Dr. Caroline Wang developed *photovoice*, an innovative approach used in participatory action research (Wang, 1999) in which interviews are stimulated and guided by photographs. These photographs empower members of marginalized groups to work together to "identify, represent and enhance their community through a specific photographic technique" (Wang & Burris, 1997). They also aid in breaking down barriers between researchers and participants. Participants use photographs, which act as prompts, to help others to see their world, and stories are told while discussing the photographs; this can be empowering to the individual. Photovoice requires that community members take on multiple roles, such as photographer, key informant, and co-researcher.

Scruby et al. (2019) used photovoice to explore social support, sport participation, and rural women's health. Participants were asked to photograph images they felt represented health in the context of the curling rink; these images were analyzed thematically.

APPRAISING THE EVIDENCE

Data-Collection Methods

Evaluating the adequacy of data-collection methods from written research reports is often problematic for new nursing research readers. Because the tool itself is not available for inspection, you may not feel comfortable judging the adequacy of the method without seeing it. However, you can ask questions to judge the method chosen by the researcher. These questions are listed in the Critiquing Criteria box.

In all studies, data-collection methods should be clearly identified. The conceptual and operational definitions of each important variable should be present in the report. Sometimes it is useful for the researcher to explain why a particular method was chosen. For example, if the study dealt with young children, the researcher may explain that a questionnaire was deemed to be an unreasonable task, and so an interview was chosen.

APPRAISING THE EVIDENCE—*cont'd*

Data-Collection Methods

Once you have identified the method chosen to measure each variable of interest, you should decide whether the method used was the best way to measure the variable. For example, if a questionnaire was used, you might wonder why the researcher decided not to use an interview. Also consider whether the method was appropriate to the clinical situation. Does it make sense to interview patients in the recovery room, for example?

Once you have decided whether all relevant variables are operationalized appropriately, you can begin to determine how well the method was carried out. For studies involving physiological measurement, determine whether the instrument was appropriate to the problem and not forced to fit it. The rationale for selecting a particular instrument should be given. For example, it may be important to know that the study was conducted under the auspices of a manufacturing firm that provided the measuring instrument. In addition, the researcher should have made provisions to evaluate the accuracy of the instrument and the skill level of the people who used it.

Several considerations are important when you read studies that involve observational methods. Who were the observers, and how were they trained? Is there any reason to believe that different observers perceived events or behaviours differently? Remember that the more inferences the observers are required to make, the more likely it is that observations will be biased. Also, consider the problem of reactivity: In any observational situation, it is possible that the mere presence of the observer will cause the participant to change the behaviour in question. Of importance is not that reactivity could occur but the extent to which reactivity could affect the data. Finally, consider whether the observational procedure was ethical. You need to consider whether the participants were informed that they were being observed, whether any intervention was performed, and whether the participants had agreed to be observed.

Interviews and questionnaires should be clearly described to allow the reader to decide whether the variables were adequately operationalized. Sometimes the researcher will reference the original report about the tool, and you may wish to read this study before deciding whether the method was appropriate for the current study. Also, the respondents' task should be clear. Thus, the researcher should have made provisions for the participants to understand both their overall responsibilities and the individual items of the interview or questionnaire. The following questions must be considered: Who were the interviewers in the interview situation? Does the researcher explain how they were trained to decrease any interviewer bias?

Available data, such as medical records, are subject to internal and external criticism. Internal criticism concerns the evaluation of the worth of the records and refers primarily to the accuracy of the data. The researcher should present evidence that the records are genuine. External criticism is concerned with the authenticity of the records. Are the records really written by the first author? The researcher may have a biased sample of all of the possible records in the problem area, which may have a profound effect on the validity of the results.

Once you have decided that the data-collection method used was appropriate for the problem and the procedures were appropriate for the population studied, the reliability and validity of the instruments themselves need to be considered. These characteristics are discussed in Chapter 15.

CRITIQUING CRITERIA

1. Is the framework for research clearly identified?

DATA-COLLECTION METHODS

1. Are all of the data-collection instruments clearly identified and described?
2. Is the rationale for their selection given?
3. Is the method used appropriate for the problem being studied?
4. Were the methods used appropriate for the clinical situation?
5. Are the data-collection procedures similar for all participants?
6. Were efforts made to ensure intervention fidelity through the data-collection protocol?

Continued

PHYSIOLOGICAL MEASUREMENT

1. Is the instrument used appropriate for the research problem and not forced to fit it?
2. Is a rationale given for why a particular instrument was selected?
3. Is there a provision for evaluating the accuracy of the instrument and the skill of the people who used it?

OBSERVATIONAL METHODS

1. Who conducted the observation?
2. Were the observers trained to minimize any bias?
3. Was an observational guide provided?
4. Were the observers required to make inferences about what they saw?
5. Is there any reason to believe that the presence of the observers affected the behaviour of the participants?
6. Were the observations performed according to the principles of informed consent?

INTERVIEWS/FOCUS GROUPS

1. Is the interview schedule described adequately enough for you to know whether it covers the purpose of the study?
2. Is it clear that the participants understood the task and the questions?
3. Who were the interviewers, and how were they trained?
4. Is any interviewer bias evident?

QUESTIONNAIRES

1. Is the questionnaire described well enough for you to know whether it covers the purpose of the study? Is evidence provided that participants were able to perform the task?
2. Is it clear that the participants understood the questionnaire?
3. Are the majority of the items appropriately closed- or open-ended?

AVAILABLE DATA AND RECORDS

1. Are the records used appropriate for the problem being studied?
2. Are the data examined in such a way as to provide new information and not summarize the records?
3. Has the author addressed questions of internal and external criticism?
4. Is there any indication of selection bias in the available records?

CRITICAL THINKING CHALLENGES

- Physiological measurements are objective, precise, and sensitive. Discuss factors that might influence their validity and feasibility.
- A student in research class asks why nurses who participate in a clinical research study in the role of a data collector or who perform a "treatment intervention" need to be trained. What important factors or rationale would you offer to support the establishment of interrater reliability?
- Observation is a data-collection method used frequently in nursing research. Discuss the factors that make nurses perfect potential candidates for this role and the disadvantages of using this method.
- Studies often use a survey to collect data. How can researchers increase their return rate for the survey, and how do they determine whether the survey return is adequate?

CRITICAL JUDGEMENT QUESTIONS

1. In a study conducted at a large long-term care facility, two data collectors examined the correct use of personal protective equipment (PPE) on 100 nurses. The examinations were independently performed on the same day. A comparison of the results indicated that the data collectors scored 90 of the 100 nurses following the correct protocol for wearing and removing PPE. What can be determined from this finding?
   a. Interrater reliability between the two data collectors was high.
   b. Interrater reliability between the two data collectors was low.
   c. The data-collection method was inappropriate for the phenomenon under investigation.
   d. In order to establish interrater reliability, the two data collectors should have examined each nurse's behaviour at the same time.
2. Which of the following is an example of a physiological measurement?
   a. Definition of a type "A" behaviour pattern
   b. Description of self-care behaviour abilities in patients with dementia.
   c. HbA1C blood levels
   d. Adjusted scores on the State-Trait Anxiety Scale

3. Which data-collection method is most appropriate for measuring postpartum depression?
   a. Assessment of estrogen levels
   b. Unstructured interview
   c. Visual analog pain scale
   d. Edinburgh Postnatal Depression Scale

KEY POINTS

- Data-collection methods are described as being both objective and systematic. The data-collection methods of a study provide the operational definitions of the relevant variables.
- Types of data-collection methods include physiological measurements, observational methods, interviews, questionnaires, and records or available data. Each method has advantages and disadvantages.
- Physiological measurements are the methods in which technical instruments are used to collect data about patients' physical, chemical, microbiological, or anatomical status. These methods are suited to studying how to improve the effectiveness of nursing care. Physiological measurements are objective, precise, and sensitive, but they may be very expensive and may distort the variable of interest.
- Observational methods are used in nursing research when the variables of interest deal with events or behaviours. Scientific observation requires preplanning, systematic recording, controlling the observations, and determining the relationship to scientific theory. This method is best suited to research problems that are difficult to view as part of a whole. Observers may be required to perform or not perform interventions, and their activity may be concealed or obvious.
- Observational methods have several advantages: (1) they provide flexibility to measure many types of situations, and (2) they enable a great depth and breadth of information to be collected.
- Observation has disadvantages as well: (1) data may be distorted as a result of the observer's presence (reactivity), (2) concealment requires the consideration of ethical issues, and (3) data from observations may be biased by the person who is doing the observing.
- Interviews are data-collection methods commonly used in nursing research. Items on interview schedules may be of direct or indirect interest. Participants may be asked either open-ended or closed-ended questions. The form of the question should be clear to the respondent, free of suggestion, and grammatically correct.
- Questionnaires, or surveys, are useful when the number of questions to be asked is finite. The questions need to be clear and specific. Questionnaires are less costly and less time consuming to administer to large groups of participants, particularly if the participants are geographically widespread. Questionnaires also can be completely anonymous and prevent interviewer bias.
- Interviews are most appropriate when a large response rate and an unbiased sample are important because the refusal rate for interviews is much lower than that for questionnaires. Interviews enable the participation of people who cannot use a questionnaire, such as children and people who are illiterate. An interviewer can clarify and maintain the order of the questions for all participants.
- Records or available data are also an important source of research data. The use of available data may save the researcher considerable time and money in conducting a study. This method reduces problems with both reactivity and ethical concerns. However, records and available data are subject to problems of availability, authenticity, and accuracy.
- A critical evaluation of data-collection methods should emphasize the appropriateness, objectivity, and consistency of the method employed.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Arnstein, P., Gentile, D., & Wilson, M. (2019). Validating the functional pain scale for hospitalized adults. *Pain Management Nursing, 20*(5), 418–424. https://doi.org/10.1016/j.pmn.2019.03.006.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.

Bottorff, J. L., Haines-saah, R., Oliffe, J. L., Struik, L. L., Bissell, L. J. L., Richardson, C. P., Gotay, C., Johnson, K. C., Hutchinson, P., Bottorff, J. L., Haines-saah, R., Oliffe, J. L., Struik, L. L., Bissell, L. J. L., Richardson, C. P., Gotay, C., Johnson, K. C., & Hutchinson, P. (2014). Designing tailored messages about smoking and breast cancer : A focus group study with youth. *Canadian Journal of Nursing Research, 46,* 66–86.

Chan, R., Booth, R., Strudwick, G., & Sinclair, B. (2019). Nursing students' perceived self-efficacy and the generation of medication errors with the use of an electronic medication administration record (eMAR) in clinical simulation. *International Journal of Nursing Education Scholarship, 16*(1). https://doi.org/10.1515/ijnes-2019-0014.

Coker, E., Ploeg, J., Kaasalainen, S., & Carter, N. (2017). Observations of oral hygiene care interventions provided by nurses to hospitalized older people. *Geriatric Nursing, 8*(1), 17–21. https://doi.org/10.1016/j.gerinurse.2016.06.018.

Corby, K., Kane, D., & Dayus, D. (2021). Investigating predictors of prenatal breastfeeding self-efficacy. *Canadian Journal of Nursing Research, 53*(1), 56–63. https://doi.org/10.1177/0844562119888363.

Covell, C. L., Primeau, M.-D., Kilpatrick, K., & St-Pierre, I. (2017). Internationally educated nurses in Canada: Predictors of workforce integration. *Human Resources for Health, 15*(1), 1–16. https://doi.org/10.1186/s12960-017-0201-8.

Currie, G., & Szabo, J. (2020). Social isolation and exclusion: the parents' experience of caring for children with rare neurodevelopmental disorders. *International Journal of Qualitative Studies on Health and Well-being, 15*(1), Article 1725362. https://doi.org/10.1080/17482631.2020.1725362.

Hart, K., Marchuk, A., Walsh, J. L., & Howlett, A. (2019). Validation of a surgical neonatal nursing workload tool. *Journal of Neonatal Nursing, 25*(6), 293–297. https://doi.org/10.1016/j.jnn.2019.06.002.

Information and Privacy Commissioner of Ontario. IPC. (2015). https://www.ipc.on.ca/.

Kaasalainen, S., Sussman, T., Durepos, P., McCleary, L., Ploeg, J., Thompson, G., & Team, SPA-LTC (2019). What are staff perceptions about their current use of emergency departments for long-term care residents at end of life. *Clinical Nursing Research, 28*(6), 692–707. https://doi.org/10.1177/1054773817749125.

Lambert, C. E., Arbuckle, S. A., & Holden, R. R. (2016). The Marlowe-Crowne Social Desirability Scale outperforms the BIDR Impression Management Scale for identifying fakers. *Journal of Research in Personality, 61,* 80–86. https://doi.org/10.1016/j.jrp.2016.02.004.

McDonald, M., Brown, J., & Knihnitski, C. (2018). Student perception of initial transition into a nursing program: A mixed methods research study. *Nurse Education Today, 64*, 85–92. https://doi.org/10.1016/j.nedt.2018.01.028.

Neyerhof, A. J. (2006). Methods of coping with social desirability bias: A review. *European Journal of Psychology, 15*, 263–280.

Oviedo, G. R., Tamulevicius, N., Onagbiye, S. O., Phidza, M., Sedumedi, C. M., Cameron, M., & Moss, S. J. (2019). Quality of life, physical activity and cardiorespiratory fitness in black African women: B-Healthy project. *Quality of Life Research, 29*, 987–997.

Pauly, B., McCall, J., Browne, A. J., Parker, J., & Mollison, A. (2015). Toward cultural safety: Nurse and patient perceptions of illicit substance use in a hospitalized setting. *Advances in Nursing Science, 38*(2), 121–135. https://doi.org/10.1097/ANS.0000000000000070.

Pesut, B., Thorne, S., Schiller, C., Greig, M., Roussel, J., & Tishelman, C. (2020). Constructing good nursing practice for medical assistance in dying in Canada: An interpretive descriptive study. *Global Qualitative Nursing Research, 7*. https://doi.org/10.1177/2333393620938686.

Scruby, L. S., Rona, H. A., Leipert, B. D., Mair, H. L., & Snow, W. M. (2019). Exploring social support, sport participation and rural women's health using Photovoice: The Manitoba Experience. *Canadian Journal of Nursing Research, 51*(4), 233–244. https://doi.org/10.1177/0844562119832395.

Verkuyl, M., & Hughes, M. (2019). Virtual gaming simulation in nursing education: A mixed-methods study. *Clinical Simulation in Nursing, 29*, 9–14. https://doi.org/10.1016/j.ecns.2019.02.001.

Vincelette, C., Quiroz-Martinez, H., Fortin, O., & Lavoie, S. (2019). Preliminary development

and validation of the Nurse Cardiopulmonary Resuscitation Survey (NCRS) among intensive care unit nurses. *Official Journal of the Canadian Association of Critical Care Nurses, 30*, 24–31.

Wang, C. (1999). Photovoice: A participatory action research strategy applied to women's health. *Journal of Women's Health, 8*(2), 185–192.

Wang, C., & Burris., M. A (1997). Photovoice: Concept, methodology, and use for participatory needs assessment. *Health Education and Behaviour, 24*(3), 369–387.

Wang, W., Seah, B., Jiang, Y., Lopez, V., Tan, C., Lim, S. T., Ren, H., & Khoo, Y. H. (2018). A randomized controlled trial on a nurse-led smartphone-based self-management programme for people with poorly controlled type 2 diabetes: A study protocol. *Journal of Advanced Nursing, 74*(1), 190–200. https://doi.org/10.1111/jan.13394.

Woodgate, R. L., Tailor, K., Tennent, P., Wener, P., & Altman, G. (2020). The experience of the self in Canadian youth living with anxiety: A qualitative study. *PLoS ONE, 15*(1), Article e0228193. https://doi.org/10.1371/journal.pone.0228193.

CHAPTER 15

Rigour in Research

Mina D. Singh | Lorraine Thirsk

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Discuss the purposes of reliability and validity.
- Define reliability.
- Discuss the concepts of stability, equivalence, and homogeneity as they relate to reliability.
- Compare the estimates of reliability.
- Define validity.
- Compare content validity, criterion-related validity, and construct validity.
- Discuss how measurement error can affect the outcomes of a research study.
- Identify the criteria for critiquing the reliability and validity of measurement tools.
- Use the critiquing criteria to evaluate the reliability and validity of measurement tools.
- Understand how to evaluate the quality of qualitative research
- Discuss the purpose of credibility, auditability, and fittingness.
- Apply the critiquing criteria to evaluate the rigour in a qualitative report.
- Discuss how evidence related to research rigour contributes to clinical decision-making.

KEY TERMS

alpha coefficient
alternate-form reliability
auditability
chance error
Cohen's kappa
concurrent validity
constant error
construct validity
content validity
contrasted-groups approach
convergent validity
credibility
criterion-related validity
Cronbach's alpha
divergent validity
equivalence
error variance
face validity
factor analysis
fittingness
homogeneity
hypothesis-testing approach
internal consistency
interrater reliability
item-to-total correlation
known-groups approach
Kuder-Richardson (KR-20) coefficient
methodological coherence
multitrait-multimethod approach
observed test score
parallel-form reliability
predictive validity
random error
reflexivity
reliability
reliability coefficient
rigour
split-half reliability
stability
systematic error
test-retest reliability
trustworthiness
validation sample
validity

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

IN BOTH QUANTITATIVE AND QUALITATIVE RESEARCH, the purpose is to collect trustworthy data that can be used for analyses to make generalizations about the population and that are transferable to other groups. Because findings need to be generalizable and transferable, measurement of nursing phenomena is a major concern of nursing researchers, and rigour is strived for. Rigour refers to the strictness with which a study is conducted to enhance the quality, believability, or trustworthiness of the study findings. Rigour in quantitative research is determined by measurement instruments that validly and reliably reflect the concepts of the theory being tested, so that conclusions drawn from a study will be valid and will advance the development of nursing theory and evidence-informed practice. Thus, psychometric assessments are designed to obtain evidence of the quality of these instruments—that is, their reliability and validity.

Issues of reliability and validity are of central concern to the researcher, as well as to you as the critiquer of research. From either perspective, the measurement instruments that are used in a research study must be evaluated. Many new constructs are relevant to nursing theory, and a growing number of established measurement instruments are available to researchers. However, researchers often face the challenge of developing new instruments and, as part of that process, establishing the reliability and validity of those tools.

In qualitative research, rigour, often called trustworthiness, is ascertained by credibility, auditability, and fittingness. The growing importance of measurement issues, tool development, and related issues (e.g., reliability and validity, qualitative rigour) is evident in issues of the *Journal of Nursing Measurement*, *Canadian Journal of Nursing Research*, *International Journal of Qualitative Methods*, and other nursing research journals. In this chapter, concepts related to quantitative rigour are discussed first, followed by factors that contribute to the trustworthiness of qualitative research.

When you read quantitative research studies and reports, you must assess the reliability and validity of the instruments used in each study to determine the soundness of the selection of these instruments in relation to the concepts or variables under investigation. The appropriateness of the instruments and the extent to which reliability and validity are demonstrated have a profound influence on the findings and on the internal and external validity of the study. Invalid measures produce invalid estimates of the relationships between variables, thus affecting internal validity. The use of invalid measures also leads to inaccurate generalizations to the populations being studied, thus affecting external validity and the ability to apply or not apply research findings in clinical practice. Thus, the assessment of reliability and validity is an extremely important skill to develop for critiquing nursing research.

Regardless of whether a new or already developed measurement tool is used in a research study, evidence of reliability and validity is crucial. Box 15.1 identifies several Internet resources that you can use to access and evaluate the reliability and validity of the measurement instruments used in research studies.

RELIABILITY

People are considered reliable when their behaviour is consistent and predictable. Likewise, the **reliability** of a research instrument is the extent to which the instrument yields the same results on

BOX 15.1

INTERNET RESOURCES FOR ACCESSING AND EVALUATING THE VALIDITY AND RELIABILITY OF MEASUREMENT INSTRUMENTS*

- **Mental Measurements Yearbook—Test Reviews Online**
  - To subscribe, visit https://marketplace.unl.edu/buros/
  - Contains online abstracts of 3,500 commercially available tests, of which more than 2,800 have been critically appraised by the Buros Institute. These reviews also appear in the print version of the Mental Measurements Yearbook series. Searches can be conducted by author or by using the 18 search categories.
- **WORLDviews on Evidence-Based Nursing**
  - To subscribe, visit https://sigmapubs.onlinelibrary.wiley.com/journal/17416787
  - Provides full-text articles and searches, hypertext navigation, links to Cumulative Index to Nursing and Allied Health Literature (CINAHL) and MEDLINE, and tables and figures.
- **the Sigma Repository (formerly the Virginia Henderson Global Nursing e-Repository)**
  - To subscribe, visit https://www.sigmarepository.org/
  - Includes the following databases: Registry of Nurse Researchers, Registry of Research Projects, and Registry of Research Results. This repository also has printed materials such as: Directory of Unpublished Experimental Mental Measures, Instruments for Clinical Health-Care Research, The Instruments of Psychiatric Research, Measuring Health: A Guide to Rating Scales and Questionnaires, and The Mental Measurements Yearbook.
- **Sigma Theta Tau International Honor Society of Nursing**
  - 550 West North Street Indianapolis, IN 46202; 317-634-8171; https://www.sigmanursing.org/
  - Supports the learning, knowledge, and professional development of nurses. Also promotes resources to bridge research and clinical practice.
- **PROQOLID: Patient-Reported Outcome and Quality of Life Instruments Database**
  - https://www.webcitation.org/getfile?fileid=b3447074665f3736126b320f83ff98304e2f43dcIdentifies and describes quality-of-life instruments.

*See Chapter 5 for detailed information about Internet resources.

repeated measures. Reliability, then, is concerned with consistency, accuracy, precision, stability, equivalence, and homogeneity. Concurrent with questions of validity, or after these questions are answered, the researcher and you, as the critiquer, ask how reliable the instrument is.

A reliable measure can produce the same results if the behaviour is measured again by the same scale. Reliability, then, refers to the proportion of accuracy to inaccuracy in measurement. In other words, if researchers use the same or comparable instruments on more than one occasion to measure behaviours that ordinarily remain relatively constant, the researchers would expect similar results if the tools are reliable.

The three main attributes of a reliable scale are **stability**, homogeneity, and equivalence. The stability of an instrument refers to the instrument's ability to produce the same results with repeated testing. The **homogeneity**, or **internal consistency**, of an instrument means that all of the items in a tool measure the same concept or characteristic. An instrument is said to exhibit equivalence if the tool produces the same results when equivalent or parallel instruments or procedures are used. Each of these attributes and the means to estimate them are discussed here. Before these are discussed, however, an understanding of how to interpret reliability is essential.

Interpretation of the Reliability Coefficient

Because all of the attributes of reliability are concerned with the degree of consistency between scores that are obtained at two or more independent times of testing, these attributes often are expressed in terms of a correlation coefficient. The **reliability coefficient**, or **alpha coefficient**, expresses the relationship between the error variance, true variance, and the observed score, and it ranges from 0 to 1. A correlation of 0 indicates no relationship, and thus the error variance is high. When the error variance in a measurement instrument is low, the reliability coefficient is closer to 1. The closer to 1 the coefficient is, the more reliable the tool is. For example, suppose that a reliability coefficient of a tool is reported to be .89. This number indicates that the error variance is small and the tool has little measurement error.

But if the reliability coefficient of a measure is reported to be .49, the error variance is high, and the tool has a problem with measurement error. For a tool to be considered reliable, a level of .70 or higher should be reported, although the intended purpose of the instrument needs to be considered if lower levels are accepted.

The interpretation of the reliability coefficient depends on the proposed purpose of the measure. Seven major tests of reliability can be used to calculate a reliability coefficient, depending on the nature of the tool: *test-retest reliability, parallel- or alternate-form reliability, item-to-total correlation, split-half reliability, Kuder-Richardson coefficient, Cronbach's alpha, and interrater reliability*. These tests are discussed as they relate to the attributes of stability, homogeneity, and equivalence (Box 15.2). In critiquing research reports, you should be aware that no single best way exists to assess reliability in relation to these attributes and that the researcher's method should be consistent with the aim of the research.

Stability

An instrument is thought to be stable or to exhibit stability when repeated administration of the instrument yields the same results. Researchers are concerned with an instrument's stability because they expect the instrument to measure a concept consistently over a period of time. Measurement over time is important in a longitudinal study because in that type of research, an instrument is used on several occasions. Stability is also a consideration when a researcher is conducting an intervention study that is designed to effect a change in a specific variable. In this case, the instrument is administered once and then again after the alteration or change intervention has been completed. The tests that are used to estimate stability are test-retest reliability and parallel- or alternate-form reliability.

BOX 15.2

MEASURES USED TO TEST RELIABILITY

STABILITY
Test-retest reliability
Parallel- or alternate-form reliability

HOMOGENEITY
Item-to-total correlation
Split-half reliability
Kuder-Richardson (KR-20) coefficient
Cronbach's alpha

EQUIVALENCE
Parallel- or alternate-form reliability
Interrater reliability

Test-Retest Reliability

Test-retest reliability is the stability of the scores of an instrument when it is administered more than once to the same participants under similar conditions. Scores from repeated testing are compared. This comparison is expressed by a correlation coefficient, usually a Pearson *r* (see Chapter 17). The interval between repeated administrations varies and depends on the concept or variable being measured. For example, if the variable that the test measures is related to developmental stages in children, the interval between test administrations should be short. The amount of time over which the variable was measured should also be recorded in the report.

An example of an instrument that was assessed for test-retest reliability is O'Keefe-McCarthy, McGillion, Nelson, and associates' (2014) Prodromal-Symptoms Screening Scale. Test-retest reliability was assessed at approximately a 2-week interval, and a high test-retest reliability coefficient ($r = .81$, $p < .01$) was obtained. The interval was adequate (2 weeks between testing), and coefficients exceeded .80 and were thus very good (Nunnally & Bernstein, 1994).

Parallel- or Alternate-Form Reliability

Parallel-form reliability is applicable and can be tested only if two comparable forms of the same instrument exist. **Parallel-form reliability**, or

alternate-form reliability, is like test-retest reliability in that the same individuals are tested more than once within a specific interval, but in the assessment of parallel-form reliability, a different form of the same test is given to the participants on the second testing. Parallel forms or tests contain the same types of items that are based on the same domain or concept, but the wording of the items is different. The development of parallel forms is desired if the instrument is intended to measure a variable for which a researcher believes that "testwiseness" will be a problem; that is, respondents might recognize the test items and try to answer them in the same way as previously, instead of spontaneously.

Practically speaking, developing alternative forms of an instrument is difficult because of the many issues of reliability and validity. If alternative forms of a test exist, they should be highly correlated if they are to be considered reliable.

Research Hint

When a longitudinal design with multiple data-collection points is being conducted, look for evidence of test-retest reliability or parallel-form reliability.

Homogeneity, or Internal Consistency

Another attribute related to reliability of an instrument is the homogeneity with which the items within the scale reflect or measure the same concept. In other words, the items within the scale are correlated with, or complementary to, each other, and the scale is *unidimensional*. A unidimensional scale measures one concept, such as exercise self-efficacy. A total score is then used in the analysis of data.

Corby et al. (2019) tested the reliability of the scales they used in their study to investigate predictors of prenatal breastfeeding self-efficacy. They found that the Cronbach's alpha for the Breastfeeding Self-Efficacy Scale-short form was .94, the perceived Stress Scale was .895, and the State-Trait Anxiety Inventory was .927. These reliability coefficients provided sufficient evidence of the internal consistency of these instruments for the sample. Another example is provided by Kennedy et al. (2015) where the estimated Cronbach's alpha for the revised 22-item NCSES (Nursing Competence Self-Efficacy Scale) with the study population was high .919 (Kennedy, et al., 2015, p. 554). Homogeneity can be assessed with one of four methods: item-to-total correlation, split-half reliability, Kuder-Richardson coefficient, or Cronbach's alpha.

Research Hint

When the characteristics of a study sample differ significantly from those of the sample in the original study, check to see whether the researcher has re-established the reliability of the instrument with the current sample.

Item-to-Total Correlation

The **item-to-total correlation** is a measure of the relationship between each scale item and the total scale. When item-to-total correlations are calculated, a correlation for each item on the scale is generated (Table 15.1). Items that do not achieve a high correlation may be deleted from the instrument. In a research study, the lowest and highest item-to-total correlations are typically reported; the other correlations are usually not reported unless the study is a methodological investigation. An example of an item-to-total correlation report is illustrated in the study by Sidani et al. (2017), who tested the reliability and validity of

TABLE 15.1

EXAMPLES OF ITEM-TO-TOTAL CORRELATIONS FROM COMPUTER-GENERATED DATA

ITEM	ITEM-TO-TOTAL CORRELATION
1	.5069
2	.4355
3	.4479
4	.4369
5	.4213
6	.4216

the Multi-Dimensional Treatment Satisfaction Measure (MDTSM). In that study, the item-to-total correlations were greater than .30 (p. 9). According to Nunnally and Bernstein (1994), these results are acceptable because the minimal mandatory correlation should be greater than .30.

Split-Half Reliability

Split-half reliability involves dividing a scale into halves and making a comparison. The halves may be, for example, odd-numbered and even-numbered items or a simple division of the first from the second half, or items may be randomly grouped into halves that will be analyzed opposite one another. Split-half reliability provides a measure of consistency in terms of sampling the content. The two halves of the test or the contents in both halves are assumed to be comparable, and a reliability coefficient is calculated. If the scores for the two halves are approximately equal, the test may be considered reliable.

The Spearman-Brown formula is one method of calculating the reliability coefficient. In a study to evaluate the reliability and validity of the Cancer Loneliness Scale (CLS) and the Cancer-related Negative Social Expectations Scale (CRNSES), Kara and Cinar (2020) used the split-half reliability method by means of Item Total Score Correlation, Cronbach's Alpha Coefficient, and Spearman-Brown Coefficient Value. There are seven items in the Cancer Loneliness Scale and five items in the Cancer-related Negative Social Expectations Scale. They found that for the CLS, the Cronbach's alpha coefficente was 0.88 and the Spearman-Brown correlation value was r = 0.81. For the CRNSES, the Cronbach's alpha coefficente was 0.82 and the Spearman-Brown correlation value was r = 0.86. For both scales, the correlation coefficient between the split-half of the scale was above 0.70, and thus the internal consistency is high (Boyle, Saklofske, & Matthews, 2015).

Kuder-Richardson Coefficient

The **Kuder-Richardson (KR-20) coefficient** is the estimate of homogeneity used for instruments that have a dichotomous response format. A *dichotomous response format* is one in which the answer to a question should be either "yes" or "no" or either "true" or "false." The technique yields a correlation that is based on the consistency of responses to all items of a single form of a test that is administered once.

Because the scale was a binary format (true/false), the Kuder-Richardson reliability for the entire scale was calculated at .75, which is acceptable, having exceeded the minimum acceptable score of .70; however, the magnitude of the correlation is not robust.

Cronbach's Alpha

The fourth and most commonly used test of internal consistency is Cronbach's alpha. Cronbach's alpha is a test of internal consistency in which each item in the scale is simultaneously compared with the others, and a total score is then used to analyze the data. Many tools used to measure psychosocial variables and attitudes have a Likert-type scale response format (Figure 15.1), which is very suitable for testing internal consistency. In a Likert-type scale format, the participant responds to a question on a scale of varying degrees of intensity between two extremes. The two extremes are anchored by responses ranging from, for example, "strongly agree" to "strongly disagree" or from "most like me" to "least like me." The points between the two extremes may range from 1 to 5 or 1 to 7. Participants are asked to circle the response that most closely represents what they believe. Examples of reported Cronbach's alpha for various studies are given in Box 15.3.

Figure 15.1 displays examples of items from a tool in which a Likert-type scale format was used to develop a nurses' perception of clinical reasoning instrument (Liou, Liu, Tsai et al., 2015). Boscart et al. (2018) tested the psychometric properties of Team Member Perspectives of Person-Centered (TM-PCC) survey. The testing revealed that there were

Directions: Please read each item and circle the number that best describes your current performance. There is no right or wrong answer.
5 = Strongly agree, 4 = Agree, 3 = Neutral, 2 = Disagree, 1 = Strongly disagree

1. I know how to collect an admitted patient's health information quickly.	5	4	3	2	1
2. I can apply proper assessment skills to collect a patient's current health information.	5	4	3	2	1
3. I can identify abnormalities from the collected patient information.	5	4	3	2	1
4. I can identify a patient's health problems from the abnormal information collected.	5	4	3	2	1
5. I can recognize possible early signs or symptoms when a patient's health deteriorates.	5	4	3	2	1
6. I can explain the mechanism and development associated with the early signs or symptoms when a patient's health deteriorates.	5	4	3	2	1
7. I can accurately prioritize and manage any identifiable patient problems.	5	4	3	2	1
8. I can correctly explain the mechanism behind a patient's problems.	5	4	3	2	1
9. I can set nursing goals properly for the identified patient problems.	5	4	3	2	1
10. I can provide appropriate nursing intervention for the identified patient problems.	5	4	3	2	1
11. I am knowledgeable of each nursing intervention provided.	5	4	3	2	1
12. I can identify and communicate vital information clearly to the doctors based on the patient's current condition.	5	4	3	2	1
13. I can anticipate the prescription ordered by the doctor according to the patient information provided.	5	4	3	2	1
14. I can accurately evaluate and identify whether a patient's condition is improved.	5	4	3	2	1
15. I know the follow-up steps to take if the patient's condition does not improve.	5	4	3	2	1

FIG. 15.1 Example of a Likert-type scale response format. From Liou, S. R., Liu, H. C., Tsai, H. M., et al. (2015). The development and psychometric testing of a theory-based instrument to evaluate nurses' perception of clinical reasoning competence. *Journal of Advanced Nursing, 72*(3), 707–717. Copyright © 2015 John Wiley & Sons Ltd.

BOX 15.3

EXAMPLES OF REPORTED CRONBACH'S ALPHA

"Inter-item correlation coefficients were reviewed for redundancy (r > .85) among items. Item-to-total correlation coefficients > .30 and alpha coefficient ≥ .70 supported the TSC (Therapeutic Self-Care) measure's internal consistency reliability" (Sidani & Doran, 2014, p. 20).
"For the ETBQ-S [Empowering Teaching Behaviours Questionnaire], Cronbach's alpha reliability coefficients for subscales ranged from .74 to .96 with an overall reliability of .89" (Babenko-Mould et al., 2012, p. 7).
"Internal consistency for survey responses in hospital using Cronbach's alpha was 0.95" (McQueen et al., 2013, p. 66).
"The estimated Cronbach's alpha for the revised 22-item NCSES [Nursing Competence Self-Efficacy Scale] with the study population was high (.919)" (Kennedy et al., 2015, p. 554).

three separate domains: supporting social relationships, familiarity with residents' preferences, and meaningful resident-staff relationships, as illustrated in Table 15.2. Cronbach's alpha ranged from .62 to .83, and the overall Cronbach's alpha was .82, thereby providing sufficient evidence of the internal consistency of the instrument.

TABLE 15.2

CRONBACH'S ALPHA SCORES FOR THE FOUR DOMAINS OF THE NURSING COMPETENCE SELF-EFFICACY SCALE

OPTIONS	CRONBACH'S ALPHA
Supporting Social Relationships	.83
Familiarity with Residents' Preferences	.71
Meaningful Resident–Staff Relationships	.62

Adapted from Boscart, V. M., Davey, M., Ploeg, J., Heckman, G., Dupuis, S., Sheiban, L., Luh Kim, J., Brown, P., & Sidani, S. (2018). Psychometric evaluation of the Team Member Perspectives of Person-Centered Care (TM-PCC) survey for long-term care homes. *Healthcare, 6*(2), 59. https://doi.org/10.3390/healthcare6020059.

Research Hint

If a research article provides information about the reliability of a measurement instrument but does not specify the type of reliability, it is probably safe to assume that internal consistency reliability was assessed with Cronbach's alpha.

Equivalence

Equivalence is either the consistency or agreement among observers who use the same measurement tool or the consistency or agreement between alternative forms of a tool. An instrument is thought to demonstrate equivalence when two or more observers have a high percentage of agreement about a certain behaviour or when alternative forms of a test yield a high correlation. Two methods to test equivalence are interrater reliability and alternate- or parallel-form reliability.

Interrater Reliability

Some measurement instruments are not self-administered questionnaires but instead are direct measurements of observed behaviour that must be systematically recorded. Such instruments must be tested for **interrater reliability** (the consistency of observations between two or more observers with the same tool). To accomplish interrater reliability, either two or more individuals should make an observation or one observer should observe the same behaviour on several occasions. The observers should score their observations with regard to the definition and operationalization of the behaviour to be observed.

When the research method of direct observation of a behaviour is required, consistency (or reliability) of the observations among all observers is extremely important. Interrater reliability concerns the reliability (or consistency) of the observer, not the reliability of the instrument. Interrater reliability is expressed either as a percentage of agreement between scorers or as a correlation coefficient of the scores assigned to the observed behaviours.

One method of calculating interrater reliability is Cohen's kappa, a coefficient of agreement between two raters that is considered to be a more precise estimate of interrater reliability. **Cohen's kappa** expresses the level of agreement that is observed beyond the level that would be expected by chance alone. A Cohen's kappa of .80 or better is generally assumed to indicate good interrater reliability. A Cohen's kappa of .68 allows tentative conclusions to be drawn when lower levels of reliability are acceptable (McDowell & Newell, 1996). In their study describing the validation of a surgical neonatal nursing workload tool, Hart et al. (2019) checked for interrater reliability. To determine the validity of the Winnipeg Assessment of Neonatal Nursing Needs Tool-Surgical Complex (WANNNT-SC), the charge nurse on each shift was asked how many nurses, based off professional judgement were asked to meet unit staffing needs. Another senior nurse (rater) was asked to complete the blinded surgical tool within 1 hour of the charge nurse (p. 295). The overall interrater Kappa was .73, with a confidence interval ranging from .60–.87.

Evidence-Informed Practice Tip

Interrater reliability is important for minimizing bias.

Parallel- or Alternate-Form Reliability

Parallel- or alternate-form reliability was described in the discussion of stability (see pp. 331–332). Use of parallel forms is thus a measure of stability and equivalence. The procedures for assessing equivalence through the use of parallel forms are the same.

VALIDITY

Validity refers to whether a measurement instrument accurately measures what it is intended to measure. To be valid, an instrument must first be reliable; without reliability, the instrument cannot have validity. However, reliability, although necessary, is not a sufficient condition for validity. Internal and external validity of a study are discussed in Chapter 10.

For example, a valid instrument that is intended to measure anxiety does so; it does not measure another construct, such as stress. A reliable measure can consistently rank participants on a given construct (e.g., anxiety), but a valid measure correctly measures the construct of interest. A measure can be reliable but not valid. Suppose that a

researcher wanted to measure anxiety in patients by measuring their body temperatures. The researcher could obtain highly accurate, consistent, and precise temperature recordings, but such a measure would not be a valid indicator of anxiety. Thus, the high reliability of an instrument is not necessarily congruent with evidence of validity. A valid instrument, however, is reliable. If an instrument is erratic, inconsistent, and inaccurate, it cannot validly measure the attribute of interest.

The three major kinds of validity—content, criterion-related, and construct validity—vary according to the kind of information provided and the investigator's purpose. In critiquing research articles, you will want to evaluate whether sufficient evidence of validity is present and whether the type of validity is appropriate to the design of the study and instruments used in the study. The sample that provides the initial data for determining the reliability and validity of a measurement tool is termed a **validation sample**.

Evidence-Informed Practice Tip

Selecting measurement instruments that have strong evidence of validity increases the reader's confidence in the study findings—that the researchers actually measured what they intended to measure.

Content Validity

Content validity is the degree to which the content of the measure represents the universe of content—that is, the domain of a given construct. The universe of content provides the framework and basis for formulating the items that will adequately represent the content. When an investigator is developing a tool and issues of content validity arise, the concern is whether the measurement tool and the items it contains are representative of the universe of content that the researcher intends to measure. The researcher begins by defining the concept and identifying the dimensions that are the components of the concept. The items that reflect the concept and its dimensions are formulated (see Practical Application box for an example).

When the researcher has completed this task, the items are submitted to a panel of judges considered to be experts on this concept. Researchers typically request that the judges indicate their level of agreement with the scope of the items and the extent to which the items reflect the concept under consideration. The formula for evaluating agreement among experts on individual items is the number agreeing with an item divided by the number of experts. When the concept being rated is relevance, there are two standard methods for computing content validity. The first method is an item-level content validity index (I-CVI), which is computed as the number of experts giving a rating 3 or 4 to the relevancy of each item, divided by the total number of experts. This is usually the first round of rating content validity. In some validation research, there is a second round of agreement where an overall scale content validity is calculated. Oetker-Black and Davis (2019) in the development of the Mock Code Self-Efficacy scale established content validity by using two experts who actively teach cardiopulmonary resuscitation (CPR). The experts assessed if the items included accurately represented CPR skills. Each item was evaluated for relevance, clarity in representing the skills, and sufficiency. The experts were asked to assess the relevance and content of items. A content validity index was used to rate relevance of each item using a 4-point scale. All items received a rating of 3 or 4 by the experts, indicating that all items had content validity (p. 37).

Practical Application

Vincelette et al. (2019) reported on the preliminary development and validation of the nurse cardiopulmonary resuscitation survey (NCRS) among intensive care unit nurses. They first reviewed the literature and found that there was a gap in this area. Items for the scale were generated by one of the authors, based on the 2015 guidelines for advanced cardiovascular life support (ACLS). Two items were deemed repetitive and removed before submitting the instrument to face validity experts. The assessment of face validity led to the adjustment of the word choice of three items that were perceived as unclear. This step led to a final version that was submitted to an independent expert panel to assess content validity.

A subtype of content validity is **face validity**, which is a rudimentary type of validity in which the instrument intuitively gives the appearance of measuring the concept. To establish face validity, colleagues or participants are asked to read the instrument and evaluate the content in terms of whether it appears to reflect the concept that the researcher intends to measure. This procedure may be useful in the tool development process in terms of determining the readability and clarity of the content. Face validity, however, should in no way be considered a satisfactory alternative to other types of validity. In the development of the Mock Code Self-Efficacy Scale for nursing students, Oetker-Black and Davis (2019) established face validity by using five senior undergraduate nursing students to review the questions and the reading level.

Evidence-Informed Practice Tip

When face validity and content validity, the most basic types of validity, are the only types of validity reported in a research article, you, as a research consumer, cannot appraise the measurement tools as having strong psychometric properties; thus, you would lack confidence in the usefulness of the study findings.

Criterion-Related Validity

Criterion-related validity is the degree of relationship between the participant's performance on the measurement tool and the participant's actual behaviour. The criterion is usually the second measure, which is used to assess the same concept being studied.

Two types of criterion-related validity are concurrent and predictive. Concurrent validity is the degree of correlation of two measures of the same construct administered at the same time. A high correlation coefficient indicates agreement between the two measures. **Predictive validity** is the degree of correlation between the measure of the concept and a future measure of the same concept. Because of the passage of time, the correlation coefficients are likely to be lower for predictive validity studies.

For example, in a study by Ford-Gilboe et al. (2016) describing the development and validation of Composite Abuse Scale (Revised)—Short Form (CASR-SF), they examined concurrent validity through measures of depression, symptoms of PTSD and coercive control. As expected, CASR-SF total scores were moderately correlated with each validation measure, providing support for concurrent validity of the instrument construct validity.

Construct Validity

Construct validity is the extent to which a test measures a theoretical construct or trait. To establish this type of validity, the researcher attempts to validate a body of theory underlying the measurement and testing of the hypothesized relationships. Empirical testing confirms or fails to confirm the relationships that would be predicted among concepts and, as such, provides more or less support for the construct validity of the instruments measuring those concepts. Establishing construct validity is a complex process, often involving several studies and approaches. The following approaches are discussed in this section: hypothesis-testing, convergent and divergent, contrasted-groups, and factor-analytical.

In their study, Sidani et al. (2017) assessed construct validity of the MDTSM discussed in Item-to-Total Correlation section earlier, by factor analysis and relationships between the variables. They found that there was a positive relationship between the MDTSM subscales and self-reported adherence to therapy, which means that high levels of satisfaction were associated with high levels of adherence (p. 9).

Hypothesis-Testing Approach

When the **hypothesis-testing approach** is used, the investigator uses the theory or concept underlying the measurement instrument to validate the instrument. The investigator accomplishes this task first by developing hypotheses about the behaviour of individuals with varying scores on the measure; then by gathering data to test the hypotheses; and, finally, on the basis of the findings, by making inferences about whether the rationale underlying the instrument's construction is adequate to explain the findings. Hypothesis-testing approaches

include convergent validity, divergent validity, and known-groups validity.

For example, Lambert et al. (2015) used a hypothesis-testing approach to establish the Appraisal of Caregiving Scale. This scale was designed to evaluate stress associated with caregiving for someone with advanced cancer. Construct validity was tested positively on the basis of two hypotheses: (1) that depressed caregivers had higher stress scores than non-depressed caregivers and (2) younger caregivers reported significantly higher scores on the General Stress subscale than older caregivers.

CONVERGENT AND DIVERGENT APPROACHES. Two strategies for assessing construct validity are convergent and divergent approaches.

Convergent validity exists when two or more tools that are intended to measure the same construct are administered to participants and are found to be positively correlated. A correlational analysis (i.e., a test of relationship; see Chapters 12 and 17) determines whether the measures are positively correlated, in which case convergent validity is said to be supported.

In contrast to convergent validity, the calculation of **divergent validity** requires measurement approaches that differentiate one construct from others that may be similar. Sometimes researchers search for instruments that measure the opposite of the construct. If the divergent measure is negatively related to other measures, the measure's validity is strengthened.

As an example, Melnyk, Oswalt, and Sidora-Arcoleo (2014) assessed the psychometric properties of scores on the Neonatal Intensive Care Unit Parental Beliefs Scale (NICU PBS) in a sample of mothers and fathers of preterm infants receiving intensive care. The NICU PBS is a rating instrument designed to assess parental beliefs about their premature infant and their role during hospitalization. For convergent and divergent (discriminant) validity assessment, correlation analysis of the Time 1 data was used for assessment of the NICU PBS with maternal demographic characteristics (age, education, employment status), mental health, stress, pregnancy history variables (gravidity, high-risk status, subsequent pregnancy in 12 months), and baby outcome variables (Clinical Risk Index for Babies [CRIB] scores, birth weight, NICU length of stay). Higher total PBS scores were associated with younger maternal age, lower education, lower income, receipt of Medicaid, minority status, mothers' employment, no biological father in the study, higher gravidity, having had another child in the past 12 months, shorter NICU length of stay, and lower stress, anxiety, and depression.

A specific method of assessing convergent and divergent validity is the **multitrait-multimethod approach**. Similar to the divergent validity approach just described, this method, proposed by Campbell and Fiske (1959), also involves examining the relationships between instruments that are intended to measure the same construct and between those that are intended to measure different constructs. A variety of measurement strategies, however, are used. In other words, this approach is a type of validation in which more than one method is used to assess the accuracy of an instrument. For example, anxiety could be measured by the following:

- Administering the State-Trait Anxiety Inventory
- Recording blood pressure readings
- Asking the participant about anxious feelings
- Observing the participant's behaviour

The results of one of these measures should then be correlated with the results of each of the others in a multitrait-multimethod matrix (Waltz, Strickland, & Lenz, 1991).

The use of multiple measures of a concept decreases systematic error. The use of a variety of data-collection methods (e.g., self-report, observation, interview, and collection of physiological data) also diminishes the effect of systematic error.

Contrasted-Groups Approach

In the **contrasted-groups approach** (sometimes called the **known-groups approach**) to the development of construct validity, the researcher

identifies two groups of individuals expected to score extremely high or extremely low in the characteristic being measured by the instrument. The instrument is administered to both groups, and the differences in scores are examined. If the instrument is sensitive to individual differences in the trait being measured, the mean performance of these two groups should differ significantly, and evidence of construct validity would be supported. A t test or analysis of variance is used to statistically measure the difference between the two groups.

Factor-Analytical Approach

A final approach to assessing construct validity is **factor analysis**. This procedure gives the researcher information about the extent to which a set of items measures the same underlying construct or the same dimension of a construct. In factor analysis, the researcher assesses the degree to which the individual items on a scale truly cluster around one or more dimensions. Items designed to measure the same dimension should load on the same factor; those designed to measure differing dimensions should load on different factors (Nunnally & Bernstein, 1994).

A factor analysis also indicates whether the items in the instrument reflect a single construct or several constructs. Several factors may be identified in a set of data. The study must have a large sample size in order to conduct a factor analysis. Nunnally and Bernstein (1994) recommended 10 observations for each variable. Thus, to develop the factor structure and reliability of the Nursing Competence Self-Efficacy Scale, Kennedy and associates (2015) used 252 students to test 22 items.

Research Hint

When validity data about a study's measurement instruments are not included in a research article, you cannot determine whether the intended concept is being captured by the measurement tool. Before you use the results, check the instrument's validity by reviewing the original source.

Evidence-Informed Practice Tip

When the tools used in a study are presented, note whether the sample used to develop the measurement instruments is similar to your patient population.

The Critical Thinking Decision Path will help you assess the appropriateness of the type of validity and reliability selected for use in a particular research study.

Researchers may be concerned about whether the scores that were obtained for a sample of participants were consistent, true measures of the behaviours, and thus an accurate reflection of the differences between individuals. The extent of variability in test scores that is attributable to error rather than a true measure of the behaviours is the error variance.

An **observed test score** that is derived from a set of items consists of the true score plus error (Figure 15.2). The error may be either chance (random) error or systematic error.

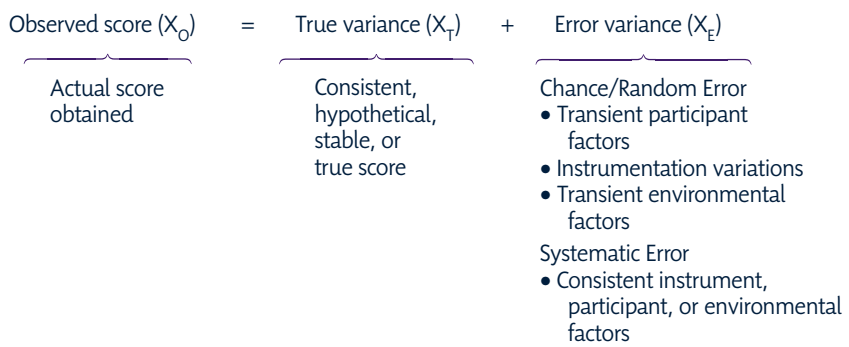


FIG. 15.2 Components of observed scores.

A chance error or a random error is an error that is difficult to control (e.g., a respondent's anxiety at the time of testing). These errors are unsystematic and are not predictable; thus, they cannot be corrected. However, awareness of the sources of these errors may help the researcher minimize their effect on measurement accuracy. These sources are as follows:

1. Transient human conditions, such as hunger, fatigue, health, lack of motivation, and anxiety, which are often beyond the awareness and control of the examiner.
2. Variations in the measurement procedure, such as misplacement of the blood pressure cuff, not waiting for a specific time period before taking the blood pressure, or placing the arm randomly in relation to the heart while measuring blood pressure; changing the wording of interview questions between administrations; or environmental factors, such as the presence of others while data are being obtained, a cold room, or discomfort with the researcher (who is part of the environment).
3. Errors in data processing, such as coding errors and incorrect inputting into the computer.

Chance errors affect an individual's observed score, so that the person's observed score may be higher than his or her true score, whereas another person's observed score may be lower than his or her true score. Instruments that are free of chance errors are considered reliable. A **systematic error** or a **constant error** is a measurement error that is attributable to relatively stable characteristics of the study population that may bias their behaviour, cause incorrect instrument calibration, or both. Such error has a systematic biasing influence on the participants' responses and thereby influences the validity of the instruments. Level of education, socioeconomic status, social desirability, response pattern, or other characteristics may influence the validity of the instrument by altering the measurement of the "true" responses in a systematic way. For example, a participant who wants to please the investigator may constantly answer items in a socially desirable way, thus making the estimate of validity inaccurate.

Systematic error also occurs when an instrument is improperly calibrated. Consider a scale that consistently gives a person's weight at 1 kg less than the actual body weight. The scale could be quite reliable (i.e., capable of reproducing the precise measurement), but the result is consistently invalid. Systematic error is considered part of the true score. The multimethod-multitrait approach is one method of decreasing systematic error. The validity of an instrument is the extent to which it is free of both chance errors and systematic errors.

The amount of detail about reliability and validity varies considerably among research articles. When the focus of a study is tool development, psychometric evaluation—including extensive reliability and validity data—is carefully documented and appears throughout the article rather than briefly in the "Instruments" section, as in other research studies.

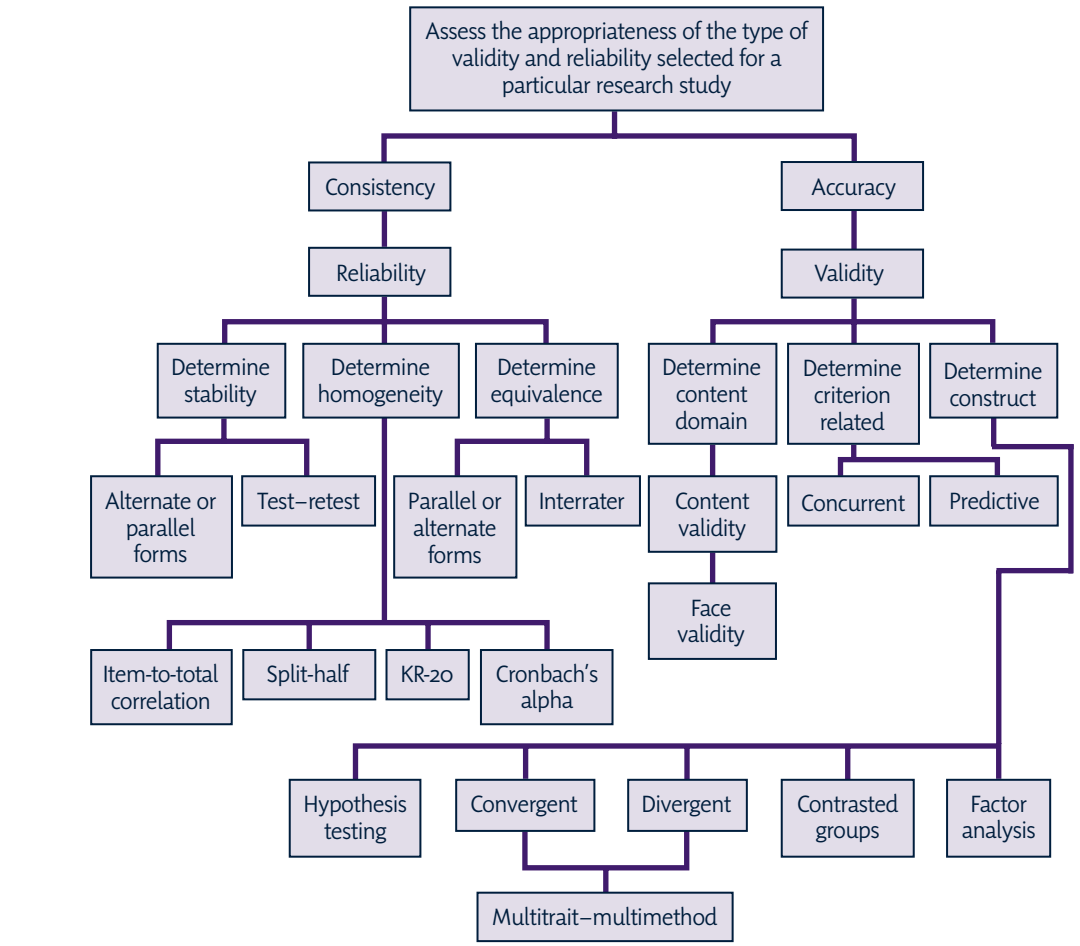
RIGOUR IN QUALITATIVE RESEARCH

It is important to keep in mind that qualitative research is an umbrella term that covers a variety of research methods. While there are general principles when evaluating rigour or trustworthiness in qualitative studies, there are also differences across methods that will influence criteria for evaluating. Although there is still debate amongst researchers about how to best evaluate qualitative research, there is consensus that given its increasing prevalence and use in evidence-informed practice, there should be some evaluation of quality (Williams et al., 2020).

As in quantitative research, the basic approach to ensure rigour in qualitative research is employing methodical research design, data collection, interpretation, and communication. Qualitative researchers seek to achieve two goals: (1) to account for the method and the data, which must be independent so that another researcher can analyze the same data in the same way and make the same conclusions; and (2) to produce a credible

CRITICAL THINKING DECISION PATH

Determining the Appropriate Type of Validity and Reliability Selected for a Study



KR-20, Kuder-Richardson coefficient.

and reasoned explanation of the phenomenon under study. Thus, the rigour in qualitative methodology is judged by unique criteria appropriate for the research approach and is often called trustworthiness. Credibility, auditability, and fittingness are some of the scientific criteria for *trustworthiness* proposed for qualitative research studies by Lincoln and Guba (1985), along with authenticity (Guba & Lincoln, 1994). The meaning of credibility, auditability, and fittingness are briefly explained in Table 15.3.

Morse et al. (2002) offered a different viewpoint and terminology to describe rigour in qualitative research. They argued that ensuring rigour is part of the process of the research, rather than a post hoc evaluation of the presented research study, and that reliability and validity are useful terms in qualitative research. In qualitative research, reliability means showing consistent support for the findings across participants; validity means the data are appropriate and provide an accurate account of participants (Spiers et al., 2018). Verification refers to the

TABLE 15.3

CRITERIA FOR JUDGING SCIENTIFIC RIGOUR: CREDIBILITY, AUDITABILITY, FITTINGNESS

CRITERIA	CHARACTERISTICS
Credibility	Truth of findings as judged by participants and others within the discipline. For example, you may find the researcher returning to the participants to share interpretation of findings and query accuracy from the perspective of the persons living the experience.
Auditability	Accountability as judged by the adequacy of information leading the reader from the research question and raw data through various steps of analysis to the interpretation of findings. For example, you should be able to follow the reasoning of the researcher step by step through explicit examples of data, interpretations, and syntheses.
Fittingness	Faithfulness to the everyday reality of the participants, described in enough detail so that others in the discipline can evaluate importance for their own practice, research, and theory development. For example, you will know enough about the human experience being reported that you can decide whether it "rings true" and is useful for guiding your practice.

process of evaluating the multiple decisions made during the research, which contribute to the validity and reliability of the qualitative study (Morse et al., 2002). Verification strategies proposed by Morse et al. (2002) are found in Table 15.4. Other terms you may encounter are transparency, reflexivity, dependability, and transferability.

Research Hint

Different qualitative methods will use different processes to ensure rigour. Understanding the different qualitative methods, as presented in Chapter 8, is an important part of assessing a qualitative research report. Evaluating adherence to the stated method is an important aspect of rigour in qualitative research; this is called *methodological coherence*. There should be enough description of the research process in the "Methods" section that you can evaluate this coherence.

TABLE 15.4

VERIFICATION STRATEGIES IN QUALITATIVE RESEARCH

STRATEGY	DESCRIPTION
Investigator responsiveness	-ongoing analysis forces purposive sampling-ensure analysis holds together; be open to new ideas, giving up those that are not supported by the data
Methodological coherence	-coherence between the research question and the methods-question should match the method, which matches the data collection and analysis procedures
Appropriate sampling	-participants should be appropriate; those who have knowledge of the research topic-adequate number to account for different aspects of phenomenon
Concurrent data collection and analysis	-analysis begins as data collection begins-ensures mutual interaction between "what is known and what one needs to know" (p. 18)
Thinking theoretically	-building a solid foundation, in incremental steps-emerging ideas are reconfirmed in new data
Theory development	-move between a micro perspective of data to a macro conceptual/theoretical understanding

Adapted from: Morse et al. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, *1*(2), 13–22. https://sites.ualberta.ca/~iiqm/backissues/1_2Final/pdf/morseetal.pdf.

Credibility

Credibility is a characteristic of qualitative research that refers to the accuracy, validity, and soundness of data. It is similar to internal validity in qualitative research. The methods to ensure credibility are prolonged engagement, persistent observation, peer debriefing, and member checks (Lincoln, 1995). In prolonged engagement and

persistent observation, the researchers spend sufficient time with the study's participants to check for discrepancies in responses. Peer debriefing is conducted with experts in the field, whose probing questions and review about the research can assist the researchers in improving trustworthiness in the data. Member checking verifies the accuracy of participants' responses by asking the study participants to review the themes and narratives to determine whether the researchers accurately described their experiences (Lincoln & Guba, 1985). Member checking is not an appropriate strategy for all qualitative methods (Morse et al., 2002).

Triangulation, crystallization, and searching for disconfirming evidence through negative case analyses are also used to ensure credibility and confirmability. In Chapter 7, triangulation—the cross-checking and verification of data through the use of different information sources, such as a variety of data sources, investigators, theoretical models, and research methods—and crystallization in both qualitative and mixed method research are discussed. Triangulation is viewed as offering completeness to naturalistic inquiry (Tobin & Begley, 2004), but may not be appropriate for all methods.

Auditability and Fittingness

Engaging in an inquiry audit establishes both the auditability and the fittingness of the data. **Auditability** is the characteristic of a qualitative study, developed by the investigator's research process, that allows another researcher or a reader to follow the thinking or conclusions of the investigator. Auditability means the researcher presents a decision trail to the reader, and this allows the reader of the research to better judge its quality or trustworthiness (Sandelowski, 1986). Fittingness is the degree to which study findings are applicable or transferable outside the study situation and the degree to which the results are meaningful to individuals not involved in the research. The audit trail was proposed by Guba (1981) to allow external auditors to follow the trail of qualitative data gathering and has been described by Lincoln and Guba (1985) as "the most important trustworthiness technique available" (p. 283). The audit trail involves reviewing all documents relating to the study, such as the research protocol, memos and correspondences, research tools, and field notes.

An example of how transferability was enhanced in a grounded theory study is provided by King-Shier et al. (2019). They were able to develop rich descriptions of the management of hypertension within a group of South Asians. Careful sampling to ensure representation, rigourous data analysis procedures, and data saturation contributed to the ability of others to see utility of the results in other contexts (p. 323).

Authenticity

Authenticity refers to fairness in the presentation in that all value conflicts, differences, and views of the participants are noted in the analysis. The reader can understand the moods and experiences of the participants while reading the thematic analyses (Guba & Lincoln, 1994). A variety of viewpoints should be offered, and not only those that align with the researchers' previously held views and opinions (Tobin & Begley, 2004).

Reflexivity. Reflexivity refers to the researcher's process of reflection and critical thinking during the research process. In particular, the researcher should reflect on whether or how they have influenced sampling, data collection, analysis, and presentation of the findings (Williams et al., 2020). This is often supported by a skeptical peer review during the research process or during the publication process (Buetow, 2019). Another technique to promote reflexivity is the keeping of field notes or reflective journals during the research process. Ethnographers, for example, will routinely document changes in their ideas, beliefs, and values as they are engaged in the research.

Evidence-Informed Practice Tip

The narrative or visual formats used by qualitative researchers can be an artistic as well as a scholarly

endeavour (Sandelowski, 2015). As a reader, it is important to be critical about the presentation of the research, and how that might affect what you take away from it. Can good writing cover up bad research? Or can good research be overlooked because of poor writing?

Tools for Appraising Qualitative Research. Critical appraisal checklists and frameworks to assess qualitative research can be found from numerous sources. It is important to keep in mind that there are two types of tools: those intended to provide guidance in appraisal of quality, and those that provide guidance and criteria for researchers and reviewers in terms of how to report a qualitative research project (Williams et al., 2020). Meyrick (2006) offered a useful model for evaluating qualitative research based on guiding principles of transparency and conducting research systematically (Figure 15.3). This model captures many of the suggestions described previously regarding rigour in qualitative research.

Practical Application

Benbow et al. (2019) incorporated several strategies to ascertain rigour in their study to better understand the unique narratives of social exclusion for mothers experiencing homelessness. In terms of credibility, the researcher ensured prolonged engagement in community agencies, which involved opportunities for follow-up interviews with participants and e-mail/telephone/face-to-face follow-up with service providers and community agencies to discuss study results. The follow-up interviews were not for member checking, but to share and co-construct current understandings and provide an opportunity for participants to review their transcripts. An audit trail was used to make clear all decision-making throughout the analysis process. Each decision was tracked and supported with raw data, while being grounded in the study purpose and the theoretical lens of the study. Fittingness was achieved during follow-up interviews where the researcher gained insight and special attention was paid to convergent and divergent perspectives within and across participants. Because the criteria of credibility, fittingness/transferability, and auditability were met, trustworthiness was achieved.

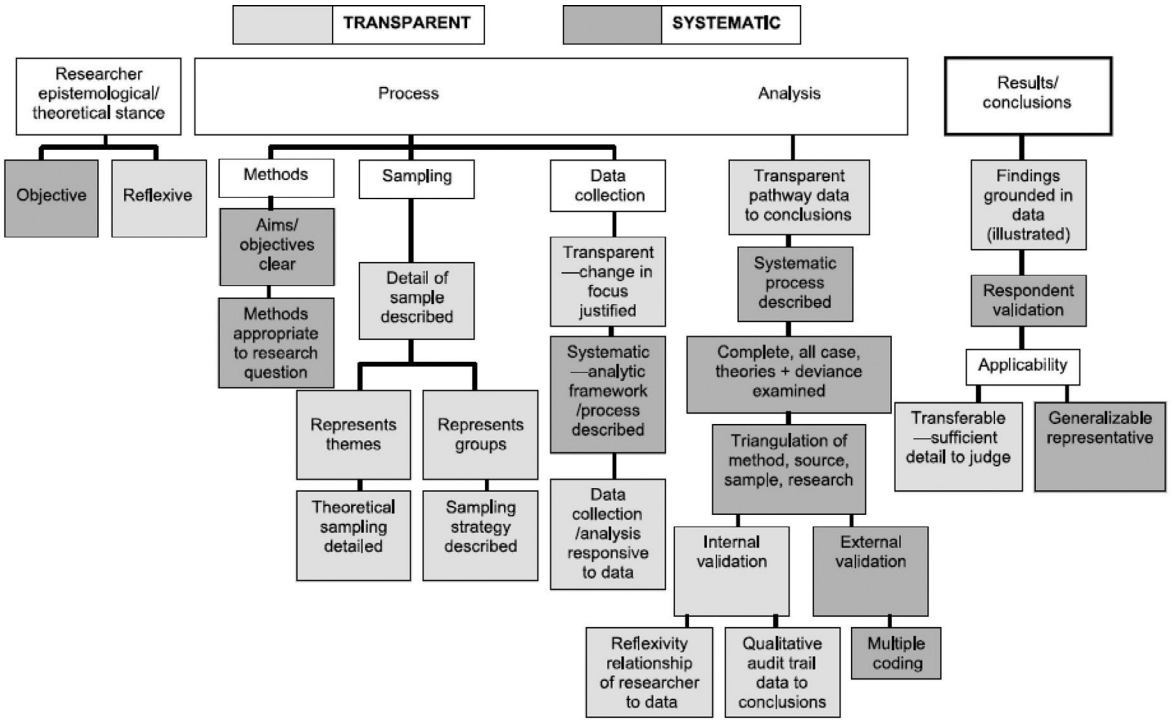


FIG. 15.3 Quality framework for qualitative research. From Meyrick, J. (2006). What is good qualitative research: A first step towards a comprehensive approach to judging rigour/quality. *Journal of Health Psychology, 11*(5), 799–808. https://doi.10.1177/1359105306066643.

APPRAISING THE EVIDENCE

Reliability and Validity

Reliability and validity are two crucial aspects in the critical appraisal of a measurement instrument. The reviewer evaluates an instrument's level of reliability and validity, as well as how they were established. In a research report, the reliability and validity for each measure should be presented. If these data have not been presented, the reviewer must seriously question the merit and use of the tool and the study's results. Criteria for critiquing reliability and validity are presented in the Critiquing Criteria box.

If reliable and valid questionnaires are not used in a study, the results cannot be credible. As a critiquer, you have an ethical responsibility to question the reliability and validity of instruments used in research studies and to examine the findings in view of the quality of the instruments used and the data presented. The following discussion highlights key areas related to reliability and validity that should be evident in a research article.

Appropriate reliability tests should have been performed by the developer of the measurement tool and should then have been included by the current user in the research report. If the initial standardization sample and the current sample have different characteristics, the reader would find either (1) that a pilot study for the present sample would have been conducted to determine whether the reliability was maintained or (2) that a reliability estimate was calculated for the current sample. For example, if the standardization sample for a tool that measures "satisfaction in an intimate heterosexual relationship" comprises undergraduate college students and if an investigator plans to use the tool with married couples, the reliability of the tool should be established with the latter group.

The investigator determines which type of reliability procedure is used in the study, depending on the nature of the measurement tool and how it will be used. For example, if the instrument is to be administered twice, you might determine that test-retest reliability should have been used to establish the stability of the tool. If an alternate form of the instrument has been developed for use in a repeated-measures design, evidence of alternate-form reliability should be presented to determine the equivalence of the parallel forms.

If the degree of internal consistency among the items is relevant, an appropriate test of internal consistency should be presented. In some instances, more than one type of reliability is presented, but you should determine whether all are appropriate. For example, the Kuder-Richardson formula implies that a single right or wrong answer exists, which makes use of the coefficient inappropriate with scales that provide a format of three or more possible responses. In such cases, another formula is applied, such as Cronbach's alpha.

Another important consideration is the acceptable level of reliability, which varies according to the type of test. Coefficients with reliability of .70 or higher are desirable. The validity of an instrument is limited by its reliability; in other words, less confidence can be placed in scores from tests with low-reliability coefficients.

Satisfactory evidence of validity is probably the most difficult determination for you as reviewer. This aspect of measurement is most likely to fall short of meeting the required criteria. Validity studies are time consuming and complex, and researchers sometimes settle for presenting minimal validity data.

Therefore, you should closely examine the item content of a tool when you evaluate its strengths and weaknesses and try to find conclusive evidence of content validity. In the body of a research article, however, it is unusual to have more than a few sample items available for review. Thus, you should determine whether the appropriate assessment of content validity was used to meet the researcher's goal.

Such procedures provide assurance that the tool is psychometrically sound and that the content of the items is consistent with the conceptual framework and the construct definitions. Construct validity and criterion-related validity are two of the more precise statistical tests of whether the tool measures what it is intended to measure. Ideally, an instrument should provide evidence of content validity, as well as criterion-related or construct validity, before a reviewer invests a high level of confidence in the tool.

You should also expect to see the strengths and weaknesses of instrument reliability and validity presented in the "Discussion," "Limitations," or

Continued

APPRAISING THE EVIDENCE—*cont'd*

Reliability and Validity

"Recommendations" section, or in all of these sections, of a research article. In this context, the reliability and validity might be discussed in relation to other tools devised to measure the same variable. The relationship of the study's findings to the strengths and weaknesses in instrument reliability and validity is another important discussion point.

Finally, the researcher should propose recommendations for improving future studies in relation to instrument reliability and validity. For example, in the "Implications for Future Research and Practice" section of a report about developing and validating the Nursing Competence Self-Efficacy Scale, Kennedy and associates (2015) noted that they will be conducting a replication of this initial psychometric assessment with a larger sample to support the factor structure in the study.

Collegial dialogue is also an approach to evaluating the merits and shortcomings of an existing instrument, as well as a newly developed one, that is reported in the nursing literature. Such an exchange promotes the understanding of methodologies and techniques of reliability and validity, stimulates the acquisition of a basic knowledge of psychometrics, and encourages the exploration of alternative methods of observation and the use of reliable and valid tools in clinical practice.

CRITIQUING CRITERIA

QUANTITATIVE STUDIES

1. Was an appropriate method used to test the reliability of the tool?
2. Is the reliability of the tool adequate?
3. Was an appropriate method used to test the validity of the instrument?
4. Is the validity of the measurement tool adequate?
5. If the sample from the developmental stage of the tool was different from the current sample, were the reliability and validity recalculated to determine whether the tool is still adequate?
6. Have the strengths and weaknesses of the reliability and validity of each instrument been presented?
7. Are the strengths and weaknesses of the research appropriately addressed in the "Discussion," "Limitations," or "Recommendations" sections of the report?

QUALITATIVE STUDIES

1. Do the authors explain the process of the research, including the specific qualitative method used?
2. Is evidence provided that the researcher's interpretation is supported by the data?
3. Can the reader follow the researcher's thinking?
4. Does the study follow the principles and processes of the stated research method?
5. Can the findings be applicable to outside the study situation?
6. Are the results meaningful to individuals not involved in the research?
7. Do the conclusions, implications, and recommendations give the reader a context in which to use the findings?
8. Do the conclusions reflect the study's findings?
9. Is there evidence of researcher reflexivity?

CRITICAL THINKING CHALLENGES

- Discuss the three types of validity that must be established before a reviewer invests a high level of confidence in the tool. Include examples of each type of validity.
- What are the major tests of reliability? Is it necessary to establish more than one measure of reliability for each instrument used in a study? Which do you think is the most essential measure of reliability? Include examples in your answer.
- Is it possible to have a valid instrument that is not reliable? Is the reverse possible? Support your answer with instruments you might use in the clinical setting with your patients.
- What are some ways in which credibility, auditability, and fittingness can be evaluated?
- How do you think the concept of evidence-informed practice has changed research utilization

models? Is the review of the literature the same when a research proposal is developed as it is when the steps of research utilization or an evidence-informed practice protocol is implemented? Support your position.

CRITICAL JUDGEMENT QUESTIONS

1. In testing a new scale to measure nursing students' clinical competency in simulation, the researchers administered repeatedly and obtained the same results, this is called?
   a. Validity
   b. Reliability
   c. Consistency
   d. Predictability
2. A newly developed instrument is found to have a Cronbach's alpha of .82. What is the correct interpretation of this finding?
   a. The instrument has no internal consistency
   b. The instrument has a low degree of internal consistency
   c. The instrument has a moderate degree of internal consistency
   d. The instrument has a relatively high degree of internal consistency
3. What is the formal term for rigour in qualitative research?
   a. Identifying metaphors from data
   b. Describing the insiders' view of data
   c. Bracketing
   d. Trustworthiness

KEY POINTS

- Reliability and validity are crucial aspects of conducting and critiquing research.
- Validity refers to whether an instrument measures what it is purported to measure. It is a crucial aspect of evaluating a tool.
- Three types of validity are content validity, criterion-related validity, and construct validity.
- The choice of a validation method is important and is made by the researcher on the basis of the characteristics of the measurement device in question and its use.
- Reliability refers to the ratio between accuracy and inaccuracy in a measurement device.
- The major tests of reliability are test-retest reliability, parallel- or alternate-form reliability, split-half reliability, item-to-total correlation, the Kuder-Richardson coefficient, Cronbach's alpha, and interrater reliability.
- The selection of a method for establishing reliability depends on the characteristics of the tool, the testing method that is used for collecting data from the standardization sample, and the kinds of data that are obtained.
- Credibility, auditability, and fittingness are criteria for judging the scientific rigour of a qualitative research study.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Benbow, S., Forchuk, C., Berman, H., Gorlick, C., & Ward-Griffin, C. (2019). Spaces of exclusion: Safety, stigma, and surveillance of mothers experiencing homelessness. *Canadian Journal of Nursing Research, 51*(3), 202–213. https://doi.org/10.1177/0844562119859138.

Boscart, V. M., Davey, M., Ploeg, J., Heckman, G., Dupuis, S., Sheiban, L., Luh Kim, J., Brown, P., & Sidani, S. (2018). Psychometric evaluation of the team member perspectives of person-centered care (TM-PCC) survey for long-term care homes. *Healthcare (Basel, Switzerland), 6*(2), 59. https://doi.org/10.3390/healthcare6020059.

Boyle, G. J., Saklofske, D. H., & Matthews, G. (2015). Criteria for selection and evaluation of scales and measures. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 3–15). Philadelphia: Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-386915-9.00001-2.

Buetow, S. (2019). Apophenia, unconscious bias and reflexivity in nursing qualitative research. *International Journal of Nursing Studies, 89*, 8–13. https://doi.org/10.1016/j.ijnurstu.2018.09.013.

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the matrix. *Psychological Bulletin, 53,* 273–302.

Corby, K., Kane, D., & Dayus, D. (2019). Investigating predictors of prenatal breastfeeding self-efficacy. *Canadian Journal of Nursing Research, 53*(1), 56–63. https://doi.org/10.1177/0844562119888363.

Ford-Gilboe, M., Wathen, C. N., Varcoe, C., MacMillan, H. L, Scott-Storey, K., Mantler, T., Hegarty, K., & Perrin, N. (2016). Development of a brief measure of intimate partner violence experiences: the Composite Abuse Scale (Revised)-Short Form (CASR-SF). *BMJ Open, 6,* 12. https://doi.org/10.1136/bmjopen-2016-012824.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic enquiries. *Educational Communication and Technology Journal, 29*, 75–91.

Guba, E., & Lincoln, Y (1994). Competing paradigms in qualitative research. In N. Denzin, & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.

Hart, K., Marchuk, A., Walsh, J.-L., & Howlett, A. (2019). Validation of a surgical neonatal nursing workload tool. *Journal of Neonatal Nursing, 25*(6), 293–297. https://doi.org/10.1016/j.jnn.2019.06.002.

Kara, E., & Cinar, I.O. (2020). Validity and reliability of The Cancer Loneliness and The Cancer-Related Negative Social Expectations Scale. *International Journal of Assessment Tools in Education, 7* (3), 392–403. https://doi.org/10.21449/ijate.711073.

Kennedy, E., Tomblin-Murphy, G., Martin-Misener, R., et al. (2015). Development and psychometric assessment of the Nursing Competence Self-Efficacy Scale. *Journal of Nursing Education, 54*(10), 550–558.

King-Shier, K. M., Dhaliwal, K. K., Puri, R., LeBlanc, P., & Johal, J. (2019). South Asians' experience of managing hypertension: A grounded theory study. *Patient Preference and Adherence, 13*, 321–329. https://doi.org/10.2147/PPA.S196224.

Lambert, S.D., Yoon, H., Ellis, K.R., & Northouse, L. (2015). Measuring appraisal during advanced cancer: psychometrica testing of the appraisal of caregiving scale. *Patient Education and Counselling, 98*(5), 633–639.

Lincoln, Y. S. (1995). Emerging criteria for qualitative and interpretive research. *Qualitative Inquiry, 3*, 275–289.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. New York: Sage.

Liou, S. R., Liu, H. C., Tsai, H. M., et al. (2015). The development and psychometric testing of a theory-based instrument to evaluate nurses' perception of clinical reasoning competence. *Journal of Advanced Nursing, 72*(3), 707–717.

McDowell, I., & Newell, C. (1996). *Measuring health: A guide to rating scales and questionnaires*. New York: Oxford University Press.

Melnyk, B. M., Oswalt, K. L., & Sidora-Arcoleo, K. (2014). Validation and psychometric properties of the Neonatal Intensive Care Unit Parental Beliefs Scale. *Nursing Research, 63*(2), 279–289.

Meyrick, J. (2006). What is good qualitative research? A first step towards a comprehensive approach to judging rigour/quality. *Journal of Health Psychology, 11*(5), 799–808. https://doi.org/10.1177/1359105306066643.

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1*(2), 13–22. https://doi.org/10.1177/160940690200100202.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Oetker-Black, S. L., & Davis, T. (2019). Psychometric evaluation of the Mock Code Self-Efficacy Scale. *Nursing Education Perspectives, 40*(1), 35–40. https://doi.org/10.1097/01.NEP.0000000000000341.

O'Keefe-McCarthy, S., McGillion, M., Nelson, S., et al. (2014). Content validity of the Toronto Pain Management Inventory–Acute Coronary Syndrome Version. *Canadian Journal of Cardiovascular Nurses, 24*(2), 11–18.

Sandelowski, M. (1986). The problem of rigor in qualitative research. *Advances in Nursing Science, 8*(3), 27–37. https://doi.org/10.1097/00012272-198604000-00005.

Sandelowski, M. (2015). A matter of taste: Evaluating the quality of qualitative research. *Nursing Inquiry, 22*(2), 86–94. https://doi.org/10.1111/nin.12080.

Sidani, S., Epstein, D. R., & Fox, M. (2017). Psychometric evaluation of a multi-dimensional measure of satisfaction with behavioral interventions. *Research in Nursing & Health, 40*(5), 459–469. https://doi.org/10.1002/nur.21808.

Spiers, J., Morse, J. M., Olson, K., Mayan, M., & Barrett, M. (2018). Reflection/commentary on a past article: Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 17*(1-2). https://doi.org/10.1177/1609406918788237.

Tobin, G. A., & Begley, C. M. (2004). Methodological rigour within a qualitative framework. *Journal of Advanced Nursing, 48*, 388–396.

Vincelette, C., Quiroz-Martinez, H., Fortin, O., & Lavoie, S. (2019). Preliminary development and validation of the Nurse Cardiopulmonary Resuscitation Survey (NCRS) among intensive care unit nurses. *Official Journal of the Canadian Association of Critical Care Nurses, 30*, 24–31.

Waltz, C., Strickland, O., & Lenz, E. (1991). *Measurement in nursing research* (3rd ed.). Philadelphia: F.A. Davis.

Williams, V., Boylan, A. M., & Nunan, D. (2020). Critical appraisal of qualitative research: necessity, partialities and the issue of bias. *BMJ Evidence-Based Medicine, 25*(1), 9–11. https://dx.doi.org/10.1136/bmjebm-2018-111132.

CHAPTER 16

Qualitative Data Analysis

Lorraine Thirsk | Sarah Stahlke

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Examine the processes of qualitative data analysis.
- Outline the steps common to qualitative data analysis.
- Describe how data are interpreted to form meaningful units (themes).
- Summarize the process of identifying themes and categories and the relationships between them.
- Compare the process of creating and presenting interpretations from select qualitative methods.
- Assess the integrity of data analysis from a qualitative study.

KEY TERMS

codes
coding
constant comparative method
data analysis
data display
data reduction
member checking
thematic analysis
themes

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

QUALITATIVE DATA ANALYSIS is a multifaceted, complex, and systematic process. The data generated through a qualitative research project can include transcripts of interviews, narratives, documents, photographs, media such as newspapers and movies, and field notes. Qualitative researchers collect enormous amounts of data, which must be managed carefully; several hundred pages of transcript can result from 25 interviews. To add to the complexity of qualitative data analysis, many researchers take different approaches to analysis, based on the purpose of the study and the conceptual framework or methodology used. This chapter expands on the discussion in Chapter 9, in which the analysis of data was introduced in the context of several qualitative research traditions, such as phenomenology, grounded theory, ethnography, and qualitative description. General principles and common types of analysis are reviewed, along with specific examples of some common analytical processes and considerations.

DATA MANAGEMENT

The open nature of qualitative inquiry typically results in the collection of more data than required. Glesne (2011) refers to the sheer volume of the data collected as "fat data." Consequently, researchers must be methodical in their organization and management of the data. Some researchers will organize all of these data by hand, but computer software can also be used to simplify the storage and retrieval of data. In addition, researchers are also required to develop a decision or audit trail, which necessitates the tracking of the participants, the original audio recordings, and original and photocopied documents. Moreover, all the data must be kept secure to maintain confidentiality.

Audio Recording Interviews

As discussed in Chapter 8, qualitative researchers gather data from a variety of sources, including interviews, observations, narratives, and focus groups (Merriam & Tisdell, 2016). Interviews are the most common source and serve as the primary source of data for many qualitative research projects. Typically, interviews are audio-recorded and transcribed to facilitate analysis. For example, Pauly et al. (2015) interviewed 34 people in a private setting and audio recorded the meeting. Although some researchers believe that a recording device inhibits the free flow of discussion, Seidman (2013) and other authors have found that most participants and interviewers forget about the presence of the device. Consequently, most researchers record interviews and then transcribe them verbatim into written text. Some researchers may consider summarizing or paraphrasing the spoken words (Seidman, 2013), but this is not commonly practised. Most researchers wish to use the original words from the participants so that the researcher's recall of the interview dialogue and their own interpretations do not become intertwined with the participant's thoughts. The presence of the original words allows the reader to check the authenticity of the data. New researchers may transcribe the recording into text themselves; however, most researchers use a transcriptionist. It is recommended that the researcher spot-check interviews to ensure accuracy of the transcription.

Electronic Data Management and Software

Regardless of the type of qualitative data collected, it is usually converted to an electronic format (Averill, 2015) to facilitate analysis and sharing with collaborators. Computer software to organize and retrieve data is referred to as computer-assisted qualitative data analysis software (CAQDAS). There are many computer programs to choose from, such as ATLAS.ti, Ethnograph, HyperRESEARCH, Inspiration, QSR NVivo, QSR XSight, and C-I-SAID.

Unlike computer programs used with quantitative data, these programs do not analyze data. Data analysis and interpretation remain largely the task of the researcher. In other words, CAQDAS cannot "think for the researcher" (Glesne, 2011, p. 207). However, using computer programs for orderly organization and grouping of data facilitates the researcher's job of analysis and interpretation.

Research Hint

There have been advances in text analysis software, called natural language processing (NLP), that may assist in the analysis of large volumes of textual data. Renz et al. (2018) suggested this might be helpful to analyze text from larger data sets and can also be used for triangulation purposes along with conventional analysis. NLP can examine the arrangement, frequency, and types of words used but does have limits. As in all research, it is important that the method fits with answering the research question and that there is methodological coherence.

All data must be backed up and stored in multiple places, such as a cloud storage site, while ensuring security of the data to protect participant confidentiality. Lost data cannot be replaced easily.

There have been some concerns raised about the use of CAQDAS and the effect on the quality of analysis. Maher et al. (2018) compared analysis techniques for a grounded theory study using traditional, manual methods (i.e., paper, highlighters, sticky notes, large display boards) with digital coding using CAQDAS. They argued that deep and insightful interaction with data is necessary for qualitative analysis and that interacting with data in a physical way allows for different perceptions of the material. Digital tools support a more sequential or mathematical mode of cognition, whereas manual methods support more relational and contextual modes of cognition (Maher et al., 2018). This slower and more kinetic mode of interacting with the data was found to be preferential in terms of developing interpretive insights; digital analysis resulted in a smaller and more fragmented view of the data that also made it more difficult to discuss and reflect on interpretations with colleagues (Maher et al., 2018).

Data Repositories and Sharing Data

Data repositories can be used to securely store research data and facilitate sharing of data. Research funding agencies as well as editors of academic journals encourage researchers to make their data available to other researchers for either replicability studies or systematic reviews (Chauvette et al., 2019). Field notes are also an important source of data in qualitative research and may also be useful in secondary data analyses, such as meta-synthesis (Phillippi & Lauderdale, 2018). There are some ethical, methodological, and legal concerns to keep in mind when considering making original qualitative data available to others (Chauvette et al., 2019), and often there is compelling ethical rationale to keep it private (Guishard, 2018). This is particularly important in research with Indigenous communities, where it may be the community and not the researcher who controls the data.

OVERVIEW OF DATA ANALYSIS

The purpose of **data analysis** is to answer the research question (Merriam & Tisdell, 2016). Although analysis among different qualitative approaches differ, there are a few general processes: preparing and organizing the data, reading (and rereading) the database, organizing or reducing the data through coding or themes, representing the findings in figures, narratives, or tables, and forming interpretations (Creswell & Poth, 2018). When does data collection end and data analysis begin? Most commonly in qualitative research, data collection and analysis are done concurrently, although obviously at least some data needs to be collected prior to analysis. Many researchers believe that the stages of data collection and data analysis should be integrated (Denzin & Lincoln, 2000; Merriam & Tisdell, 2016; Miles et al., 2014; Streubert & Carpenter, 2011), whereas others believe that these stages should be separate (Seidman, 2013). This will be largely dependent on the method of qualitative research, and the researcher should explain this process in the study.

Many researchers begin a preliminary analysis as the material accumulates. Typically, the qualitative researcher transcribes all of the interviews, field notes, and observations as they are collected. As each piece of data is transcribed, researchers begin a preliminary analysis during which they determine what additional data need to be collected.

Qualitative researchers look for "insight, meaning, understanding, and larger patterns of knowledge, intent, and action" in the data (Averill, 2015, p. 1). Patton (2002) encouraged researchers to do their "very best with . . . full intellectual capacity to fairly represent the data and communicate what the data reveal given the purpose of study" (p. 433). Recall from previous chapters that qualitative research is often an inductive process—where specific details are examined to generate theory or overarching explanations. As described earlier, qualitative analysis is not a linear process; rather, it is cyclical, transformative,

reciprocal, and iterative. Miles et al. (2014, p. 10) have identified some common features among different approaches to qualitative data analysis:

1. Affixing codes or themes to a set of field notes, interview transcripts, or documents
2. Sorting and shifting though these coded materials to identify similar phrases, relationships between variables, patterns, themes, distinct differences between subgroups, and common sequences
3. Isolating these patterns and processes, and commonalities and differences, and taking them out to the field in the next wave of data collection
4. Noting reflections of other remarks in the margins
5. Gradually elaborating a small set of assertions, propositions, and generalizations that cover the consistencies discerned in the database
6. Confronting those generalizations with a formalized body of knowledge in the form of constructs or theories

Guidelines such as these are useful, but they serve only as recommendations. Each qualitative study is unique and is reliant on the creativity, intellect, style, and experience of the researcher. Another example of the steps of qualitative analysis is presented in Figure 16.1.

During the data analysis phase, all researchers fully immerse themselves in the data over a period of weeks to months. This process requires constant reading and rereading of the text until an understanding is reached about what the data convey (Polit & Beck, 2017). Many researchers also listen to the recorded interviews several times to increase their understanding and to align the emotive component. For example, during the interviews in the study by Pesut et al. (2020; Appendix A), the transcripts included the emotions, such as crying, that were evident in the interviews. In addition, the principal researcher listened to the audio recordings. Observations written by the researcher during the interviews can capture these important elements as well. An important part of the data analysis is the interplay between data gathering or questioning and verifying what is heard and understood. Researchers continue to ask whether what they understood before is still relevant after subsequent interviews, observations, and reading of related documents. This "cyclic nature of questioning and verifying is an important aspect of data collection and analysis" (Streubert & Carpenter, 2011, p. 46).

Miles et al. (2014) refer to three discrete stages of data analysis: data reduction, data display, and conclusion drawing and verification (Figure 16.2). Many of the common methods used in nursing research fit into this general view of qualitative analysis.

Data Reduction

According to Miles et al. (2014), **data reduction** is "the process of selecting, focusing, simplifying, abstracting, and transforming the data that appear in written-up field notes or transcriptions" (p. 10). This process is ongoing as data are collected. Glesne (2011) suggested three steps to help researchers analyze qualitative data. The first step is to write memos during the data-collection stage,

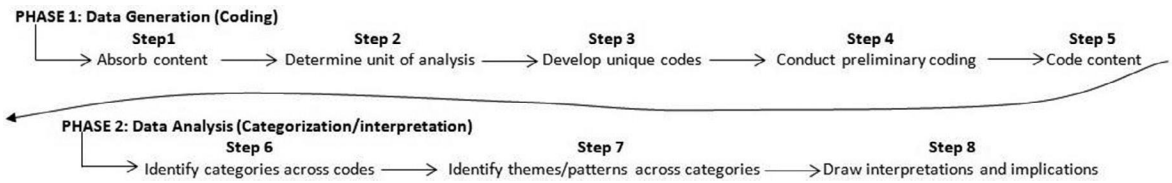


FIG. 16.1 Phases of data analysis. From: Roller, M. R. (2019). A quality approach to qualitative content analysis: Similarities and differences compared to other qualitative methods. *Forum: Qualitative Social Research, 20*(3), 1–21. https://o-doi-org.aupac.lib.athabascau.ca/10.17169/fqs-20.3.3385.

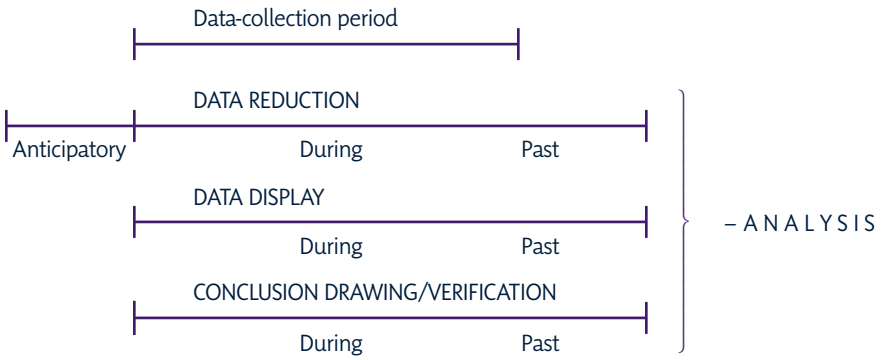


FIG. 16.2 Components of data analysis: Interactive model. From Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage (Display 1.1, p. 14).

which allows researchers to record thoughts about the data as these thoughts occur. Analytical files are developed to sort data into general categories, such as interview questions, people, and places, as well as useful quotations from the interviews and relevant quotations from the literature. These files help organize researchers' thoughts and those of others. This is sometimes referred to as reflective journaling and can be helpful to track decision-making, reflect on emotions, and note thoughts about what is happening in the research (Polit & Beck, 2017).

Next, Glesne (2011) recommends the development of rudimentary coding schemes. **Coding** is a progressive marking, sorting, resorting, and defining and redefining of the collected data. Coding allows researchers to transform the "unstructured and messy data to ideas about what is going on in the data" (Richards & Morse, 2007, p. 133). The data can then be organized into meaningful clusters of data by grouping related or similar data. Often, these clusters or groups of data are labelled as **themes**, or structured meaning units of data that occur frequently in the text. **Thematic analysis**—the process of recognizing and recovering the emergent themes—is an important aspect of organizing data. Coding is a large part of qualitative data analysis and will be described further in the next section.

Evidence-Informed Practice Tip

"Coding is nothing more than assigning some sort of shorthand designation to various aspects of your data so you can easily retrieve specific pieces of data" (Merriam & Tisdell, 2016, p. 199).

Lastly, Glesne (2011) recommends that researchers write themselves monthly field reports as a way of systematically reviewing the progress and determining the next steps. Aside from helping researchers keep track of their progress and communicate progress with other members of the research team, monthly summaries often result in new insights and new ways of approaching the research. This also helps to develop the audit trail and track decision-making, which is an important part of rigour in qualitative research (see Chapter 15).

Coding

There is considerable variety in the process of coding, depending on the method, disciplinary orientation, and philosophical basis for the research. Denzin and Lincoln (2000) describe the fundamental steps in coding of data as sampling, identifying themes, building code-books, and marking texts. Richards and Morse (2007) describe three types of coding: descriptive, topic, and analytic. Researchers use some, or all, of these types of coding. *Descriptive coding* helps

the researcher keep track of factual knowledge (e.g., gender). In *topic coding*, used most commonly, the data are grouped together by topic "to reflect on all the different ways people discuss particular topics, to seek patterns in their responses, or to develop dimensions of that experience" (Richards & Morse, 2007, p. 134). As the categories become more complicated, the topic coding becomes analytic. *Analytic coding* is more theoretical and leads to the development of themes. Although coding may sound complicated to you, remember that this process is evolutional, and it varies from project to project and from researcher to researcher. For example, many researchers conducting narrative inquiry do not use coding, data reduction, and some of the other commonly used methods of data analysis.

Miles et al. (2014) described the process of coding taking place in two steps. The first is described as *first cycle coding*, in which the data are assigned to data chunks. While there are many types of coding, Saldaña (2013) suggests three foundational types of coding in this first step:

- Descriptive coding: labels are assigned, composed as a short phrase or word
- In vivo coding: short phrases or words are drawn from the participants' own language
- Process coding: "ing" is used to describe observations or actions (e.g., knowing)

The second step, finding themes, occurs during and after data collection. These themes or basic units of analysis can be entire texts (e.g., interview transcripts, responses to surveys), grammatical segments (words, phrases, sentences, paragraphs), formatting units (rows, pages), or clusters of texts that reflect a single theme. Most researchers try to divide data into units of analysis that do not overlap with others. Researchers approach this step in a variety of ways; for example, experts in grounded theory recommend that the researcher read the text line by line. Miles et al. (2014) describe this second phase of coding as pattern codes. While the first cycle summarized chunks of data, in this second cycle the first-cycle codes are organized into a smaller number of categories, themes, or constructs.

The coding process itself is analysis (Miles et al., 2014). **Codes** are simply tags or labels that are assigned to the themes; often, the code itself is only one to four words long. Major codes may exist along with subcodes. Codes evolve during the analysis; more may be added, and others may be blended. They mean something to the researcher and are not typically included in the research report. As the coding and themes are fine-tuned and finalized, much of the analysis is completed.

Richards and Morse (2007) described two primary steps to data analysis: categorizing and conceptualizing: "Categorizing is how we understand and come to terms with the complexity of data in everyday life" (p. 155). Coding is one method for categorizing the data: however, other researchers in qualitative studies can think about data without coding. Conceptualizing moves up the ladder of abstraction (see Chapter 2) to build frameworks of concepts or theory. It is a process of forming theoretical definitions to "make sense" or organize the data. Phenomenology, ethnography, and grounded theory are all methods that necessitate conceptualization.

In Table 16.1, the differences in the methods of abstraction are described by means of the following questions:

- When does abstraction occur?
- Where does abstraction come from?
- How is abstraction done?
- What analytical outcome is being sought?

While not all qualitative methods are concerned with identifying themes, thematic analysis is common across several qualitative approaches. Thorne (2020) argued that finding themes is only an initial step in data analysis and that further analysis and interpretation are needed in order for the research to make a useful contribution to nursing, although it may not always be necessary to generate a new theory. Braun and

TABLE 16.1

DOING ABSTRACTION IN THREE DIFFERENT METHODS

METHOD	WHEN DOES ABSTRACTION OCCUR?	WHERE DOES ABSTRACTION COME FROM?	HOW IS ABSTRACTION DONE?	WHAT ANALYTICAL OUTCOME IS BEING SOUGHT?
Phenomenology	Not until one has the data: previous ideas and knowledge may be bracketed	Themes and meanings in accounts, texts	Deep immersion, focus, thorough reading	To describe the essence of a phenomenon
Ethnography	Prior knowledge of site, situation; understanding develops during field research	Knowledge of social and economic setting; observation and learning from the setting	Rich description; combination of qualitative and quantitative patterning, coding, comparing, reviewing field notes	To identify themes and patterns; to explain and account for a social and cultural situation
Grounded theory	Abstraction is from the data but can be informed by previously derived theories	Categories derived from data (observations or line-by-line analysis of texts); constant comparison with other situations or settings	Theoretical sensitivity; seeking concepts and their dimensions; open coding, dimensionalizing, memo writing, diagramming	To identify a core category and theory grounded in data

From Morse, Janice M.; Richards, Lyn. (2007). *Read me first for a user's guide to qualitative methods.* Thousand Oaks, CA: SAGE Publications Inc.

Clark (2006) described two levels of thematic analysis: semantic and latent. In semantic analysis, there is little beyond what the participant said, although the interpretation may also include theorizing about the significance of patterns that were selected from the data. In latent analysis, the goal is to "identify or examine underlying ideas, assumptions, and conceptualizations—and ideologies—that are theorized as shaping or informing the semantic content" (Braun & Clarke, 2006, p. 84). When reading qualitative research, notice the level of analysis that has been conducted. Do the findings reflect themes that are still close to the description provided by participants, or is there a significant amount of interpretation? Do the researchers focus on understanding the hidden assumptions behind what is said, or is the data taken at face value? Understanding this will help you to determine the process and level of analysis.

Evidence-Informed Practice Hint

Thematic analysis may lend itself to track repeated patterns and frequency or prevalence of concepts found in qualitative data (Braun & Clark, 2006). These may be described in the findings as common themes, or that several participants reported a similar idea. In some qualitative research methods, such as hermeneutics, ideas and concepts that are unique to one participant may be the foundation of analysis. This is called the "fecundity of the individual case" (Jardine, 1992).

Rigour in Data Reduction and Abstraction

Increasingly, there is an expectation that researchers explicitly describe the process of data analysis. For example, Bourque Bearskin et al. (2016; Appendix C) wrote about her own experiences and background, to position herself in context of the research. Indigenous research methodologies were explained and referenced in the section titled "Research Framework." The "Data

Analysis" section described the iterative process of analysis, as well as the use of member checking and peer debriefing. The focus of the research study presented in Appendix C was the finding of *Ontological Beginnings and Epistemological Openings.* This included two subthemes of *Early roots of Indigenous knowledge* as well as *Integrating roots of knowledge into nursing practices.* In the presentation of the analysis, Bourque Bearskin et al. included participant quotes as well as referencing other published information to help explain and expand the analysis.

In the study by Pesut et al. (2020; Appendix A), the authors stated they followed interpretive description in their analysis. They further described how the audio recordings were handled, who developed initial codes, and how subsequent analysis refined these codes. They found three themes in their data: *Systems, Teams,* and *Processes.* The results section presents their analysis according to these themes and includes several participant quotes that help to further explain and explore the findings. All of the participant quotes are coded so that the participants' confidentiality is maintained, and the reader gets a sense of the numerous perspectives represented.

In their focused ethnography of managers' roles in supporting teamwork, Stahlke and Dahlke (2020) followed line-by-line coding, found persistent words and phrases, and grouped these ideas into themes. These themes, along with exemplary quotes from participants, can be found in Table 16.2.

One of the major pitfalls in qualitative and, particularly, thematic analysis is the use of interview questions to organize the analysis (Braun & Clark, 2006). Braun and Clark (2006) claim that this represents a failure to do *any* analysis. They described three types of questions that are used in qualitative research:

1. The main research question that guides the inquiry.
2. The questions asked of participants during the interviews.
3. The questions asked of the data during the analysis.

Thorne (2020) suggested that this third set of questions should include things like "Why have we noticed these patterns?" and not just a reporting of the patterns. Our cognitive processes are attuned to find patterns and meaning in events and experiences, and whether these are significant,

TABLE 16.2

EXAMPLES OF THEMES FROM FOCUSED ETHNOGRAPHY

THEME	ILLUSTRATIVE QUOTES
Manager visibility	"I just really think [name of supervisor] is amazing. I truly do. She is our manager, but I have seen her in the kitchen, cooking food, preparing things…She doesn't have to; she could stay in her office. I think that helps the team. It's like that [idea of] walk a mile in my shoes." (Activity Aide)"The willingness is there, which is amazing…It definitely motivates staff; I think even frontline staff see it as 'Oh they don't just sit in their office.'" (Recreation Therapist)"The entire leadership team is continually trying to figure out how we can be more approachable." (Nurse Educator)
Conflict management	"So, when you start sweeping things underneath the carpet, you're going to get a great big bump and someone's going to trip…We need strong management that says 'this is the way it is.'" (HCA)
Organizational values	"What interferes with teamwork? Being totally frustrated with the system. Seeing the business side of it…having the business side run over the care side." (HCA)"The policy here, even my director of care will say, is residents first…before any paperwork or any doctor's order, residents first." (LPN)

Source: From Stahlke, S., & Dahlke, S. (2020). The relational role of managers in support of teamwork. *Nursing Leadership, 33*(1), 112–121.

or exist, needs careful reflection and attention (Buetow, 2019).

Research Hint

Qualitative researchers will often include their interview guide in the publication of the research report. Review the interview guide and see how it compares to the results section of the research study. Did the authors use their interview questions to organize their themes, or did they move beyond this analysis, asking different questions of the data and presenting new information?

Data Display

The next major step in data analysis is the data display. Miles et al. (2014) define data display as "a visual format that presents information systematically so the user can draw conclusions and take needed action" (p. 108). This display helps the researcher understand the data and can be in the form of graphs, flowcharts, matrices, or any other visual representation. Like the rest of the analysis, the data display changes as more is known about the phenomenon under study. For example, Thirsk et al. (2021) developed a taxonomy from their data examining supports and barriers to family-centered care in adult critical care areas. Two broad domains were identified (people and structure), and each of these domains included four concepts and numerous subconcepts. This framework evolved as more data were collected and analyzed and through many discussions with the research team. The final data representation can be seen in Figure 16.3.

Although many researchers use figures and charts as part of their data display, profiles or vignettes can also display what is to be learned from the participant's experience. Vignettes of the participant's experience can summarize what was learned from each participant and can then be shared with each participant for validation (Seidman, 2013). This narrative form transforms the text into a story—a compelling way of sharing meaning.

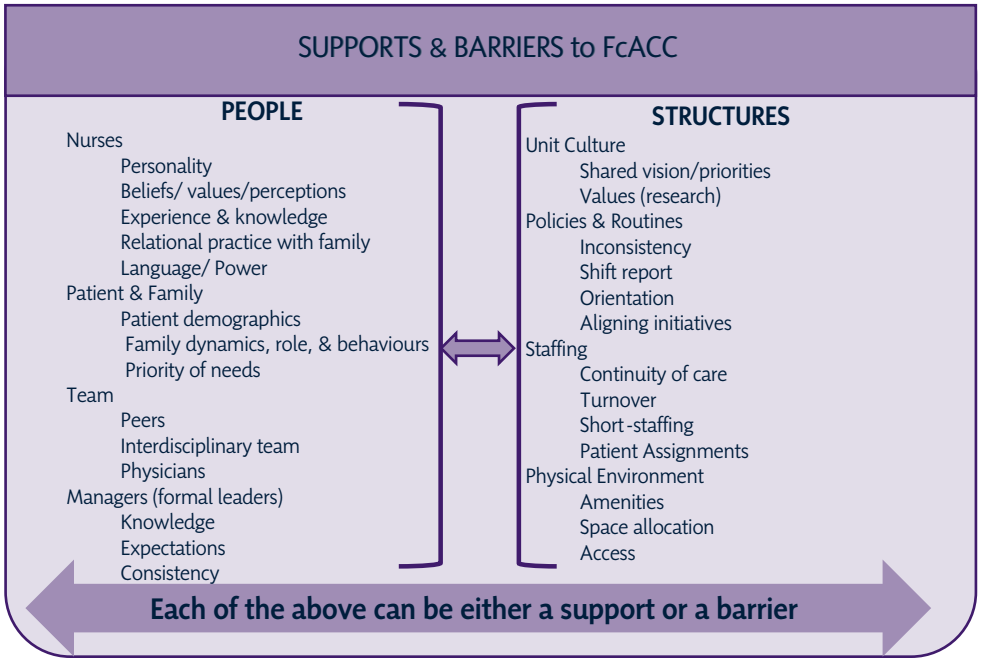


FIG. 16.3 Taxonomy of supports and barriers. From Thirsk, L.M., Vandall-Walker, V., Rasiah, J., & Keyko, K. (2021). A taxonomy of supports and barriers to family-centered adult critical care: A qualitative descriptive study. *Journal of Family Nursing*. https://doi.org/10.1177/1074840721999372.

For example, Smith et al. (2018) studied the experiences of mothers whose children were in long-term treatment for substance use disorders. This narrative inquiry engaged four women to understand their experiences and explore the "dominant and competing stories of motherhood, family, and substance abuse" (p. 512). Vignettes were developed with each of the four participants as a way of introducing the reader to the participants and their story. An example of one vignette is as follows:

> Mary's son was 7 months into treatment when we first met. Substance abuse entered Mary's life in an unexpected way. Mary grew up living the story of a happy child in a loving carefree family. She explained, "I thought this is how all families were supposed to be—happy, loving, supportive, and carefree—I didn't know addiction as a child." Life changed for Mary when she married a man who later became dependent on drugs. Mary described feeling confused and rejected. She told of a painful journey of a wife and mother who struggled to understand her husband's and later her son's substance abuse. Mary did not know how to make sense of her identity as a wife and mother amid growing family troubles. Mary began treatment embarrassed by her inability to "fix the problem," and she felt too ashamed to ask for help. In treatment, Mary became more aware of substance abuse and of herself as she connected with support. She said, "I was given permission to stop trying to fix, to change, or control my son. I am now focused on what I feel." Mary left treatment looking ahead to the ongoing emotional work of living amid substance abuse in her family context. (Smith et al., 2018, pp. 514–515)

Rich descriptions, such as those found in narratives or direct quotations, enliven the data and give meaning to people's experiences. Most qualitative research includes selected quotations to illustrate the themes and to provide readers with the opportunity to understand and validate the themes chosen by the researcher. For example, Thirsk et al. (2021) selected the following quotes to demonstrate how the relationship between the nurse and the family influences family-centered care and that not all nurses make connections with all families.

I-5: You'll see certain family members not want to leave the patient's side when certain nurses are working with them. And then, if you have a comfortable dynamic, they'll go home to sleep or have a shower…they feel comfortable [enough] to leave.

I-5: I find that families will find a nurse that they connect with. And then, although that might not be the nurse assigned that particular day, they'll seek them out to ask questions.

As another example of rich description, Table 16.3 includes selected quotations from participants to support the themes emerging from Burns et al.'s (2019) study examining Mi'kmaq women's access to prenatal care.

When the data are presented, the most important consideration for the research is to ensure that the presentation supports the findings and relays what needs to be known (Streubert & Carpenter, 2011). The analysis and presentation need to be convincing to the reader, who has not read the entire data set (Braun & Clark, 2006). The purpose of the study determines how the story is told. If the method is descriptive phenomenology, the focus is on the description of the lived experiences, whereas in a grounded theory study, the focus is on a more careful description of how the narrative gives rise to the analysis and interpretation, which results in theory development.

CONCLUSION DRAWING AND VERIFICATION

Conclusion drawing starts at the beginning of data collection but is not finalized until the project is completed. Although qualitative research is inductive, it is tempting to draw conclusions prematurely. The challenge for the researcher is to remain amenable to new ideas, themes, and concepts as they appear.

Conclusion drawing is essentially the description of the relationship between the themes. Richards and Morse (2007) describe this process

TABLE 16.3

EXAMPLES OF SELECTED QUOTATIONS TO SUPPORT THE THEMES

THEMES	SUBTHEME	EXAMPLE OF A QUOTATION
Closing the gaps in prenatal care	Traveling the distance	*This is her first baby. Yeah, so she's going into it blind* (referring to the lack of prenatal classes in the community).
Social support networks during pregnancy	Family support	*So we talk a lot about our own experiences and what we did, and what happened to us, and what we would suggest for other women.*
	Professional support	*If I had any concerns about my pregnancy, I would go to the Health Centre first because we have a Community Health Nurse and the Family Support Worker. They're really accommodating. If you need them, they're there.*
Cultural beliefs and preferences	Importance of traditions	*Prenatal care means taking care of your body, but also your mind and your spirit when you're pregnant.* *It is an honour to sweat with a pregnant woman because she is bringing life into the world.*
	Providing culturally safe care	*I found at the hospital here they're very open and embrace our differences. There was no limit to how many people in the room when I was in labour. Which I know, you know, they don't allow for everyone but it's important for us. Just very respectful.*

From Burns, L., Whitty-Rogers, J., & MacDonald, C. (2019). Understanding Mi'kmaq women's experiences accessing prenatal care in rural Nova Scotia. *Advances in Nursing Science, 2*, 139. https://doi.org/10.1097/ANS.0000000000000248.

as "doing abstraction" (p. 158), in which data are moved from categories (codes and themes) to concepts and constructs. As discussed earlier and shown in Table 16.1, the ways of abstracting vary with the type of method. Grounded theory formalizes this stage through the development of models, which lead to theory. Verification occurs as the data are collected; this process can vary from questioning one's own conclusion through the rechecking of the text to verification by colleagues and to finding new cases and applying the model to them. In grounded theory and many other qualitative methods, researchers use the **constant comparative method**, in which new data are compared as they emerge with data previously analyzed.

Miles et al. (2014) have stated that this process of making sense of the data is a skill that all nurses have. People make sense of the world around them by organizing and interpreting it; this skill is applied to drawing and verifying conclusions. Miles et al. (2014) list the following 13 tactics for analyzing the data (pp. 277–278):

1. Noting patterns and themes (repetitive or recurring patterns among many separate pieces of data)
2. Clustering (grouping together things that seem to share characteristics)
3. Making metaphors (using a literary device in which different things are compared to make sense of the experience)
4. Counting (noting that something is happening a number of times)
5. Making contrasts or comparisons (comparing sets of things)
6. Partitioning variables (breaking down the themes into smaller units)

7. Subsuming particulars into the general (using a higher level of abstraction)
8. Factoring (generating words [factors] to express common findings)
9. Noting relationships between variables (depicting the relationships between the findings)
10. Finding intervening variables (discerning other variables that may link findings together)
11. Building a logical chain of evidence (validating each of the relationships identified)
12. Making conceptual or theoretical coherence (linking the findings into an overarching "how" and "why" of the phenomenon under study)

Refer to Miles et al. (2014) for more detail about these tactics. Merriam and Tisdell's (2016) text on qualitative research also described the step-by-step process.

To verify their findings, Thirsk et al. (2021) moved back and forth between the emerging categories of the taxonomy and the interview transcripts. This involved rereading transcripts and revising the taxonomy through numerous discussions with the research team until all of the reported barriers and supports were documented. Other researchers can use different methods to validate their themes. Pesut et al. (2020) drew on information gathered from a systematic review and analysis of policy documents to help situate their qualitative findings and explain the consequences of legislative decisions. Stahlke and Dahlke (2020) built variation into their sample, created thick descriptions of their participants' experiences, documented their decision-making during analysis, and discussed emerging ideas with each other.

No matter what method is used, researchers ask themselves, "What have I learned? How do I understand this, make sense of it and see the connections in it?" (Seidman, 2013). The conclusions drawn are simply to "describe, make contributions and contribute to greater understanding, or at least, more informed questioning" (Glesne, 2011, p. 210). As discussed in Chapter 7, through the processes of reflexivity, researchers constantly compare their findings with their own personal beliefs and knowledge to ensure that the analysis reflects the participants' beliefs rather than their own.

SPECIFIC ANALYTICAL PROCEDURES

The processes of data analysis vary according to the type of qualitative research. Table 16.4 summarizes the methods of analysis in qualitative methods, including phenomenology, ethnography, grounded theory, and qualitative description. Excerpts from Canadian studies are included to exemplify the methods.

Research Hint

In participatory action research the participants, or collaborators, in the project are often involved in the process of data analysis. For example, Doré et al. (2018) involved hemodialysis nurses in their PAR on nursing burnout. After phase one focus groups, two nurses from the advisory team were included in the initial discussion to strengthen the analysis and interpretations. This first round of analysis was then presented in subsequent focus groups for further discussion, and the final analysis resulting from these discussions was also confirmed with the advisory team. Engaging participants in this manner means the participants have some control of the analysis, rather than this lying solely with the research team.

TRUSTWORTHINESS

As described in Chapter 15, rigour in qualitative research is determined by credibility, auditability, and fittingness as the criteria for evaluation, although there is still a debate about the most appropriate terminology (see Morse et al., 2015). Trustworthiness is important for determining the integrity, or soundness, of the data interpretation or analysis. To ensure the trustworthiness of their findings, qualitative researchers must ask themselves the following questions (Hollway & Jefferson, 2000):

- What do you notice? The researcher has captured some impressions about the data; however, information may be missing. Detailed or thick descriptions of the phenomenon also allow the reader to assess whether the account "rings true."
- Why do you notice what you notice? Researchers must consider their own biases and

TABLE 16.4

METHODS OF ANALYSIS AND EXEMPLARS

TRADITION	METHOD OF ANALYSIS	EXAMPLE
Phenomenology: includes a variety of traditions	• Immersion in the data: listen to recordings, read and reread transcripts • Identify and extract significant statements • Determine relationships among the extracted statements (themes) • Prepare exhaustive description of the phenomenon and the relationships among the themes • Synthesize the themes into a consistent description or statement of the phenomenon under study (essence)	Lamb et al. (2019) "Conscientious objection and nurses: Results of an interpretive phenomenological study. "These authors explored nurses' experiences of making a conscientious objection to further understand the moral choices and encounters with ethical issues. In this phenomenological approach, analysis is continuous and builds off previous interpretation. They organized the data into six themes: encountering the problem, knowing oneself, taking a stand, alone and uncertain, caring for others, perceptions of support. The authors offer an example of a quote that supports the theme of knowing oneself: *"Adding the word objection to it I feel like it has so much more power . . . you know, kind of doing what is wrong and questioning what's right, but I feel like this has more positive connotation to it, like you have a choice, you're not feeling this residual distress and you just have to accept that's the way it is. I think it almost, it's a little empowering.*
Ethnography	• Immersion in the data • Identify patterns and themes • Complete a cultural inventory • Interpret the findings • Compare the findings with those in the literature	Sutherland et al. (2017) "Structural impact on gendered expectations and exemptions for family caregivers in hospice palliative home care" Pauly et al. (2015) collected data through interviews, observations, and document reviews. Initial codes were developed through multiple readings of the data, and codes were compared to develop themes. All of the co-authors were involved in every aspect of the data analysis, which resulted in the development of three themes.
Grounded theory	• Examine data carefully line by line • Divide data into discrete parts • Compare data for similarities and differences • Compare data with other data continuously in a process: constant comparative method • Cluster codes to form categories • Expand and develop categories or collapse them into one another • Determine relationships between categories	King-Shier et al. (2019) "South Asians' experience of managing hypertension: a grounded theory study. "The authors implemented the main features of grounded theory such as theoretical sampling, continual comparison of data and codes, which influenced questions asked of participants, line-by-line reading, axial and selective coding. The results were modelled in a diagram showing the process of managing hypertension.
Qualitative Description	• Identify themes, primarily from textual data from interviews • Create a description that would be meaningful to others. • Provide representative quotes to explore complexity of phenomenon.	Stahlke Wall (2018) "The impact of regulatory perspectives and practices on professional innovation in nursing". The author analyzed interviews and categorized findings into several themes: problematic regulatory processes; specific process issues; and impacts of conservative regulation. In addition, a negative case was included in the analysis because one participant's experience with the nursing regulator was different than the other participants. This exploration of a negative case helps to recognize the variance in and complexity with the phenomenon.

predispositions as they interpret the data to produce trustworthy interpretations. Many researchers use a journal to document their reflections to monitor their own developing interpretations.

- How can you interpret what you notice? Credibility stems from prolonged engagement and persistent observation. To be able to complete a full interpretation, the researcher must spend a sufficient amount of time in the field to build sound relationships with the participants.
- How can you know that your interpretation is the "right" one? In some qualitative methods, researchers will use **member checking.** Member checking is the collaborative process of involving participants in data analysis by verifying and refining data and research findings for congruency with their experiences and meanings (Gillis & Jackson, 2002, p. 216). It is used more in descriptive qualitative research, where the analysis stays "close" to the participants language, and is often important in research with Indigenous peoples and communities. The further the abstraction, the less useful member checking is. The researcher is also checking whether the connections between the categories or themes are logical. Inviting other experts to review the data analysis is another option for many researchers. This occurs primarily in research teams when collaborators review, discuss, and make decisions about the analysis. Reviewing the data and checking the interpretations also occurs during the publication process as peers review the presented research study. In addition, some researchers analyze their data from several different frameworks (a form of triangulation) to increase the trustworthiness of the data analysis.

Finally, it is important to consider the limitations of the study. Many researchers describe the issues they faced so that readers will understand the research in the proper context (Glesne, 2006).

APPRAISING THE EVIDENCE

Qualitative Data Analysis

The general criteria for critiquing qualitative data analysis are proposed in the Critiquing Criteria box; however, remember that many different approaches to data analysis exist. The data analysis is consistent with the research methodology, the question, and the design. For example, researchers using grounded theory build a case for substantive theory, explaining the phenomenon under study, whereas a researcher in phenomenological studies is interested in expressing the meaning of the phenomenon itself.

Regardless of the study's research method, several commonalities exist among methods used in qualitative data analysis. For example, analysis is conducted alongside the data collection, and in most cases the two processes are interrelated. Researchers become immersed in the data; they listen over and over to the interviews, read and reread the transcripts, and spend substantial time in the field. Although the methods may differ, the text is coded to search for themes and categories through a process of data reduction. As themes emerge, logical connections and relationships between the themes are identified to form a whole picture. The results are displayed in such a manner that the reader can understand and validate the conclusions that the researcher has drawn through the use of diagrams, tables, charts, direct quotations from the participants, and rich descriptions of the findings. In summary, qualitative data analysis involves much disparate data and transforms them into a coherent whole or story to provide and explanation of human experience.

CRITIQUING CRITERIA

1. The method of data analysis should be clearly stated.
2. The strategy of data analysis should be appropriate for the methodology of the study.
3. The steps of analysis should be listed for readers to follow.
4. The researcher should provide evidence that his or her interpretation captures the phenomenon under study.
5. The researcher should address the credibility, auditability, and fittingness of the data.

CRITICAL THINKING CHALLENGES

- Is it important for the researcher to personally transcribe the interviews?
- Why do some researchers reread the literature as themes emerge from the data?
- Often, data analysis takes place as data are collected. How can analysis of the data change the data collection?
- Some researchers validate their interpretation of the data through a process of member checking. What happens if the participants indicate that the analysis does not reflect their experience?

CRITICAL JUDGEMENT QUESTIONS

1. Why is it important for qualitative researchers to listen to the recorded interviews?
   a. It validates the participants' stories
   b. It allows for the addition of paraverbal data
   c. It is required by research ethics boards
   d. It improves triangulation of data
2. Which of the following is NOT part of qualitative data analysis?
   a. Concurrent data collection and data analysis
   b. Coding systems
   c. Immersive reading of data
   d. Testing hypotheses
3. How do researchers determine that data analysis is complete?
   a. When sufficient participants have been interviewed
   b. When the researcher is saturated
   c. When substantial analysis has been created
   d. When the research grant has been spent

KEY POINTS

- Qualitative data are text derived from transcripts of interviews, narratives, documents, media such as newspapers and movies, and field notes.
- Computer software can be used to simplify the storage and retrieval of data.
- Qualitative research data can be managed through the use of computers, but the researcher must interpret the data.
- Data analysis and data collection are parallel processes.
- Qualitative analysis is not a linear process; rather, it is a cyclical and iterative process.
- The three discrete stages of data analysis are data reduction, data display, and conclusion drawing and verification.
- Data are organized into meaningful chunks of data through a clustering of related or similar data and are labelled as themes.
- Coding is the process of progressively marking, sorting, resorting, and defining and redefining the collected data.
- Data display involves the use of graphs, flowcharts, matrices, or any other visual representation to assemble data and to allow for conclusion drawing.
- Grounded theorists use the constant comparative method, in which new data are compared with data previously analyzed.
- Member checking, used in some qualitative methods, is the process of sharing findings with the participants in order to check whether the interpretation of the findings is accurate.

FOR FURTHER STUDY

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

REFERENCES

Averill, J. (2015). Qualitative data analysis. In M. De Chesnay (Ed.), *Nursing research using data analysis: Qualitative designs and methods in nursing*. New York: Springer Publishing Company.

Bourque Bearskin, L., Cameron, B. L., King, M., & Pillwax, C. W. (2016). Mâmawoh Kamâtowin, "Coming together to help each other in wellness": Honouring Indigenous nursing knowledge. *International Journal of Indigenous Health, 11*(1), 18–33. https://doi.org/10.18357/ijih111201615024.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. http://dx.doi.org/10.1191/1478088706qp063oa.

Buetow, S. (2019). Apophenia, unconscious bias and reflexivity in nursing qualitative research. *International journal of nursing studies, 89*, 8–13. https://doi.org/10.1016/j.ijnurstu.2018.09.013.

Burns, L., Whitty-Rogers, J., & MacDonald, C. (2019). Understanding Mi'kmaq women's experiences accessing prenatal care in Rural Nova Scotia. *Advances in Nursing Science*, *2*, 139. https://doi.org/10.1097/ANS.0000000000000248.

Chauvette, A., Schick-Makaroff, K., & Molzahn, A. E. (2019). Open data in qualitative research. *International Journal of Qualitative Methods, 18*. https://doi.org/10.1177/2F1609406918823863.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage Publications.

Denzin, N., & Lincoln, Y. (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Doré, C., Duggett-Leger, L., McKenna, M., Salsberg, J., & Breau, M. (2018). Participatory action research to empower hemodialysis nurses and reduce risk of burnout. *The CANNT Journal, 28*(3), 40–53. https://cannt-acitn.ca/cannt-journal/.

Gillis, A., & Jackson, W. (2002). *Research for nurses: Methods and interpretation*. Philadelphia: F.A. Davis.

Glesne, C. (2006). *Becoming qualitative researchers: An introduction* (3rd ed.). Don Mills, ON: Longman.

Glesne, C. (2011). *Becoming qualitative researchers: An introduction* (4th ed.). Toronto: Pearson.

Guishard, M. A. (2018). Now's not the time! Qualitative data repositories on tricky ground: Comment on Dubois et al. (2018). *Qualitative Pyschology, 5*(3), 402–408. http://dx.doi.org/10.1037/qup0000085.

Hollway, W., & Jefferson, T. (2000). *Doing qualitative research differently: Free association, narrative and the interview method*. Thousand Oaks, CA: Sage.

Jardine, D. W. (1992). The fecundity of the individual case: Considerations of the pedagogic heart of interpretive work. *Journal of Philosophy of Education, 26*(1), 51–61. https://doi.org/10.1111/j.1467-9752.1992.tb00264.x.

King-Shier, K. M., Dhaliwal, K. K., Puri, R., LeBlanc, P., & Johal, J. (2019). South Asians' experience of managing hypertension: A grounded theory study. *Patient Preference and Adherence, 13*, 321. https://doi.org/10.2147/PPA.S196224.

Lamb, C., Babenko-Mould, Y., Evans, M., Wong, C. A., & Kirkwood, K. W. (2019). Conscientious objection and nurses: Results of an interpretive phenomenological study. *Nursing Ethics, 26*(5), 1337–1349. https://doi.org/10.1177/F0969733018763996.

Maher, C., Hadfield, M., Hutchings, M., & de Eyto, A. (2018). Ensuring rigor in qualitative data analysis: A design research approach to coding combining NVivo with traditional material methods. *International Journal of Qualitative Methods, 17*, 1–13. https://doi.org/10.1177/1609406918786362.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation*. San Francisco: Jossey-Bass.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: Methods source book* (3rd ed.). Thousand Oaks, CA: Sage.

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research, 25*(9), 1212–1222. https://doi.org/10.1177/1049732315588501.

Patton, M. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Pauly, B., McCall, J., Browne, A. J., et al. (2015). Toward cultural safety: Nurse and patient perceptions of illicit substance use in a hospitalized setting. *Advances in Nursing Science, 38*(2), 121–135.

Pesut, B., Thorne, S., Schiller, C. J., Greig, M., & Roussel, J. (2020). The rocks and hard places of MAiD: A qualitative study of nursing practice in the context of legislated assisted death. *BMC Nursing, 19*(1), 1–14. https://doi.org/10.1186/s12912-020-0404-5.

Phillippi, J., & Lauderdale, J. (2018). A guide to field notes for qualitative research: Context and conversation. *Qualitative Health Research, 28*(3), 381–388. https://doi.org/10.1177/1049732317697102.

Polit, D. F., & Beck, C. T. (2017). *Nursing research: Generating and assessing evidence for nursing practice* (10th ed.). Philadelphia: Wolters Kluwer.

Renz, S. M., Carrington, J. M., & Badger, T. A. (2018). Two strategies for qualitative content analysis: An intramethod approach to triangulation. *Qualitative Health Research, 28*(5), 824–831. https://doi.org/10.1177/1049732317753586.

Richards, L., & Morse, J. M. (2007). *Read me first for a user's guide to qualitative methods* (2nd ed.). Thousand Oaks, CA: Sage.

Roller, M. R. (2019). A quality approach to qualitative content analysis: Similarities and differences compared to other qualitative methods. *Forum: Qualitative Social Research, 20*(3), 1–21. https://0-doi-org.aupac.lib.athabascau.ca/10.17169/fqs-20.3.3385.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). New York: Oxford University Press.

Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (4th ed.). New York: Teachers College Press.

Smith, J. M., Estefan, A., & Caine, V. (2018). Mothers' experiences of supporting adolescent children through long-term treatment for substance use disorder. *Qualitative Health Research, 28*(4), 511–522. https://doi.org/10.1177/1049732317747554.

Stahlke, S., & Dahlke, S. (2020). The relational role of managers in support of teamwork. *Nursing Research, 33*(1), 112–121.

Stahlke Wall, S. (2018). The impact of regulatory perspectives and practices on professional innovation in nursing. *Nursing Inquiry, 25*(1), e12212. https://doi.org/10.1111/nin.12212.

Streubert, H. J., & Carpenter, D. (2011). *Qualitative research in nursing: Advancing the humanistic imperative* (5th ed.). Philadelphia: Wolters Kluwer.

Sutherland, N., Ward-Griffin, C., McWilliam, C., & Stajduhar, K. (2017). Structural impact on gendered expectations and exemptions for family caregivers in hospice palliative home care. *Nursing Inquiry, 24*(1), e12157. https://doi.org/10.1111/nin.12157.

Thirsk, L. M., Vandall-Walker, V., Rasiah, J., & Keyko, K. (2021). A taxonomy of supports and barriers to family-centered adult critical care: A qualitative descriptive study. *Journal of Family Nursing*. https://doi.org/10.1177/1074840721999372.

Thorne, S. (2020). Beyond theming: Making qualitative studies matter. *Nursing Inquiry, 27*(1), e12343. https://doi.org/10.1111/nin.12343.

Quantitative Data Analysis

Mina D. Singh

LEARNING OUTCOMES

After reading this chapter, you will be able to do the following:

- Differentiate between descriptive and inferential statistics.
- State the purposes of descriptive statistics.
- Identify the levels of measurement in a research study.
- Describe a frequency distribution.
- List measures of central tendency and their use.
- List measures of variability and their use.
- Identify the purpose of inferential statistics.
- Distinguish between a parameter and a statistic.
- Explain the concept of probability as it applies to the analysis of sample data.
- Distinguish between type I and type II errors and their effects on a study's outcome.
- Distinguish between parametric and nonparametric tests.
- List the commonly used statistical tests and their purposes.
- Critically analyze the statistics used in published research studies.

KEY TERMS

alpha
analysis of covariance (ANCOVA)
analysis of variance (ANOVA)
chi-square (χ^2)
confidence interval
correlation
degree of freedom
descriptive statistics
Fisher's exact probability test
frequency distribution
inferential statistics
interval measurement
kurtosis
level of significance (alpha level)
levels of measurement
mean (M)
measurement
measures of central tendency
measures of variability
median
modality
mode
multiple regression
nominal measurement
nonparametric statistics
nonparametric tests of significance
normal curve
null hypothesis
odds ratio
ordinal measurement
p value
parameter
parametric statistics
Pearson correlation coefficient (Pearson r)
percentile
population
post hoc analysis
power

probability
range
ratio measurement
sampling error
scatter plots
scientific hypothesis
semiquartile range (semi-interquartile range)
skew
standard deviation
standard error of the mean
statistic
t statistic
type I error
type II error
Z score

STUDY RESOURCES

Go to Evolve at http://evolve.elsevier.com/Canada/LoBiondo/Research for the Audio Glossary.

STATISTICS ARE USED EXTENSIVELY IN health care research literature. Descriptive and inferential statistics are described in the "Methods" section, the "Results" section, or both sections of a research article.

As a reader, you do not analyze the data yourself, but it is important to understand the researcher's challenge in analyzing the data. After carefully collecting data, the researcher is faced with the task of organizing and analyzing the individual pieces of information so that the meaning of study results is clear. The researcher must choose methods of organizing and analyzing the raw data on the basis of the design, the type of data collected, and the hypothesis or question that was tested. Statistical procedures are used to organize and give meaning to the data.

The "Results" section of a research article contains the data generated from the testing of the hypothesis or research questions. These data are the result of analysis with both *descriptive* and *inferential statistics*. An example of what may be found is as follows: "Overall, the majority (n = 41, 53%) of women in the lanolin group were 'very satisfied' with the effects of lanolin in treating their nipple pain…This difference in maternal satisfaction between the groups was statistically significant," (Jackson and Dennis, 2017: see Appendix B). The data in Table 2 of Appendix B (Jackson and Dennis) are known as descriptive statistics, which are usually the first set of statistical results in a report or published article.

Descriptive statistics are used to summarize and organize data. The techniques used allow researchers to arrange data visually to display meaning and to help in understanding the sample characteristics and variables before the researchers engage in inferential data analyses. In some studies, descriptive statistics may be the only results sought from statistical analysis. Descriptive statistical techniques include measures of central tendency, which describe the average member of a sample, such as mode, median, and mean; measures of variability, such as range and standard deviation (SD); and some correlation techniques, such as a scatter plots, which are a visual representation of the strength and magnitude of the relationship between two variables.

In contrast to descriptive statistics, inferential statistics allow researchers to estimate how reliably they can make predictions and generalize findings on the basis of the data. Inferential statistics are statistical details that combine mathematical processes and logic to test hypotheses about a population with the help of sample data. Through the use of inferential statistics, researchers can draw conclusions that extend beyond the immediate data of the study. An example of inferential statistics is in the study by Havaie et al. (2016): The LPNS [Licensed Practical Nurses] who were likely or very likely to leave had higher emotional exhaustion levels (mean = 33.9,

SD = 11.5) than LPNs who were unlikely or very likely to leave (mean = 27.0, SD = 13.9). This difference was statistically significant [$t(132) = -3.0$, $p < 0.01$] (p. 396).

The purpose of this chapter is to demonstrate how researchers use descriptive and inferential statistics in nursing research studies so that you, as a reader, will be better able to determine the appropriateness of the statistics used and to interpret the strength and quality of the reported findings, their clinical significance, and their applicability to practice. Basic concepts and terminology common in evidence-informed practice publications are presented in Chapter 21. The information in this chapter will help you begin to make sense of the statistics used in research papers.

DESCRIPTIVE STATISTICS

Levels of Measurement

Measurement is the assignment of numbers to variables or events according to rules. Every variable in a research study that is assigned a specific number must be similar to every other variable assigned that number. For example, male participants may be assigned the number 1 and female participants the number 2. The measurement level is determined by the nature of the object or event being measured. Levels of measurement—categorization of the precision with which an event can be measured—from low to high are nominal, ordinal, interval, and ratio (Table 17.1). The levels of measurement help determine the type of statistics to be used in analyzing data. The higher the level of measurement, the greater the flexibility the researcher has in choosing statistical procedures. Every attempt should be made to use the highest level of measurement possible so that the maximum amount of information will be obtained from the data. The Critical Thinking Decision Path illustrates the relationship between levels of measurement and appropriate choice of specific descriptive statistics.

In nominal measurement, variables or events are classified into categories (see Table 17.1). The

TABLE 17.1

SCALES OF MEASUREMENT

MEASUREMENT LEVEL	DESCRIPTION	EXAMPLE
Nominal (may be dichotomous or categorical)	Variables or events are classified into categories; the categories are mutually exclusive, there is no ranking. Dichotomous variables are mutually exclusive and have two true values: e.g., true/false, male/female. Categorical variables are mutually exclusive and have more than two true values: e.g., marital status may be single, married, divorced, separated, widowed.	Gender Hair colour Marital status Religious affiliation
Ordinal	Sorting on relative rankings of variables or events	High school education and less/more than high school education
Ordinal Scale treated as Interval	Rank ordering on an attribute and specifies the difference between the ranks, then a value is assigned to each category	Highly disagree -1, disagree-2, neutral (neither agree or disagree)-3, agree-4, highly agree-5
Interval	Rank ordering on an attribute and specifies the difference between the ranks, assume equivalent distance between the ranks	Body temperature - the distance between 95 C and 100 C is the same as 101 C and 106 C
Ratio	Highest level of measurementAbsolute zero, so can divide, multiply	A person weighing 100 kg is twice as heavy as one who weighs 50 kg

categories are mutually exclusive; a variable or an event either has or does not have the characteristic of a particular category. The numbers assigned to each category are nothing more than labels; such numbers do not indicate more or less of a characteristic. Nominal measurement can be used to categorize a sample with regard to such information as gender, hair colour, marital status, or religious affiliation.

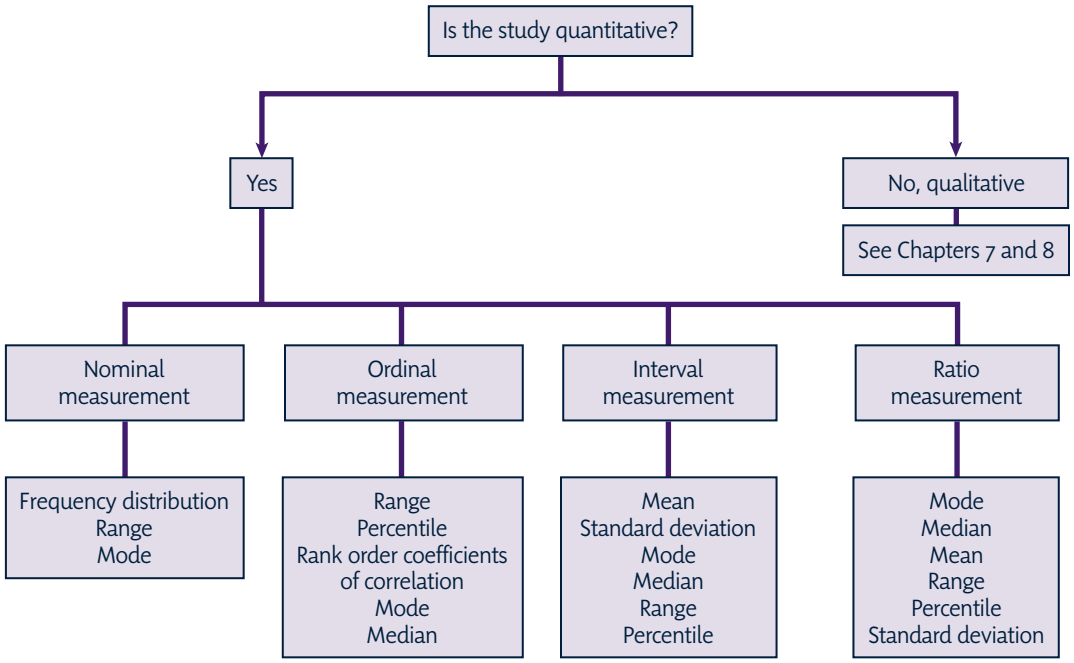
Hisel's (2019) study examining the level of work engagement among four groups of nurses: Veteran-aged, Baby Boomer, Generation X, and Millennial registered nurses—involved nominal measurement. The nominal level of measurement allows the least amount of mathematical manipulation. Most commonly, the frequency of each event is counted, as is the percentage of the total that each category represents.

A variable at the nominal level can also be considered a *dichotomous* or a *categorical* variable. A dichotomous nominal variable has only two true values, such as true/false or gender (male/female) (see Table 17.1). Nominal variables that are categorical still have mutually exclusive categories but have more than two true values, such as marital status (single, married, divorced, separated, or widowed). In both cases, the nominal variables are mutually exclusive. The gender variable of the undergraduate nurses in the study by Goldsworthy et al. (2019) in Appendix D would be considered a dichotomous nominal variable (male/female).

Ordinal measurement reveals relative rankings of variables or events. The numbers assigned to each category can be compared, and the members of a higher ranked category can be said to have more of an attribute than members of a lower ranked category. The intervals between numbers on the scale are not necessarily equal, and zero is not absolute but arbitrary. For example, ordinal

CRITICAL THINKING DECISION PATH

Descriptive Statistics



measurement is used to formulate class rankings, in which one student can be ranked higher or lower than another. However, the actual grade point averages of students may differ widely. Another example is ranking individuals by their level of education, as in the study by Brown et al. (2018) with the variable "age," with categories < 25, 25–35, 36–45, 46–55, and > 55 years of age. Age in categories is an example of an ordinal variable.

The New York Heart Association's classification of cardiac failure adopted by the Canadian Cardiovascular Society (Ezekowitz et al., 2017) consists of four classifications. Classification I represents no symptoms, whereas classification IV represents symptoms at rest with any minimal activity; however, an individual in class IV cannot be said to be four times sicker than an individual in class I. Lauck and colleagues (2016) used this classification in their study to explore factors influencing patients' decision making to undergo TAVI (transcatheter aortic valve implantation) eligibility assessment to inform practice, programme development, health policy, and future research.

With ordinal-level data, the amount of mathematical manipulation possible is limited. In addition to what is possible with nominal-level data, medians, percentiles, and rank-order coefficients of correlation can be calculated (Table 17.2). In most cases, ordinal variables in a scale are treated as interval measurements when converted to numerical codes. For example, when patients are asked to rate their level of satisfaction with life as "not satisfied," "satisfied," or "very satisfied," their responses are an ordinal measurement. When their ratings are treated numerically and coded as 1, 2, and 3, respectively, their ordinal responses are treated as interval measurement. For example, Doktorchik et al. (2018) investigated whether the patterns of change in anxiety and depression during pregnancy can predict preterm birth. State anxiety was measured on the Spielberger State Anxiety Scale, which is a self-report questionnaire including 20 questions on a 4-point scale, with higher numbers corresponding to higher anxiety, the highest score being 80.

In interval measurement, events or variables are ranked on a scale with equal intervals between the numbers. The zero point remains arbitrary and not absolute. For example, interval measurements are used in measuring temperatures on the Fahrenheit scale. The distances between degrees are equal, but the zero point is arbitrary and does not represent the absence of temperature. Test scores also represent interval-level data. The differences between test scores represent equal intervals, but a score of zero does not represent the total absence of knowledge.

In many areas in the social sciences, including nursing, the classification of the level of

TABLE 17.2

LEVEL OF MEASUREMENT SUMMARY TABLE

MEASUREMENT	DESCRIPTION	MEASURES OF CENTRAL TENDENCY	MEASURES OF VARIABILITY
Nominal	Classification	Mode	Modal percentage, range, frequency distribution
Ordinal	Relative rankings	Mode, median	Modal percentage, range, frequency, percentile, semiquartile range, frequency distribution
Interval	Rank ordering with equal intervals	Mode, median, mean	Modal percentage, range, percentile, semiquartile range, standard deviation
Ratio	Rank ordering with equal intervals and absolute zero	Mode, median, mean	All

measurement of intelligence, aptitude, and personality tests is controversial; some researchers regard these measurements as ordinal and others as interval. You need to be aware of this controversy and to examine each study individually in terms of how the data are analyzed. Interval-level data allow more manipulation of data, including the addition and subtraction of numbers and the calculation of means. Because of this additional manipulation, many authorities argue for the higher classification level. The Clinical Self-Efficacy Scale used by Goldsworthy et al. (2019) is an example of ordinal measurements but is used as an interval measurement (see Appendix D).

In ratio measurement, events or variables are ranked on scales with equal intervals and absolute zeros (see Table 17.2). The number represents the actual amount of the property the object possesses. Ratio measurement is the highest level of measurement but is usually achieved only in the physical sciences. Examples of ratio-level data are height, weight, pulse, and blood pressure. All mathematical procedures can be performed with data from ratio scales. Therefore, the use of any statistical procedure is possible as long as it is appropriate for the design of the study.

Research Hint

Descriptive statistics assist in summarizing the data. The descriptive statistics calculated must be appropriate for both the purpose of the study and the level of measurement.

Frequency Distribution

One of the most basic ways of organizing data is in a frequency distribution. In a frequency distribution, the number of times each event occurs is counted, or the data are grouped and the frequency of each group is reported. For example, an instructor reporting the results of an examination could report the number of students receiving each individual grade or could group the grades in ranges and report the number of students who received each group of grades. When reviewing a frequency distribution, symmetry and kurtosis are noted. A distribution can be symmetrical (shaped like a bell) or asymmetrical, where most of the information is to one side, either to the left or the right. Kurtosis is the peakedness of the distribution. Table 17.3 shows the results of an examination given to a class of 51 students. The results are reported in two ways. The columns on the left give the raw data tally and the frequency for each grade, whereas the columns on the right give the grouped data tally and grouped frequencies. In research studies, the results are grouped rather than reported individually for each participant.

When data are grouped, the researcher needs to define the size of the group or the interval width so that no score is categorized into two groups and all groups are mutually exclusive. The groupings of the data in Table 17.3 prevent overlap; each score is categorized into only one group. If the grouping had been 70 to 80 and 80 to 90, scores of 80 would have been categorized into two categories. The grouping should allow for a precise presentation of the data without serious loss of information. Very large interval widths lead to loss of data information and may obscure patterns in the data. If the test scores in Table 17.3 had been grouped as 40 to 69 and 70 to 99, the pattern of the scores would have been obscured.

Information about frequency distributions may be presented in the form of a table, such as Table 17.3, or in the form of a graph. Figure 17.1 illustrates the most common graph forms: the histogram and the frequency polygon. These two methods are similar in that in both, scores or percentages of occurrence are plotted against frequency. The greater the number of points plotted, the smoother is the resulting graph. The shape of the resulting graph allows for observations that further describe the data.

Bar graphs present categorical data (nominal and ordinal variables) with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. The bars do not touch each other;

TABLE 17.3

FREQUENCY DISTRIBUTION

INDIVIDUAL			GROUP		
SCORE	TALLY	FREQUENCY	SCORE	TALLY	FREQUENCY
90	|	1	>89	|	1
88	|	1	80–89		15
86	|	1			
84	|	6			
82	||	2			
80		5			
78		5	70–79	|||	23
76	|	1			
74	||	7			
72	||||	9			
70	|	1			
68	|||	3	60–69		10
66	||	2			
64	||||	4			
62	|	1			
60		0			
58	|	1	<59	||	2
56		0			
54	|	1			
52		0			
50		0			
Total		51	Total		51

Mean, 74.51; standard deviation, +12.1; median, 74; mode, 72; range, 36 (54–90).

this illustrates that the variables are continuous as in a histogram.

Measures of Central Tendency

Measures of central tendency answer questions such as "What does the average nurse think?" and "What is the average temperature of patients on a unit?" These measures yield a single number that describes the middle of the group and summarizes the members of a sample. In statistics, the three measures of central tendency are the mode, the median, and the mean. Depending on the distribution, these measures may not all give the same answer to the question "What is the average?" Each measure of central tendency has a specific use and is most appropriate for specific kinds of measurement and types of distributions. Of the measures of central tendency, the mean is the most stable and the median the most typical. If the distribution of a sample is symmetrical and unimodal, the mean, median, and mode coincide.

Research Hint

Measures of central tendency are descriptive statistics that describe the characteristics of a sample.

MODE. The mode is the most frequent score or result and can be obtained by inspection of the frequency distribution table or graph. Note that a sample distribution can have more than one mode. The number of modes, or peaks, contained in a distribution is called the modality of the distribution. The mode is the type of descriptive statistic most appropriately used with nominal-level data but can be used with all levels of measurement (see Table 17.2). The mode cannot be used for any subsequent calculations and is unstable; in other words, the mode can fluctuate widely from sample to sample from the same population. A change in just one score in Table 17.3 would change the mode from 72.

MEDIAN. The median is the middle score: of the other scores, 50% are higher and 50% are lower. The median is not sensitive to extremes in high and low scores; thus, it is a more accurate estimator of central tendency in non-normal distributions. In the series of scores in Table 17.3, the twenty-sixth score is always the median, regardless of how much the high and low scores change. The median is best used when the data are skewed (see the "Normal Distribution" section) and the

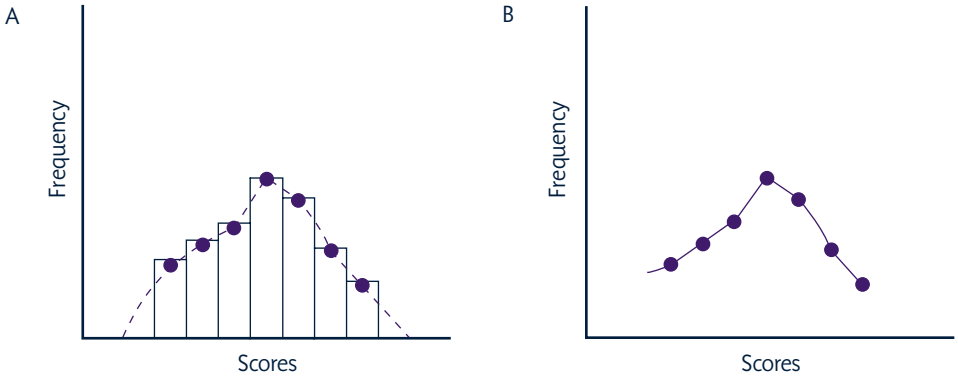


FIG. 17.1 Frequency distributions. **A,** Histogram. **B,** Frequency polygon.

researcher is interested in the "typical" score. For example, if age is a variable, and if a wide range with extreme scores may affect the mean, it would be appropriate to also report the median. The median is easy to find either by inspection or by calculation and can be used with ordinal or higher data, as shown in Table 17.2.

MEAN. The mean (M) is the arithmetical average of all scores and is used with interval- or ratio-level data (see Table 17.2). Most statistical tests of significance refer to the mean, the most widely used measure of central tendency, which is referred to in general conversations as the average. Because the mean is affected by every score, it is affected by extreme scores; however, the larger the sample size, the less effect a single extreme score will have on the mean. For normally distributed populations, the mean is an appropriate measure of central tendency and is generally considered the single best point for summarizing data.

Research Hint

Of the three measures of central tendency, the mean is the most stable, the least affected by extremes, and the most useful for other calculations. The mean can be calculated only with interval- and ratio-level data.

Héon and associates (2016) used a table to describe the sample characteristics in their study that they used to rule out confounding variables. The summary statistics in Appendix D about the sample, comparing the experimental group with the control group, were reported in narrative form; for example, "mothers had an average age of 29.3 years (SD = 5.4; EG: mean = 28.6 ± 5.7; CG; mean = 30.0 ± 5.1)" (p. 575). They did inferential statistics to determine whether there was a difference in ages and found no statistical difference: $p = .419$ (p. 575).

Normal Distribution

The theoretical concept of normal distribution is based on the observation that data from repeated interval or ratio measurements will gather at a midpoint in a distribution, approximating the normal curve illustrated in Figure 17.2. In addition, if the means of a large number of samples of the same interval- or ratio-level data are calculated and plotted on a graph, that curve also approximates the normal curve. This tendency of the means to approximate the normal curve is termed the *sampling distribution of the means*. The mean of the sampling distribution of the means is the mean of the population.

In visual representations of statistics, the normal curve is unimodal and symmetrical about the mean. The mean, median, and mode are equal. An additional characteristic of the normal curve is that a fixed percentage of the scores is located within a given distance of the mean. As shown in Figure 17.2, about 68% of the scores or means

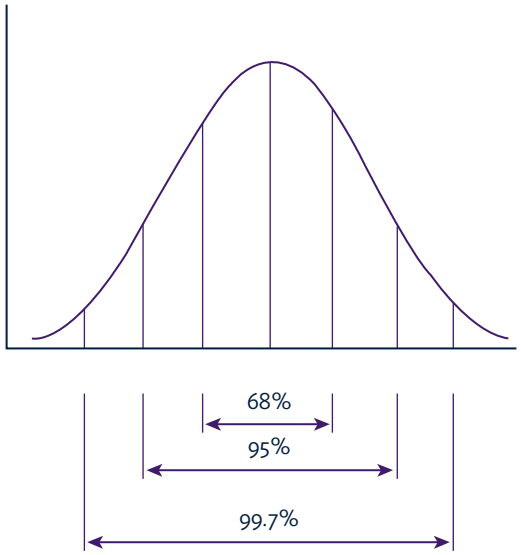


FIG. 17.2 The normal distribution and associated standard deviations.

are within 1 standard deviation of the mean, 95% within 2 standard deviations of the mean, and 99.7% within 3 standard deviations of the mean.

SKEWNESS. Skew is a measure of the asymmetry of a set of scores. Not all samples of data approximate the normal curve. Some samples are nonsymmetrical, and the peak is off centre. For example, worldwide individual income has a positive skew: Most individuals have incomes in the low-to-moderate range and few in the upper range. In a positive skew, the peak of the distribution curve would be to the left of a normal curve, and the mean is to the right of the median. In contrast, age at death in Canada has a negative skew because most deaths occur at older ages. In a negative skew, the peak of the distribution curve would be to the right of a normal curve, and the mean is to the left of the median. Figure 17.3 illustrates positive and negative skew. In each diagram, the peak is off centre, and one "tail" of the curve is longer.

If the distribution is skewed, the mean will be pulled in the direction of the long tail of the distribution. With a skewed distribution, all three statistics should be reported.

Evidence-Informed Practice Tip

The descriptive statistics for a sample indicate whether the sample data are skewed.

Interpreting Measures of Variability

Variability or dispersion is concerned with the spread of data. Measures of variability—statistical procedures that describe the level of dispersion in sample data—answer questions such as "Is the sample homogeneous or heterogeneous?" and "Is the sample similar or different?" If a researcher measures oral temperatures in two samples, one sample drawn from a healthy population and one sample from a hospitalized population, it is possible that the two samples will have the same mean. However, a wider range of temperatures is

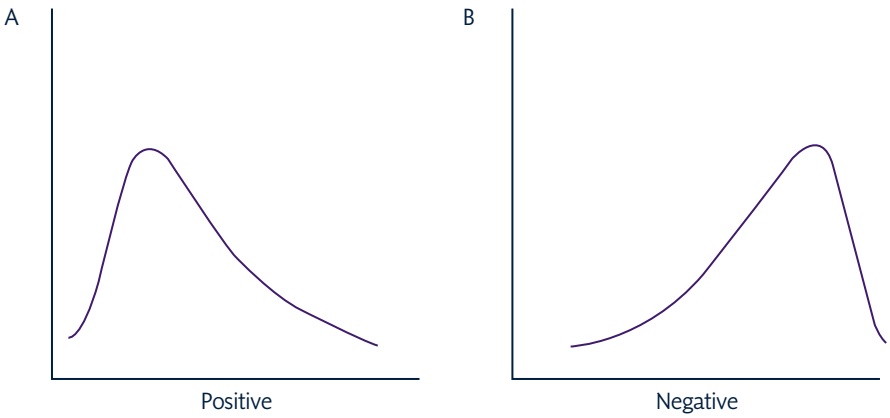


FIG. 17.3 Positive and negative skew. A, Positive skew. B, Negative skew.

more likely to be found in the hospitalized sample than in the healthy sample. Measures of variability are used to describe these differences in the dispersion of data. As with measures of central tendency, the various measures of variability are appropriate to specific kinds of measurement and types of distributions.

Research Hint

Descriptive statistics related to variability enable you to evaluate the homogeneity or heterogeneity of a sample.

RANGE. The range is the simplest but most unstable measure of variability. Range is the distance between the highest and lowest scores. A change in either of these two scores would change the range. The range should always be reported with other measures of variability. For example, Castonguay et al. (2017) tested body-related shame and guilt as predictors of breast cancer survivors' (BCS') moderate to vigorous intensity physical activity (MVPA) during six months. They found that there was quite a spread in of ages in their sample, from 30 years old to 79 years old (Table 17.4). Range affects the standard deviation, as discussed later. The range in Table 17.4 could easily change with an increase or decrease in the high scores or the low scores with a different sample.

SEMIQUARTILE RANGE. The semiquartile range (semi-interquartile range) is the range of the middle 50% of the scores. It is more stable than the overall range because it is less likely to be changed by a single extreme score. The semiquartile range lies between the upper and lower quartiles; the upper quartile consists of the top 25% of scores, and the lower quartile consists of the lowest 25% of the scores. In Table 17.3, the middle 50% of the scores are between 68 and 78, and the semiquartile range is 10.

PERCENTILE. A percentile represents the percentage of scores that a given score exceeds. The median

TABLE 17.4

DESCRIPTIVE STATISTICS AND SCORE RANGES OF STUDY VARIABLES (N = 149)

VARIABLE	X-BAR	SD	RANGE
Age (years)	55.34	10.5	30–79
Body mass index (kg/m-sq)	25.86	5.33	18–43
Depression	1.48	0.7	0–3
Months since diagnosis	16.46	9.46	2–26
Months since treatment	9.48	8.41	0–13
Body-related shame	2.64	0.91	1–5
External regulation	0.39	0.58	0–4
Introjected regulation	1.17	1.07	0–4
Autonomous regulation	2.61	0.98	0–4
Identified regulation	2.62	0.85	0–4
Intrinsic regulation	2.54	1.07	0–4
MVPA at T1	1.08	1.51	0–7
MVPA at T2	0.81	1.22	0–5
Variable		n	%
Current smoker		9	6
Education			
Some high school		7	5
High school diploma		20	13
Some college		12	8
College or technical diploma		29	20
Undergraduate degree		41	28
Graduate degree		40	27
Stage of cancer			
I		62	42
II		63	42
III		24	16

MVPA - moderate to vigorous physical activity; T1- time 1 (baseline); T2 - time 2 (six months later)

Adapted from Castonguay, A., Wrosch, C., Pila, E., & Sabiston, C. (2017). Body-related shame and guilt predict physical activity in breast cancer survivors over time. *Oncology Nursing Forum*, *44*(4), 465–475. https://doi.org/10.1188/17.ONF.465-475.

is the fiftieth percentile, and in Table 17.3, it is a score of 74. A score in the ninetieth percentile is exceeded by only 10% of the scores. The zero percentile and the hundredth percentile are usually not used.

STANDARD DEVIATION. The standard deviation is the most frequently used measure of variability and is based on the concept of the normal curve (see Figure 17.2). The standard deviation is a measure of average deviation of the scores from the mean and, as such, should always be reported with the mean. The standard deviation accounts for all scores and can be used to interpret individual scores. For the examination in Table 17.3, the mean was 74.51 and the standard deviation was 12.1; thus, a student should know that 68% of the grades were between 86.61 and 62.41. If the student received a grade of 88, he or she would know that this grade was better than those of most of the class, whereas a grade of 58 would indicate that the student did not do as well as most of the class. Table 2 in Appendix D from the study by Goldsworthy and colleagues (2019) reports the mean and standard deviation of the study variables' on the Clinical Self-Efficacy Scale within- and between-group pre-post intervention in the treatment and control groups. As illustrated in this table, the mean difference between the pre and post scores, for the treatment group on "recognizing a patient with no pulse" was 8.04 (SD = 12.41), whereas the mean difference score for the control group was 2.45 (SD = 10.44). This means that 68% of the of the change in the treatment group was between a score of 3.37 and 20.45 on self-efficacy and 68% of the control group was between 7.99 and 12.89. This table allows the reader to inspect the data and see the variation in the data, and these data shows that the self-efficacy of treatment group after the intervention. In assessing SD, it is also important to note the size of the SD, a small standard deviation indicates that the data points tend to be very close to the mean; a large standard deviation indicates that the data points are spread out over a large range of values. If two samples had a mean equal to 2, but one had a SD of 8.0 and the other had a SD of 3.0, the first sample would be more heterogeneous.

The standard deviation is used in the calculation of many inferential statistics. One limitation of the standard deviation is that it is expressed in terms of the units used in the measurement and cannot be used to compare means that have different units. If researchers were interested in the relationship between height measured in centimetres and weight measured in kilograms, it would be necessary to convert the height and weight measurements to standard units, or *Z* scores. The *Z* score is used to compare measurements in standard units. Each of the scores is converted to a *Z* score, and then the *Z* scores are used to examine the relative distance of the scores from the mean. A *Z* score of 1.5 means that the observation is 1.5 standard deviations above the mean, whereas a score of −2 means that the observation is 2 standard deviations below the mean. By using *Z* scores, a researcher can compare results from scales that use different measurement units, such as height and weight.

Research Hint

Many measures of variability exist. The standard deviation is the most stable and useful because it provides a visual image of how the scores are dispersed around the mean.

INFERENTIAL STATISTICS

Inferential statistics combine mathematical processes with logic and allow researchers to test hypotheses about a population by using data obtained from probability samples. Statistical inference is generally used for two purposes: to estimate the probability that statistics found in the sample accurately reflect the population parameter and to test hypotheses about a population.

In the first purpose, a parameter is a characteristic of a population—a well-defined set that has

certain specified properties—whereas a statistic is a characteristic of a sample. Statistics are used to estimate population parameters. Suppose that a researcher randomly selects 100 people with chronic lung disease and uses an interval-level scale to study their knowledge of the disease. A mean score of 65 for these participants represents the sample statistic. If the researcher were able to study every participant with chronic lung disease, he or she also could calculate an average knowledge score, and that score would be the parameter for the population. Researchers are rarely able to study an entire population, but inferential statistics provide evidence that allow them to make statements about the larger population from studying the sample.

Both parametric and nonparametric inferential tests can be used in data analyses (Tables 17.5 and 17.6). Parametric statistical models are based on assumptions about the distributions of sample values and parameters; thus, in these models, means and variances are used to test significance. Nonparametric tests are used when populations have non-normal distributions or when

TABLE 17.5

TESTS OF DIFFERENCES BETWEEN MEANS

LEVEL OF MEASUREMENT	ONE GROUP	TWO GROUPS		MORE THAN TWO GROUPS
		RELATED	INDEPENDENT	
NONPARAMETRIC				
Nominal	Chi-square	Chi-square Fisher's exact probability test	Chi-square	Chi-square
Ordinal	Kolmogorov-Smirnov test	Sign test Wilcoxon matched-pairs test	Chi-square	Chi-square
PARAMETRIC				
Interval or ratio	Correlated t test ANOVA (repeated measures)	Correlated *t* test	Independent *t* test	ANOVA
			ANOVA	ANCOVA MANOVA

ANCOVA, analysis of covariance; *ANOVA*, analysis of variance; *MANOVA*, multivariate analysis of variance.

TABLE 17.6

TESTS OF ASSOCIATION

LEVEL OF MEASUREMENT	TWO VARIABLES	MORE THAN TWO VARIABLES
NONPARAMETRIC		
Nominal	Phi coefficient Point-biserial correlation	Contingency coefficient
Ordinal	Kendall's tau Spearman's rho	Discriminant function analysis
PARAMETRIC		
Interval or ratio	Pearson *r*	Multiple regression Path analysis Canonical correlation

researchers wish to explore associations among variables. In these tests, no assumptions about the distribution of the data are made.

The example of the study of patients with lung disease alludes to two important qualifications of how a study must be conducted so that inferential statistics may be used. First, the sample was selected randomly—that is, through the use of probability methods (see Chapter 13). Because you are already familiar with the advantages of probability sampling, you know that in order to make generalizations about a population from a sample, that sample must be representative. All procedures for inferential statistics are based on the assumption that the sample was drawn with a known probability. Second, the scale had to reflect the interval level of measurement. The mathematical operations involved in inferential statistics require this level of measurement. Note that researchers who use nonprobability methods of sampling also use inferential statistics. To compensate for the use of nonprobability sampling methods, researchers use techniques such as sample size estimation through power analysis. The following two Critical Thinking Decision Paths provide algorithms that reflect inferential statistics and that researchers use for statistical decision making.

Evidence-Informed Practice Tip

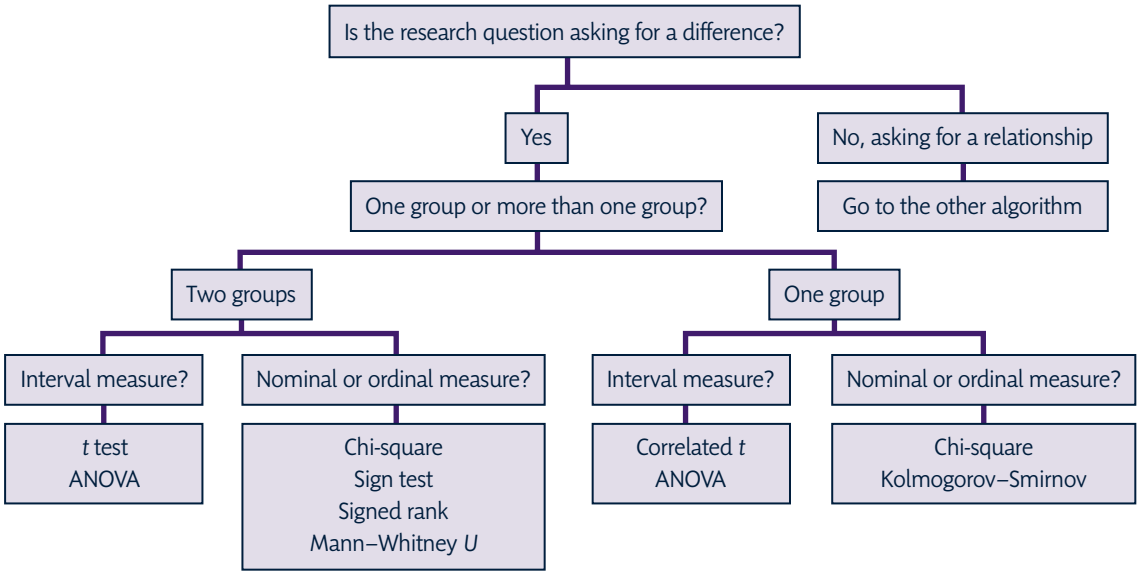
Try to determine whether the statistical test chosen was appropriate for the design, the type of data collected, and the level of measurement.

Hypothesis Testing

The second and most commonly used purpose of inferential statistics is hypothesis testing. Statistical hypothesis testing allows researchers to make objective decisions about the outcome of their study and to answer questions such as "How much of this effect is a result of chance?"; "How strongly are these two variables associated with each other?"; and "What is the effect of the intervention?"

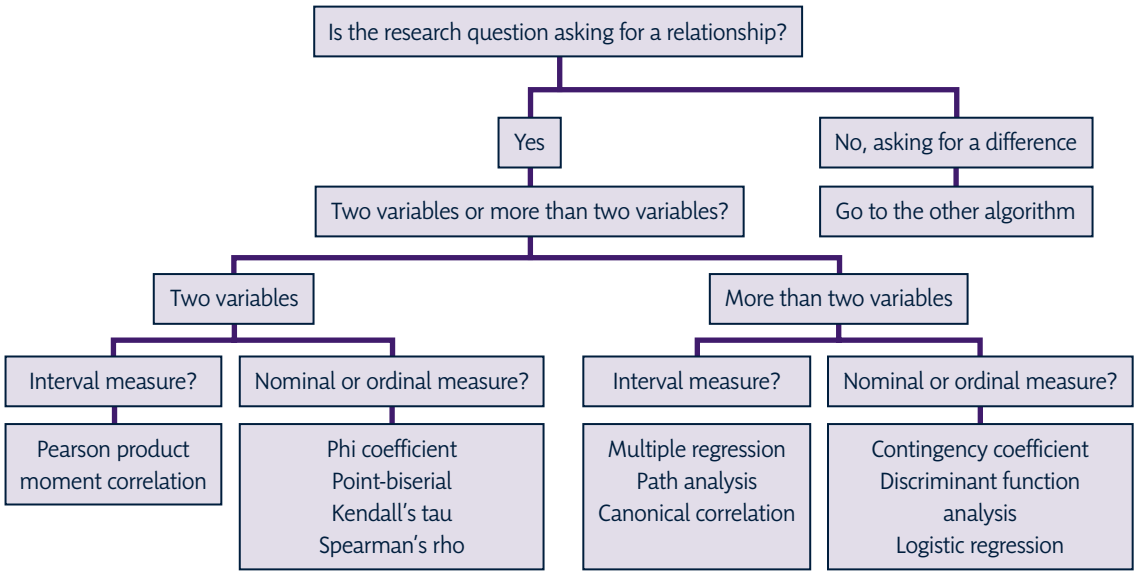
CRITICAL THINKING DECISION PATH

Inferential Statistics—Difference Questions



CRITICAL THINKING DECISION PATH

Inferential Statistics—Relationship Questions



The procedures used to make inferences are based on principles of negative inference. For example, to study the effect of a new educational program for patients with chronic lung disease, the researcher would actually have two hypotheses: the scientific hypothesis and the null hypothesis. The research or scientific hypothesis (H1) is what the researcher believes the outcome of the study will be. In this example, the scientific hypothesis would be that the educational intervention would have a marked effect on the outcome in the experimental group in comparison with that in the control group. The null hypothesis (also called the *statistical hypothesis* or H_0), which is the hypothesis that actually can be tested by statistical methods, would be that no difference exists between the groups. In inferential statistics, the null hypothesis is used to test the validity of a scientific hypothesis in sample data. According to the null hypothesis, no relationship exists between the variables, and any observed relationship or difference is merely a function of chance fluctuations in sampling.

The concept of the null hypothesis is often confusing. An example may help clarify this concept. Sherrard, Duchesne, Wells, and colleagues (2015) used an interactive voice response to follow patients with acute coronary syndrome best practice guidelines and compared this to the usual care offered by the best practice guidelines. On the basis of this hypothesis, Sherrard and colleagues wanted to determine whether the differences found in the dependent variables of medication adherence and adverse effects differed significantly between the intervention group and the control group. The authors had to use the null hypothesis—that no difference would exist between the intervention and control groups—to test the scientific hypothesis. They found a significant improvement in medication adherence and a decrease in unplanned medical visits by the group that received interactive voice response. In other words, the differences between the control and intervention group scores were large enough to conclude that they were unlikely to be caused by chance. Thus, the null hypothesis was rejected.

In another example, Bilik et al. (2020) examined the effects of web-based concept mapping education on nursing students' concept mapping and critical thinking skills. Their research hypotheses were: 1) there is a difference in mean scores for concept evaluation keys between students receiving web-based concept mapping education and those not receiving this education, and 2) there is a difference in mean scores on the Critical Thinking Motivational Scale (CTMS) between students receiving web-based concept mapping education and those receiving this education (p. 2). They reported significant differences between the groups ($p = 0.00$) for hypothesis 1, thus the null hypothesis was rejected. In regards to hypothesis 2, the experimental and control groups differed significantly in their scores for the subscales of CTMS expectancy ($p = 0.037$), attainment ($p = 0.015$), and utility ($p = 0.015$), but they did not differ significantly in terms of the subscales value ($p = 0.225$) and cost ($p = 0.070$) (p. 4); thus, hypothesis 2 was partially met. See information on the interpretation of p values in the "Level of Significance" section.

All statistical hypothesis testing is a process of disproof or rejection. It is impossible to prove that a scientific hypothesis is true, but it is possible to demonstrate that the null hypothesis has a high probability of being incorrect. To reject the null hypothesis, therefore, is to show support for the scientific hypothesis, which is the desired outcome of most reports of inferential statistics.

Research Hint

Remember that most samples used in clinical research are samples of convenience, but most researchers use inferential statistics. Although such use violates one of the assumptions of such tests, the tests are robust enough to not seriously affect the results unless the data are skewed in unknown ways.

Probability

The researcher can never *prove* the scientific hypothesis but can show support for it by rejecting the null hypothesis—that is, by showing that the null hypothesis has a high probability of being incorrect. The theory underlying all of the procedures discussed in this chapter is probability theory. Probability is a concept that people talk about all the time, such as the chance of rain, but have a difficult time defining. The probability of an event is the event's long-run relative frequency in repeated trials under similar conditions. In other words, the statistician does not think of the probability of obtaining a single result from a single study but rather of the chances of obtaining the same result from an idealized study that can be carried out many times under identical conditions. The notion of repeated trials allows researchers to use probability to test hypotheses.

Statistical probability is based on the concept of sampling error. The use of inferential statistics is based on random sampling. However, even when samples are randomly selected, the possibility of errors in sampling always exists. Therefore, the characteristics of any given sample may be different from those of the entire population.

Suppose that a large group of patients with decubitus ulcers is available for study and that researchers wish to learn the average length of time for such ulcers to heal with the usual nursing care. If the researchers studied the entire population, they might obtain an average healing time of 50 days, with a standard deviation of 10 days. Now, suppose that the researchers did not have the money necessary to study all the patients but wished to conduct several consecutive studies of this condition. For this study, the researchers would first select a sample of 25 patients, calculate the mean and standard deviation, and then select the next sample. If this process is repeated many times in different samples, a different mean for each sample would probably result. For example, the researchers might find that one sample's mean might be 50.5 days, the next 47.5, and the next 62.5. The tendency for statistics to fluctuate from one sample to another is known as sampling error.

Sampling distributions are theoretical. In practice, researchers do not routinely draw consecutive samples from the same population; they usually compute statistics and make inferences on the basis of data from one sample. However, the knowledge of the properties of the sampling distribution—if these repeated samples are hypothetically obtained—enables the researcher to draw a conclusion on the basis of data from one sample. Such a conclusion is possible because the sampling distribution of the means has certain known properties.

The sampling distribution of the means is shaped like a normal curve, and the mean of the sampling distribution is the mean of the population. As discussed in the earlier "Normal Distribution" section, because the sampling distribution of the means is normal, several other important characteristics are revealed. When scores are normally distributed, 68% of them are between +1 standard deviation and −1 standard deviation, or the probability is 68 per 100 that any one randomly drawn sample mean is within the range of values between +1 standard deviation and −1 standard deviation (see Figure 17.2). In the example described earlier, if only one sample were selected, the chance of finding a sample mean between 40 and 60 would be 68%. The standard deviation of a theoretical distribution of sample means is called the standard error of the mean. The word *error* is used because the various means that make up the distribution contain an error in their estimates of the population mean. The error is considered to be standard because it implies the magnitude of the average error, just as a standard deviation implies the average variation from one mean. The *smaller* the standard error, the *less* variable are the sample means and the *more accurate* are those means as estimates of the population value.

Although researchers rarely construct sampling distributions, standard error can be estimated because it bears a systematic relationship to the sample standard deviation and the size of the sample. Thus, increasing the size of the sample will increase the accuracy of estimates of population parameters. It is intuitive that an increase in the size of a sample will decrease the likelihood that one outlying score will dramatically affect the sample mean (see Chapter 13). The other reason that the sampling distribution is so important is that all statistics have sampling distributions. Researchers consult these distributions when making determinations about rejecting the null hypothesis.

Evidence-Informed Practice Tip

Remember that the strength and quality of evidence are enhanced by repeated trials that have consistent findings, thereby increasing the generalizability of the findings and applicability to clinical practice.

Type I and Type II Errors

The researcher's decision to accept or fail to accept (reject) the null hypothesis is based on a consideration of the probability that the observed differences are a result of chance alone. Because data on the entire population are not available, the researcher cannot flatly assert that the null hypothesis is or is not true. Thus, statistical inference is always based on incomplete information about a population, and errors can occur when such inferences are made. These errors are classified as type I and type II.

A type I error is the researcher's incorrect decision to reject the null hypothesis (Kline, 2005); that is, the researcher has found that results are statistically significant, but in fact they are not, and has accepted the alternate hypothesis. If, however, the researcher had found that the groups did not differ perhaps because only a few patients had been studied or the design of the study was poor for determining differences, a type II error might occur. In a type II error—also known as beta (β)—the results from the sample data lead to the failure to reject the null hypothesis when it is actually false; that is, no statistically significant differences between groups were found but there

are indeed real differences. Power is the conditional prior probability that the researcher will decide correctly to reject the null hypothesis when it is actually false (Kline, 2005). A standard value of power of .8 is used to conduct power analyses in studies to determine sample size before the study begins; this means that the researcher is accepting 20% risk of Type II error. Power and beta are complementary and sum to 1.00. When power is increased, type II error is decreased, and vice versa.

In Campbell-Yeo, Johnston, Joseph, and colleagues' (2015) study on cobedding and recovery time after heel lance in preterm twins, one null hypothesis of the study was that there would be no differences in pain response and time to return to physiological measures between the experimental and control groups. Campbell-Yeo and colleagues reported a significant difference in recovery time; that is, the time was shorter in the cobedding group, mean = 75.6 seconds (SD, 70.0), compared with the usual care group, mean = 142.1 seconds (SD, 138.1, $p = .001$). If the differences found were truly a function of chance (because this group of participants was unusual in some way) and if the number of participants was too small, a type II error would occur. Thus, the simplest way of reducing Type II error is to increase the sample size.

The relationship of the two types of errors is shown in Figure 17.4. When you critique a study to determine whether a type I error has occurred (rejecting the null hypothesis when it is actually true), you should consider the reliability and validity of the instruments used. For example, if the instruments did not accurately and precisely measure the intervention variables, the conclusion could be that the intervention made a difference, but, in reality, it did not. It is critical to consider the reliability and validity of all of the measurement instruments reported (see Chapter 15). In a practice discipline, type I errors usually are considered more serious because if a researcher declares that differences exist where none are present, then patient care can potentially be affected adversely. Type II errors (accepting the null hypothesis when it is false) may occur if the sample in the study is too small, thereby limiting the opportunity to measure the *treatment effect*, a true difference between two groups. A larger sample size improves the ability to *detect the treatment effect*—that is, the differences between two groups. If no significant difference is found between two groups with a large sample, this finding provides stronger evidence (than with a small sample) not to reject the null hypothesis.

Level of Significance

The researcher does not know when an error in statistical decision making has occurred. It is possible to know only that the null hypothesis is indeed true or false if data from the total population are available. However, the researcher can control the risk of making type I errors by setting the level of significance before the study begins (a priori). The level of significance (alpha level) is the probability of making a type I error—in other words, the conditional probability of rejecting the

	REALITY	
Conclusion of test of significance	Null hypothesis is true	Null hypothesis is not true
Not statistically significant	Correct conclusion	Type II error
Statistically significant	Type I error	Correct conclusion

FIG. 17.4 Outcome of statistical decision making.

null hypothesis when it is actually true. Alpha, or the level of significance, is considered an a priori probability because it is set before the data are collected, and it is a conditional probability because the null hypothesis is assumed to be true. The minimum level of significance acceptable for nursing research is .05. If the researcher sets alpha at .05, the researcher is willing to accept the fact that if the study were done 100 times, the decision to reject the null hypothesis would be wrong in 5 of those 100 trials, only if the null hypothesis is true.

Sometimes the researcher wants to have a smaller risk of rejecting a true null hypothesis; in that case, the level of significance may be set at .01. In this case, the researcher is willing to make the wrong decision only once in 100 trials. The decision as to how strictly the alpha level should be set depends on how important it is not to make an error. For example, if the results of a study are to be used to determine whether a great deal of money should be spent in an area of nursing care, the researcher may decide that the accuracy of the results is so important that an alpha level of .01 is chosen. In most studies, however, alpha is set at .05.

Another concept, the *p* value, is needed to interpret the alpha value. The *p* value, or probability value, is the probability of obtaining, from the study data, a test statistic, such as the mean, a result equal to or "more extreme" than what was actually observed, when the null hypothesis is true. The *p* value is different from alpha because it is calculated from the sample data and is considered the *exact level of significance*. Thus, if this exact level of significance is less than the conditional a priori probability of making a type I error ($p <$ alpha), then the null hypothesis is rejected, and the result is considered statistically significant at that alpha level. For example, if the alpha is set at .05 and the *p* value is found to be .04, then the results are considered statistically significant.

Whatever level of significance is set, the researcher either rejects or accepts the null hypothesis when comparing the statistical results with the preset alpha. For example, in Ingram et al.'s (2016) study, the hypothetical null hypothesis regarding no change in level of knowledge about care of delirious clients after participating in an education session was rejected, as the researchers found statistical significance, because the variables of the hypothesis were significant at the .05 level or lower; in other words, the *p* values were less than alpha. Jackson and Dennis (2017), however, failed to reject the null hypothesis and found that pain scores were not significantly different between the experimental and control groups (Appendix B).

Perhaps you are thinking that researchers should always use the lowest alpha level possible because it makes sense that they would like to keep the risk of both types of errors at a minimum. Unfortunately, decreasing the risk of making a type I error increases the risk of making a type II error; that is, the stricter the researcher is in preventing the rejection of a true null hypothesis, the more likely the researcher is to accept a false null hypothesis. Therefore, researchers always have to accept more of a risk of one type of error when setting the alpha level.

Another method of determining the level of significance and whether to accept or reject the null hypothesis is called the *critical values method*. In this method, by calculating the estimates of population mean and standard deviation, a range of values is determined from which the researcher can compare the sample mean findings and decide whether to reject the null hypothesis.

Suppose researchers want to know the importance of support groups for caregivers of older adults. They ask 100 caregivers to rate the importance of support groups to them by using an instrument that ranges from 0 (not important at all) to 100 (very important). If Figure 17.2 represents the theoretical distribution for this study (a normal distribution with a mean of 50), 68% of the population would score between 40 and 60, and 95% would score between 30 and 70. Thus,

the null hypothesis would be that the mean score for the population of caregivers would be 50, and the scientific hypothesis would be greater or less than 50. After measurements with this sample are completed, the researchers find that the sample mean score is 75. This mean is consistent with the scientific hypothesis, and the researchers can be 95% sure that, most of the time, the sample mean score would fall under this cut-off; thus, they would have confidence in rejecting the null hypothesis. In other words, only 5 of 100 times would they obtain this result by chance alone.

Research Hint

Decreasing the alpha level acceptable for a study increases the chance that a type II error will occur. When a researcher is conducting many statistical tests, the probability that some of the test results will be significant increases as the number of tests increases. Therefore, when a large number of tests are being conducted, many researchers decrease the alpha level to .01.

Practical and Statistical Significance

Statistical significance and practical significance are not the same. When a researcher finds a hypothesis statistically significant, this finding is unlikely to have happened by chance. In other words, if the level of significance has been set at .05, the odds are 95% that the researcher will make the correct conclusion on the basis of the results of the statistical test performed on sample data. The researcher would reach the wrong conclusion only 5 times in 100.

Suppose that a researcher is interested in the effect of loud rock music on the behaviour of laboratory mice. The researcher could design an experiment to study this question and find that loud music makes the mice act strangely. A statistical test suggests that this finding is not the result of chance. However, such a finding may or may not have practical significance, even though the finding has statistical significance. Whereas some authorities would argue that this study might have relevance to understanding the behaviour of teenagers, others would argue that the study has no practical value. Thus, the findings of a study may have statistical significance, but they may have no practical value or significance.

Although researchers should consider the practicality of a problem in the early stages of a research project (see Chapter 3), a distinction between the statistical and practical significance of the findings also should be made in the discussion of the results of a study. Some authorities believe that if the findings are not statistically significant, they have no practical value. In Jackson and Dennis' (2017) study, in Appendix D, the research hypothesis was not statistically supported, but nonsupported hypotheses provide as much information about the intervention as do the supported hypotheses. They found that application of lanolin to painful/damaged nipples in the immediate postpartum period does not significantly decrease nipple pain or improve breastfeeding outcomes when compared with usual care. The data allowed the researchers to return to the previous literature in the area and discern from those findings both statistical and practical significance.

Evidence-Informed Practice Tip

You study the results to determine the effectiveness of the new treatment and the size and clinical importance of the effect.

Tests of Statistical Significance

Tests of statistical significance may be parametric or nonparametric. In most studies in nursing research literature, investigators use parametric tests that have the following three attributes:

1. The estimation of at least one population parameter
2. Measurement at the interval level or higher
3. Assumptions about the variables being studied

One assumption is usually that the variable is normally distributed in the overall population.

In contrast to parametric tests, nonparametric tests of significance are not based on the estimation of population parameters, so their assumptions about the underlying distribution are less

restrictive. Nonparametric tests are usually applied when the variables have been measured on a nominal or ordinal scale.

Some debate surrounds the relative merits of the two types of statistical tests. The moderate position taken by most researchers and statisticians is that nonparametric statistics—also called *distribution-free tests*—are best used when the data cannot be assumed to be at the interval level of measurement or when the sample is small and the normality of the underlying distribution cannot be inferred. If these assumptions can be made, however, most researchers prefer to use parametric statistics, which are more powerful and more flexible than nonparametric statistics. Because stringent assumptions for parametric tests makes them more powerful than nonparametric tests, researchers are able to formulate simple sample statistics, such as the mean and the standard deviation, which enables them to accurately estimate population parameters with standard sampling distributions to obtain probabilities regarding the null hypotheses.

Researchers use many different statistical tests of significance to test hypotheses; however, the procedure and the rationale for their use are similar from test to test. Once the researcher has chosen a significance level and collected the data, the data are used to compute the appropriate test statistic. Each test has a related theoretical distribution that shows the probable and improbable values for that statistic. On the basis of the statistical result and the values in the distribution, the researcher either accepts or rejects the null hypothesis and then reports both the statistical result and its probability. Thus, a researcher may perform a t test, obtain a value of 8.98, and report that it is statistically significant at the $p < .05$ level. This means that in 100 tests, the researcher had five chances to conclude wrongly that this result could not have been obtained by chance.

The likelihood of finding a statistic that is high enough to be statistically significant is increased as the sample size increases. This likelihood is indicated by the degrees of freedom, which are often reported with the statistic and the probability value. Usually abbreviated as *df*, the degree of freedom is the freedom of a score's value to vary depending on the other scores and the sum of these scores; thus, $df = N - 1$. For example, imagine you have four numbers represented by letters (*a, b, c,* and *d*) that must add up to a total of *x*; you are free to randomly choose the first three numbers, but the fourth must be chosen to make the total equal to *x*, and thus your degree of freedom is 3.

To make statistical inferences from data, many types of tests can be conducted. Tables 17.5 and 17.6 list the tests most commonly used for inferential statistics. The test used depends on the level of the measurement of the variables in question and the type of hypothesis being studied. These statistics test two types of hypotheses: that difference exists between groups (see Table 17.5) and that a relationship exists between two or more variables (see Table 17.6). In addition, many types of regression analyses are available to predict the dependent variable. Simple regression analyses (one independent variable) and multiple regression analyses (several independent variables) are used when the dependent variable is at the interval level or higher.

Research Hint

The use of nonparametric statistics in a study does not mean that the study is useless. The use of nonparametric statistics is appropriate when measurements are not made at the interval level or the variable under study is not normally distributed.

Evidence-Informed Practice Tip

Try to discern whether the test for analyzing the data was chosen because it gave a significant p value. A statistical test should be chosen on the basis of its appropriateness for the type of data collected, not because it gives the answer that the researcher hoped to obtain.

Tests of Differences

The type of test used for any particular study depends primarily on whether the researcher

examines differences in one, two, or three or more groups and whether the data to be analyzed are nominal, ordinal, or interval (see Table 17.5). Suppose that a researcher constructs an experimental study with an after-only design (see Chapter 11). What the researcher hopes to determine is that the two randomly assigned groups are different after the introduction of the experimental treatment. If the measurements taken are at the interval level, the researcher would use the t test to analyze the data. If the t statistic was found to be high enough to be unlikely to have occurred by chance, the researcher would reject the null hypothesis and conclude that the two groups were indeed more different than would have been expected on the basis of chance alone. In other words, the researcher would conclude that the experimental treatment had the desired effect.

Osahor et al.'s (2019) study on the relationship between math personality, math anxiety, test preparation strategy, and medication dose calculations in first-year nursing students illustrated the use of the t statistic. In this study, the t test was used to determine differences in math anxiety between collaborative students and students in the compressed program. The results showed that students in the collaborative program were more anxious (t (161) = −2.67, p = .008) and used more test preparation strategies (t (161) = −2.67, p = .008) than students in the compressed program.

Evidence-Informed Practice Tip

Tests of difference are most commonly used in experimental and quasiexperimental designs that provide level II and level III evidence.

PARAMETRIC TESTS. The t statistic is commonly used in nursing research. This statistic reflects whether two group means are different. Thus, the t statistic is used when the researcher has two groups, and the question is whether the mean scores on some measure are more different than would be expected by chance. To use this test, the variables must have been measured at the interval or ratio level, and the two groups must be independent, meaning that nothing in one group helps determine what is in the other group. If the groups are related in some way, as when samples are matched (see Chapter 13), and the researcher also wants to determine differences between the two groups, a paired, or correlated, t test would be used.

The t statistic illustrates one of the major purposes of research in nursing: to demonstrate that differences exist between groups. Groups may be naturally occurring collections, such as age groups, or they may be experimentally created, such as treatment and control groups. Sometimes a study has more than two groups, or measurements are taken more than once.

ANALYSIS OF VARIANCE (ANOVA). Analysis of variance (ANOVA) is a test similar to the t test, but the procedure is testing for differences when there are three or more groups. For example, Subedi et al. (2019) used five levels of education—illiterate, literate, primary, secondary, and college—to determine how general well-being is affected by Bhutanese refugees. They found a statistical difference (F (4,103) = 3.02, p = .02). Tukey's honestly significant difference post hoc test indicated that there was a minimal significant difference between "college" and "literate" (p = .05) and "college" and "illiterate" (p = .05) (p. 172). These researchers used analysis of variance (ANOVA), because there were five groups. Like the t statistic, the ANOVA statistic is used to test whether group means differ, but instead of testing each pair of means separately, ANOVA accounts for the variation between groups and within groups. The ANOVA is usually performed with two or more groups by an F test rather than multiple pairs of t tests (see Practical Application box). If multiple pairs of t tests are done, the type I error rate would increase.

There are many ANOVA tests; for example, a one-way ANOVA, as in the Subedi et al. (2019) study, is used to test the relationship between one

categorical independent variable (levels f education) and one continuous variable (well-being). A two-way ANOVA is used to test the relationship between two categorical independent variables, each with more than one level such as gender (male and female) and level of education (illiterate, literate, primary, secondary, and college) and one continuous variable (well-being).

Practical Application

Dahlke et al. (2019) conducted a study to explore student nurses' perceptions about older people. Nursing students' perceptions about working with older people was measured by Burbank's Perceptions of Caring for Older People's scale. There was statistical difference at the $p = 0.05$ level in the Burbank scores for the six different clinical rotation groups (F (5,364) = 2.6, $p = 0.024$.

In another example, Hroch et al. (2019) examined preregistration nursing students' knowledge and attitudes about the assessment and management of pain in four education sites from two post-secondary institutions. They used the Knowledge and Attitudes Survey Regarding Pain (KASRP) instrument. One of their analyses consisted of examining institution and program on KASRP scores. The ANOVA results, F test= 20.5, $p = .011$, indicated that there were differences in knowledge and attitudes regarding pain based on two variables, type of program and institution; thus, this is a two-way ANOVA.

When more than two groups are compared over time, a repeated-measures ANOVA is used, because this variation of the ANOVA takes into account the fact that multiple measures at several times affect the potential range of scores. As an example, Murray et al. (2019) explored new graduate registered nurses' (NGRNs) knowledge and attitudes concerning medical error and patient safety, at three time points (commencement, three months, six months) during their first 6 months of professional practice. A one-way repeated measures ANOVA was used to investigate the effect of time on self-reported knowledge and attitudes regarding medical errors and patient safety in NGRNs.

POST HOC ANALYSIS. When the decision according to the ANOVA is to reject the null hypothesis, this indicates that at least one of the means is not the same as the other means, as in Rajacich and associates' (2014) study. To determine where the difference in means lies, a post hoc analysis is conducted; in this analysis, pairs of means in the main effects and interaction effects are compared to determine whether they are statistically different. Many post hoc analyses are available; the most common include Tukey's Honestly Significant Difference (HSD), the Scheffé analysis, and the Bonferroni analysis. This type of post hoc analysis is also known as paired comparisons. In the Hroch et al. study (2019), post hoc comparisons (Tukey's HSD) revealed that the statistically significant difference was between the B.SC.N. program in situation A and both programs in institution b-the B.Sc.N and Practical Nursing (P.N.). There was no statistically significant difference in KASRP scores between the B.Sc.N. and P.N. programs in institution B.

In the study by Dahlke et al. (2019), post hoc comparisons using Tukey HSD test that the mean Burbank score for students in their fourth clinical rotation was significantly different from students who had not had a clinical rotation and from students who had one clinical rotation.

Research Hint

A research report may not always refer to the test that was done. The reader can find this information by looking at the tables. For example, a table with t statistics contains a column for t values, and an ANOVA table lists F values.

In Premji et al.'s (2018) study examining what it means to be a mother of late preterm infant, including a mother's level of confidence in caring for her late preterm infant over time (3–4 weeks and 6–8 weeks) , and the effect of maternal depression of this experience, several paired t-tests were done on maternal confidence. There was a significant decrease in confidence between Time 1 and Time 2, on such the knowledge subscale in relation to "*I know when my baby wants me to play with him/her,*" "*when my baby is cranky, I know the reason,*" and "*I can tell when my baby is*

tired and needs to sleep." In other cases, particularly in experimental work, researchers use t tests or ANOVA to determine whether random assignment to groups was effective in creating groups that are equivalent before the experimental treatment is introduced. In this case, a researcher wants to show that no difference exists among the groups.

In many cases, researchers check whether groups are different at the beginning of a study or baseline by using the technique of analysis of covariance (ANCOVA). ANCOVA also entails measuring differences among group means and helps researchers equate the groups under study on an important variable, to answer the question: are the observed mean differences real or are they false? This false difference is called spurious. ANCOVA allows researchers to control for confounding variables statistically. For example, Tryphonopoulos and Letourneau (2020) tested a video-feedback interaction guidance intervention designed to improve maternal–infant interaction, depressive symptoms, and cortisol patterns of depressed mothers and their infants. They used the ANCOVA to control for pre-test scores on each of the outcome variables (maternal and infant cortisol, maternal depression, Nursing Child Assessment Teaching Scale (NCATS), and CARE-Index (p. 6). Results supported two of the three proposed hypotheses, where maternal and infant cortisol and maternal depression decreased due to the intervention. In the unsupportive hypothesis related to the NCATS scale, the researchers found that there were significant differences favouring the intervention group in scores for *Sensitivity to Cues*, *Cognitive Growth Fostering*, and *Caregiver Total Contingency* subscales (p. 8).

NONPARAMETRIC TESTS. When data are at the nominal or ordinal level and the researcher wants to determine whether groups are different, the chi-square, another commonly used statistic, is helpful. The chi-square (X^2) is a nonparametric statistic used to determine whether the frequency in each category is different from what would be expected by chance. Adhikari et al. (2017) studied whether neighbourhood socioeconomic status predicts the risk of preterm birth. They conducted several chi-square tests to explore maternal characteristics across preterm birth status. One such test was to explore the association between the categorical variable "parity" (primiparous and multiplarous) and another categorical variable "type of birth" (preterm or term birth); therefore the chi-square test was performed and the results were statistically significant. Another example is Benzies et al. (2019), who determined the effect of an enhanced parenting kit given at birth on a) early parenting experiences and (b) use of educational resources and community services; again, these two variables are categorical. They performed a chi-square test between the intervention and comparison group and found that parents in the intervention group were more likely to be aware of the Healthy Parents Healthy Children (HPHC) books than parents in the comparison group (X^2 (1, N = 367) = 5.78, p = .016).

As with the t test and ANOVA, if the calculated chi-square is high enough, the researcher would conclude that the frequencies found would not be expected on the basis of chance alone, and the null hypothesis would be rejected. Although this test is robust and can be used in many different situations, it cannot be used to compare frequencies when samples are small and expected frequencies are less than six in each cell. In those instances, Fisher's exact probability test is used.

When the data are ranks, or are at the ordinal level, several other nonparametric tests may be used: the Kolmogorov-Smirnov test, the sign test, the Wilcoxon matched-pairs test, the signed-rank test for related groups, the median test, and the Mann-Whitney *U* test for independent groups. Explanation of these tests is beyond the scope of this chapter; readers who desire further information should consult a general statistics book.

In nursing research studies, several different statistical tests are often used. Pike et al.'s (2019)

study illustrated the use of several of these statistical tests. They examined the relationship between candidate variables (e.g., academic performance, demographics) on their NCLEX-RN outcome (pass/fail). Data measured at the nominal level were gender, program type (regular vs, fast-track); continuous level variables were age and high school and nursing GPA. For data measured at the interval level, such as age and high school and nursing GPA, the t test was used. Finally, to test the differences between the two groups, the chi-square method was used for nominal variables, such as gender and NCLEX-RN outcome.

Tests of Relationships

Researchers often are interested in exploring the *relationship* between two or more variables. In such studies, they use statistics that determine the correlation, or the degree of association, between two or more variables. Tests of the relationships between variables are sometimes considered to be descriptive statistics when they are used to describe the magnitude and direction of a relationship of two variables in a sample and when the researcher does not wish to make statements about the larger population. Such statistics also can be inferential when they are used to test hypotheses about the correlations that exist in the target population.

In tests of the null hypothesis, no relationship is assumed to exist between the variables. Thus, when a researcher rejects this type of null hypothesis, the conclusion is that the variables are, in fact, related. Suppose that a researcher is interested in the relationship between the age of patients and the length of time it takes them to recover from surgery. As with other statistics discussed, the researcher would design a study to collect the appropriate data and then analyze the data by using measures of association. In this example, age and length of time until recovery can be considered interval measurements. The researcher would use the Pearson correlation coefficient (Pearson *r*; also called the *Pearson product-moment correlation coefficient*) in which the calculation reflects the degree of relationship between two interval variables. The distribution of the Pearson r enables the researcher to determine whether the value obtained is likely to have occurred by chance. Again, the research reports both the value of the correlation and its probability of occurring by chance.

Correlation coefficients can range in value from −1.0 to +1.0 and also can be zero. A zero coefficient means that no relationship exists between the variables. *A perfect positive correlation* is indicated by a coefficient of +1.0 and *a perfect negative correlation by a coefficient* of −1.0. The meaning of these coefficients is illustrated by the example from the previous paragraph. If no relationship exists between the age of the patient and the time required for the patient to recover from surgery, the correlation would be zero. However, a correlation of +1.0 would mean that the older the patient is, the longer the recovery time is. A negative coefficient would imply that the younger the patient is, the longer the recovery time is. Figure 17.5 illustrates a perfect positive correlation, a perfect negative correlation, and a zero correlation. A correlation value of 0 to .2 is considered extremely weak, a value of .2 to .4 is weak, a value of .4 to .6 is moderate, a value of .6 to .8 is strong, and a value of .8 to 1.0 is very strong (Bluman, 2014).

Of course, relationships are rarely perfect. The magnitude of the relationship is indicated by how close correlation is to the absolute value of 1 (see Practical Application box). Thus, a correlation of −.76 is just as strong as a correlation of +.76, but the direction of the relationship is opposite. In addition, a correlation of .76 is stronger than a correlation of .32. In testing hypotheses about the relationships between two variables, the researcher considers whether the magnitude of the correlation is large enough not to have occurred by chance. This is the meaning of the probability value, or the *p* value, reported with correlation coefficients. As with other statistical tests of

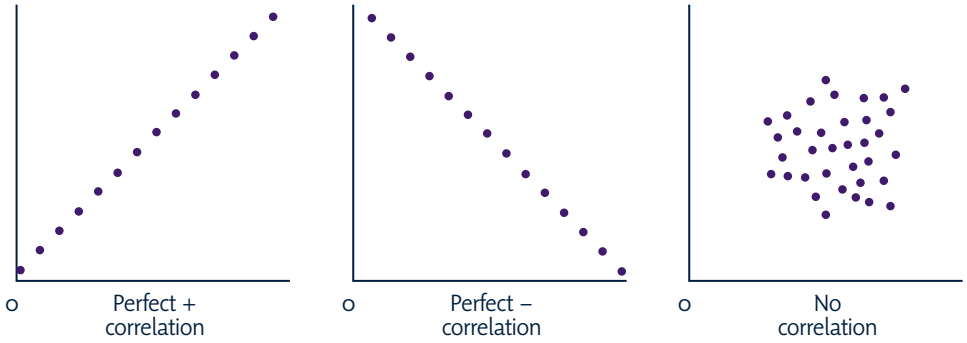


FIG. 17.5 Scatter plots illustrating the different types of correlations.

significance, the larger the sample is, the greater the likelihood of finding a significant correlation. Therefore, researchers also report the degrees of freedom associated with the test performed.

Practical Application

An example of a descriptive, correlational study is that of Levya-Moral et al. (2019), describing nursing faculty attitudes and beliefs about persons living with HIV. There were correlations between stereotype and prejudice ($r = .73$, $p < .001$), stereotype and discrimination ($r = .91$, $p < .001$), and discrimination and prejudice ($r = .83$, $p < .001$).

Winsett et al. (2016) conducted correlations between two interval variables and found that as perception of staff adequacy declined, reasons for missed care increased in importance: communication ($r = -.272$, $p = .006$), material resource ($r = -.240$, $p = .006$); and the labor resource ($r = -.255$, $p = .001$). All of these variables were measured with validated scales.

Nominal- and ordinal-level data also can be tested for relationships by nonparametric statistics. When two variables being tested are only dichotomous (e.g., male/female; yes/no), the phi coefficient can be used to express relationships. When the researcher is interested in the relationship between a nominal variable and an interval variable, the point-biserial correlation is used. Spearman's rho is used to determine the degree of association between two sets of ranks, as is Kendall's tau. All of these correlation coefficients may range in value from −1.0 to +1.0. These tests are listed in Table 17.6.

Nursing problems are rarely so simple that they can be explained by only two variables. When researchers are interested in studying complex relationships among more than two variables, they use techniques other than those discussed thus far. When researchers are interested in understanding more about a problem than just the relationship between variables and in making predictions they often use regression. There are several types of regression including simple, multiple and logistic regression. Simple regression is looking at the relationship of two continuous variables, with one predicting the other, while multiple regression is conducted when the researcher is exploring the relationship between one dependent variable at the interval level and several independent variables is measured. Multiple regression is the expansion of correlation to include more than two variables and is used when the researcher wants to determine what variables contribute to the explanation of the dependent variable and to what degree. Researchers also use logistic and ordinal regression for prediction when variables are binary/ordinal, but these statistical tests are outside the scope of this textbook.

An example of multiple regression is a researcher may be interested in determining what factors help women decide to breastfeed their infants. A number of variables—such as the mother's age, previous experience with breastfeeding, number of other children, and knowledge of the advantages of breastfeeding—might