EVALUATION OF HEALTH CARE QUALITY for DNPs
Evaluation of Health Care Quality for DNPs
Joanne V. Hickey, PhD, RN, FAAN, FCCM, holds of the Patricia L. Starck/PARTNERS Professorship in Nursing and the Coordinator of the Doctor of Nursing Practice program at the University of Texas Health Science Center at Houston, School of Nursing. She received a baccalaureate degree in nursing from Boston College, an MSN from the University of Rhode Island, an MA in counseling from Rhode Island College, a PhD from the University of Texas at Austin, and a post-master’s certificate as a nurse practitioner from Duke University. Dr. Hickey is a fellow of the American Academy of Nursing and a fellow of the American College of Critical Care Medicine. She is a member of numerous professional organizations including the American Nurses Association, Sigma Theta Tau International, Society of Critical Care Medicine, and American Association of Neuroscience Nursing.

Christine A. Brosnan, DrPH, RN, recently retired as an associate professor from the University of Texas Health Science Center at Houston, School of Nursing, where in addition to teaching and clinical research, she held a number of administrative positions, including Track Director for the MSN in Education Program, Assistant Dean for Professional Development, and Associate Dean for Academic Affairs. She received a BSN from Georgetown University, an MSN from the University of Texas Medical Branch in Galveston, and a doctorate in public health from the University of Texas Health Science Center at Houston School of Public Health. The major area of study in her doctoral work was health services organizations. Dr. Brosnan has collaborated in research studies on the cost and effectiveness of newborn screening, screening for hypertension and obesity in children, and the early detection of type 2 diabetes. She has over 35 published articles, abstracts, chapters, proceedings, and letters, and has delivered numerous presentations on health care evaluation and related topics. Her article “Type 2 Diabetes in Children and Adolescents: An Emerging Disease” received the Excellence in Writing Award from the Journal of Pediatric Health Care.
Evaluation of Health Care Quality for DNPs

Second Edition

Joanne V. Hickey, PhD, RN, FAAN, FCCM
Christine A. Brosnan, DrPH, RN

Editors
We dedicate this book to:

- Pioneers and thought leaders . . . the pioneers in the field of evaluation, including Florence Nightingale, Ernest Codman, Avedis Donabedian, and Lu Ann Aday, who have elevated evaluation into a science-based process; and the thought leaders who are demanding that a solid evaluation must be an integral part in all health care endeavors.

- Students and colleagues (current and future) . . . and particularly advanced practice nurses, who have an unprecedented opportunity to impact new models of health care quality that will influence health policy for an ultimate transformation of the health care system.

- Our contributing authors . . . recognized experts in their disciplines, these leaders are much sought after for their expertise in practice, education, research, and consultation. Our request to share their knowledge through writing was met with gracious acceptance and production of excellent manuscripts. Their passion for quality health care and especially the role of evaluation in achieving quality health care is the heart and soul of this book.

- Our husbands . . . who supported our passion for creating this book even though it took considerable time away from other home and family-based activities. Jim Hickey and Pat Brosnan served as in-house editors and wise listeners and counsels as we grappled with the best ways to organize ideas and find the right words to illuminate the work of evaluation.
CONTENTS

Contributors    ix
Preface        xi
Acknowledgments xv

SECTION I: UNDERPINNINGS OF EVALUATION

1. Evaluation and DNPs: The Mandate for Evaluation    3
   Joanne V. Hickey and Christine A. Brosnan

2. The Nature of Evidence as a Basis for Evaluation    37
   Joanne V. Hickey

   Christine A. Brosnan

4. Evaluation and Outcomes    87
   Christine A. Brosnan and Patrick G. Brosnan

5. Economic Evaluation    103
   Christine A. Brosnan and J. Michael Swint

SECTION II: SPECIAL AREAS OF EVALUATION IN HEALTH CARE

6. Evaluation of Organizations and Systems    127
   Nancy Manning Crider and Elizabeth Ulrich

7. Evaluation of Health Care Information Systems and Patient Care Technology    143
   Sharon McLane and Juliana J. Brixey
Contents

8. Program Evaluation 171
   Joanne V. Hickey and Christine A. Brosnan

9. Quality Improvement 189
   Eileen R. Giardino

10. Evaluation of Patient Care Based on Standards, Guidelines, and Protocols 207
    Ronda G. Hughes

11. Health Care Teams 227
    Joanne V. Hickey

SECTION III: EVALUATION OF POPULATIONS AND HEALTH POLICY

12. Evaluating Populations and Population Health 265
    Deanna E. Grimes and Nancy F. Weller

13. Translating Outcomes From Evaluation to Health Policy 281
    Deanna E. Grimes, Richard M. Grimes, and Christine A. Brosnan

14. Future Trends and Challenges in Evaluation 297
    Christine A. Brosnan and Joanne V. Hickey

Index 311
CONTRIBUTORS

Juliana J. Brixey, PhD, MPH, RN  Associate Professor of Biomedical Informatics and Nursing, Department of Family Health, University of Texas Health Science Center at Houston, School of Nursing and School of Biomedical Informatics, Houston, Texas

Christine A. Brosnan, DrPH, RN  Associate Professor, retired, University of Texas Health Science Center at Houston, School of Nursing, Houston, Texas

Patrick G. Brosnan, MD  Professor of Pediatrics, retired, Department of Pediatrics, University of Texas Health Science Center at Houston, School of Medicine, Houston, Texas

Nancy Manning Crider, DrPH, MS, RN, NEA-BC  Assistant Professor, Department of Family Health, University of Texas Health Science Center at Houston, School of Nursing, Houston, Texas

Eileen R. Giardino, PhD, RN, APRN, NP-C, ANP-C  Associate Professor, Department of Family Health, University of Texas Health Science Center at Houston, School of Nursing, Houston, Texas

Deanna E. Grimes, DrPH, RN, FAAN  Professor, Division Head, Community Health Nursing Director, MSN/MPH Program, University of Texas School of Public Health, University of Texas Health Science Center at Houston, School of Nursing, Houston, Texas
Contributors

Richard M. Grimes, PhD, MBA  Adjunct Professor, University of Texas Health Science Center at Houston, Medical School, Houston, Texas

Joanne V. Hickey, PhD, RN, FAAN, FCCM  Patricia L. Starck/PARTNERS Professor of Nursing, University of Texas Health Science Center at Houston, School of Nursing, Department of Family Health, Houston, Texas

Ronda G. Hughes, PhD, MHS, RN, FAAN  Associate Professor, Marquette University, College of Nursing, Milwaukee, Wisconsin

Sharon McLane, PhD, MBA, RN-BC  Retired informatician and nursing services administrator, Lakeland, Florida

J. Michael Swint, PhD  Professor of Health Economics, George McMillan Fleming Professor, Director, Division of Management, Policy, and Community Health, University of Texas Health Science Center at Houston, School of Public Health, Houston, Texas

Elizabeth Ulrich, EdD, RN, FACHE, FAAN  Professor, Department of Family Health, University of Texas Health Science Center at Houston, School of Nursing, Houston, Texas; Editor, Nephrology Nursing Journal

Nancy F. Weller, DrPh, MPH, MS, RN  Assistant Professor, Department of Nursing Systems, University of Texas Health Science Center at Houston, School of Nursing, Houston, Texas
Although evaluation is not a new concept to nurses, it is most often associated with the nursing process designed to evaluate the outcomes of health care for an individual patient. Competency in evaluation of patient outcomes is an expectation of all professional nurses. However, what is new is the expectation of high-level competency for doctor of nursing practice (DNP) graduates in evaluating health care, including health professional groups, patient populations, organizations, systems, programs, health informatics, practice guidelines/protocols, health policy, and other health-related entities, from a systematic and comprehensive evaluation perspective. The bar for evaluation has been raised for DNPs. Along with the higher expectations comes a new opportunity to influence high-level decision making in health care.

Recognizing the need for DNPs to be prepared in evaluation beyond individual patients, we searched for textbooks and other resources that would be helpful for DNPs as they conducted their work. Surprisingly, little was found beyond case studies and general principles, although there were several books on program evaluation. Although helpful, these resources did not address the full scope of evaluation required in DNP advanced practice. With this background, we began planning the second edition of this text, guided with a focus on DNPs with the assistance of input from our colleagues, students, and practicing DNPs. Some of our esteemed colleagues graciously agreed to share their expertise through the written word and have contributed chapters to this book.

The intended audiences are students enrolled in advanced practice nursing programs, especially DNP programs; DNPs and other advanced practice nurses at the master and doctoral levels; nurse administrators; directors of quality improvement; faculty teaching evaluators; and others interested in evaluation of health care from a practice and clinical perspective. In selecting content, it was our intention to provide an overview of the state of the science of evaluation and what is known about evaluation and its application to common practice issues in which DNPs will lead or participate. As we reviewed the literature, it became clear that evaluation as applied to health care is underdeveloped and evolving. It is a non-linear and messy process; there is no one right way to conduct an evaluation. The form of an evaluation is based on its intended purpose and use. The clear and
urgent message is that all aspects of health care entities must be evaluated systematically for effectiveness and to guide and inform decision making. DNPs can contribute to developing the science, processes, and uses of evaluation in health care. Although the consensus is in agreement with a national call to action, the processes and timelines for evaluation are poorly established. The intent of this book is to lay a foundation for DNPs to assume their important role in evaluation.

Section I addresses the underpinnings of evaluation. Chapter 1 elaborates on the role of DNPs in evaluation. Through a brief history, overview, mandate, and other aspects of high-level evaluation, the DNP is brought to the table of evaluation. Chapter 2 addresses the nature of evidence, the basic building block of evaluation, and provides a critical review of characteristics, sources, and quality of evidence as it applies to rigorous evaluation. The conceptual foundations for evaluation are discussed in Chapter 3. A number of frameworks are described to provide the reader with different models for addressing evaluation. Chapter 4 is new to this edition and addresses evaluation and outcomes. The national imperative for cost effectiveness is addressed in Chapter 5 through an overview of economic evaluation.

Evaluation of organizations, systems, and standards for practice is the focus of Section II. Chapter 6 examines the evaluation of organizations and systems, while Chapter 7 addresses health care informatics and patient care technology within health care. With the redesign of health care delivery, organizations and systems are being restructured and redesigned to be more responsive to patient–family-centered care models. An integral part of health care is health informatics, as well as patient care technology integration and evaluation. The current national trend toward electronic medical records is creating challenges because of the far reaching impact on organizations, systems, and individual patients. Chapter 8 is new and addresses program evaluation, a common focus for DNP professional work. Chapter 9, another new chapter, focuses on quality improvement as a developing science and important work in all health care delivery. Chapter 10 discusses the important area of patient care standards, guidelines, and protocols from the perspective of how they are developed, implemented, and evaluated. Finally, Chapter 11 examines the critical role of teams as instrumental in the delivery of high quality and safe care.

Section III addresses the evaluation of population health, health policies, and the future. From a lens of populations, Chapter 12 addresses characteristics, risk factors, determinants, and the evaluation of population health. Chapter 13 discusses the important step of translating outcomes from evaluation into health policy. DNPs are encouraged to seek opportunities for advocacy and leadership in influencing health policy development, implementation, and evaluation. Chapter 14 examines challenges and trends for the future, including the increased demand for comprehensive high-level evaluation by DNPs and related competencies.

The content from the first edition has been updated and chapters added as previously described. Some of the unique features of the book are key definitions of terms, examples to illustrate a point, and case studies to provide exemplars of comprehensive evaluations, including clinical applications and recommended resources for perusal and reference. As educators and practitioners, we were keenly aware of the multiple definitions of key terms. Unless evaluators and users of evaluations are clear about terminology, confusion and misunderstandings abound, leading to underutilized or misdirected evaluations. The examples in the book
come from a variety of practice settings and foci to provide the reader with an appreciation of the multiple uses of evaluation. This is also true of the case studies, which provide a more comprehensive overview of the evaluation process, outcomes, and uses. Readers may wish to explore other resources to augment their understanding and broaden their perspective of aspects of evaluation. Selected resources are thus provided for these purposes.

Our sincere hope is that this book will meet our primary aim of providing a useful and helpful resource to assist DNPs in assuming responsibility and accountability for competency in the conduct of high-level evaluation that will inform decision making for those engaged in health care delivery and practice. It speaks loudly to the recommendation outlined in the Institute of Medicine’s report *The Future of Nursing: Leading Change, Advancing Health* that nurses should be full partners with other health professionals in redesigning health care.

Joanne V. Hickey  
Christine A. Brosnan
ACKNOWLEDGMENTS

We are indebted to many wonderful people who helped make this book possible. We especially wish to acknowledge Dr. Janet C. Meininger, PhD, RN, FAAN, Lee and J.D. Jamail Distinguished Professor in the UT Health School of Nursing, for her helpful review of evaluation and outcomes discussed in Chapter 4.
CHAPTER FOUR

EVALUATION AND OUTCOMES

Christine A. Brosnan and Patrick G. Brosnan

True genius resides in the capacity for evaluation of uncertain, hazardous, and conflicting information.
—Winston Churchill

The previous chapters reviewed the major issues driving health care evaluation in the United States and the context in which doctor of nursing practice (DNP) graduates and other health professionals seek to improve the quality of that care. The chapters discussed acquisition and use of evidence, and the various approaches to quality improvement. Chapter 4 builds on that content to explore the association between evaluation and outcomes in greater depth. This chapter describes recent changes in and challenges to quality measurement, the development of patient-centered outcomes, and sources for valid and reliable measures. The role of comparative effectiveness research (CER) in establishing effective treatment in a real-world context is reviewed and examples provided.

EXPLORING QUALITY IMPROVEMENT

During the past 50 years, health service researchers, health care professionals, consumers, and stakeholders have continued to search for the elusive path to quality improvement. Researchers became increasingly interested in identifying clinical trials in which health-related variables could be compared across studies through either qualitative synthesis or statistical testing of aggregated data. Literature synthesis and meta-analysis studies now provide an evidence base for practice and for developing health care guidelines (Chassin & Loeb, 2011).

Utilization review committees and professional standards review organizations, which grew out of the Medicare legislation of 1965, came to be viewed as conduits for improving quality in participating hospitals (Chassin & Loeb, 2011). In 1994, the Clinton administration proposed legislation known as the Health Security Act (Mariner, 1994). Although the bill was not approved by Congress, the Health Security Act provided an impetus for an expanded discussion of more comprehensive quality improvement activities. These activities included the development of quality measures, the creation of computer centers to store data
collected from hospitals and other clinical facilities, and the formation of health care standards (Sadeghi, Barzi, Mikhail, & Shabot, 2013).

The Joint Commission (TJC), originally called The Joint Commission on Accreditation of Health Care Organizations, encouraged the use of evidence-based performance metrics and the application of Donabedian’s structure-process-outcome indicators in evaluating hospital quality (Chassin & Loeb, 2011). In 1999, the Agency for Health Care Policy and Research was renamed the Agency for Healthcare Research and Quality (AHRQ) and was charged with gathering and disseminating evidence on best health care practices. The National Quality Forum, a not-for-profit-organization formed in the same year, was tasked with supporting the application and dissemination of quality metrics (Sadeghi et al., 2013). In 2001, the report Crossing the Quality Chasm recommended redesigning the health care system to place quality improvement in the forefront (Institute of Medicine [IOM], 2001). The Affordable Care Act of 2010 contained major initiatives designed to improve health outcomes, along with incentives to help accomplish the goal.

Today, there are numerous governmental and nonprofit entities striving to construct the most direct path to quality improvement. Governmental entities such as AHRQ, Centers for Medicare & Medicaid Services (CMS), and the National Quality Measures Clearinghouse (NQMC) contain sections devoted to reviewing and disseminating valid and reliable metrics. Nonprofit organizations including the IOM, TJC, the National Committee for Quality Assurance, and the Institute for Health Improvement offer guidance to providers, patients, and health care facilities who seek to measure quality care (Sadeghi et al., 2013).

DEVELOPING PATIENT-CENTERED OUTCOMES

In 1966, Donabedian stated that outcomes were “the ultimate validators of the effectiveness and quality of medical care” (Donabedian, 1966, p. 169). Practically, however, outcome measures are more difficult to develop and collect than process measures. Outcome is more patient centered than process, and thus harder to identify and link with a specific treatment or intervention (Chin, 2014). Changes in health status often occur because of complex interactions and cannot always be attributed to a linear association of one provider to one patient. Donabedian discussed the concept of attribution and the possibility of weighting all factors that contribute to a change in health status. Thus, under certain circumstances, the contributions of physicians, nurses, therapists, pharmacists, families, and patients might be identified, measured, and analyzed to determine their impact on outcomes (Donabedian, 2003).

Process includes provider interactions, which are often easier to identify and describe than outcomes. They are also more apt to align with health care guidelines, which tend to focus on what a caregiver does. Process measures may be obtained by documenting the number of procedures billed, the amount of immunizations provided, or the number of patients screened for hypertension. Porter (2010) asserted that the majority of measures currently collected are process measures; and that, while they may reflect current quality standards, they do not necessarily correlate with patient health status. Over time, experts in the field have
reached the general consensus that more needs to be done to develop and use valid and reliable patient-centered outcome measures (Cassel et al., 2014; Lee, 2010; Porter, 2010).

An outcome refers to a change in patient health status that results from health care delivered (Donabedian, 1988, 2003). In the earlier example, an outcome measure might refer to disability or death among those patients who were billed for a specific procedure, who received immunizations, or who were screened for hypertension. Donabedian's guidelines for developing outcomes (Donabedian, 2003) are listed in Exhibit 4.1. While appearing straightforward, the guidelines are complex and require a serious commitment of resources. For example, the fourth guideline admonishes the evaluator to monitor outcomes not only for how great a difference an intervention made, but also for how long the change in health status was maintained. And the sixth guideline recommends collecting data for the amount of time it takes for the change to become apparent. The guidelines also include the need to make judgments. The fifth guideline suggests that some individuals will want to trade quantity of life for quality of life and that the individual's desires must be considered. Finally, the cost of the intervention must be considered because dollars spent on evaluating one health care intervention will not be available to spend on another alternative. This is known as an opportunity cost and will be discussed in the next chapter.

Donabedian (2003) listed four types of outcomes: (a) alteration in health status; (b) alteration in patient knowledge; (c) alteration in patient or family activities; and (d) patient and family satisfaction. He classified outcomes as clinical, physiological-biochemical, physical, psychological, social, integrative, and evaluative.

### EXHIBIT 4.1

**Guidelines for the Development of Outcomes as an Indicator of Quality**

1. The outcome selected should be relevant to the objective of care; it stands for what the clinician is aiming for.
2. The outcome must be achievable by good care. This means that the methods for this are available and under the control of the health care system.
3. The outcome, whether good or bad, must be attributable first to health care, and then, to the contribution of the practitioner or other person whose performance is being assessed.
4. The duration of the outcome as well as its magnitude should be taken into account.
5. As a corollary, the trade-off between levels and duration of alternative outcomes may be considered. For example, a shorter life at a higher level of function may have to be weighed against a longer life with greater disability.
6. As another corollary, information on the relevant outcome must be available, which is not an easy matter, especially when obtaining the information requires follow-up over long periods.
7. It is necessary to track not only the consequences of taking action but also the consequences of not taking action in order to obtain a complete picture of performance.
8. Finally, the outcome cannot stand alone. The means used to achieve the outcome also have to be considered, unless it is assumed that resources are unlimited, which almost always is far from true.

*Source: Donabedian (2003). Used with permission.*
There are a variety of perspectives through which to view outcomes. Aday, Begley, Lairson, and Balkrishnan (2004) categorized outcomes from a population perspective and a clinical perspective. They further divided clinical outcomes into system, institution, and patient subcategories. An example of a population outcome is the incidence of measles in a city between January 1st and December 31st. In a system or health care facility, the outcome might be the diagnosis of the disease on the first visit. For an individual patient, the outcome may be recovery without sequelae.

Regardless of the perspective, if an intervention involves more than one person, the only way to aggregate and analyze outcomes is to measure them. The following section presents measures that are frequently used in health care evaluation.

**CREATING MEANINGFUL OUTCOME MEASURES**

**Measuring Health Care Outcomes**

Evaluators quantify health care outcomes using counts, means, or medians. For example, they may calculate the number of adverse events, the mean systolic and diastolic blood pressure, or the median length of hospital stay. They also use proportions and percentages. Proportions are calculated by counting the number of persons with the outcome of interest and dividing that number by the number of all persons treated during a specified period of time (Romano, Hussey, & Ritley, 2010). For example, 600 out of 1,000 patients (60%) surveyed during 2015 at a city hospital said they were satisfied with the care they received during their hospital stay. Other outcomes associated with percentages include symptoms, pressure ulcers, falls, and readmission within 30 days.

Rates, proportions, and ratios are used as measures of disease frequency and are often used to evaluate health outcomes, particularly in large health care facilities, communities, and populations. A description of commonly used frequencies is provided in Table 4.1. Two major types of frequency are prevalence and incidence. Prevalence refers to the proportion or percent of the total number of individuals in a population who are known to have an existing condition at a particular point in time (Greenberg, 2015; Hennekens & Buring, 1987). Prevalence is useful in determining if a chronic condition such as diabetes or cardiovascular disease is becoming more common and in estimating the total number of patients who need treatment.

Cumulative incidence refers to new events in a population at risk during a certain time interval. The population at risk is defined as all members of that population who are susceptible to the outcome of interest. For instance, a member of the population that is a prevalent case is not at risk of becoming an incident (i.e., new) case. Incidence rates are essential to studies of causes of disease because they are used to make inferences about risk or probability of disease (Greenberg, 2015; Hennekens & Buring, 1987).

Morbidity and mortality are particular types of incidence. Morbidity refers to the appearance of diagnosable disease and mortality refers to death. All-cause mortality, cause-specific mortality, and case fatality provide different kinds of information. All-cause mortality includes everyone who died in
### TABLE 4.1  Selected Measures Used in Evaluating Treatment Outcomes

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative incidence</td>
<td>Number of new occurrences in a population</td>
</tr>
<tr>
<td></td>
<td>All individuals at risk in the population</td>
</tr>
<tr>
<td>Prevalence</td>
<td>Number of persons who have an existing condition</td>
</tr>
<tr>
<td></td>
<td>Entire population</td>
</tr>
<tr>
<td>All-cause mortality</td>
<td>Number of persons who died from all causes</td>
</tr>
<tr>
<td></td>
<td>All persons in the population</td>
</tr>
<tr>
<td>Cause-specific mortality</td>
<td>Number of persons who died from a specific disorder</td>
</tr>
<tr>
<td></td>
<td>All persons in the population</td>
</tr>
<tr>
<td>Case fatality rate</td>
<td>Number of persons who died from a specific disorder</td>
</tr>
<tr>
<td></td>
<td>Number of persons with the disorder</td>
</tr>
<tr>
<td>Relative risk</td>
<td>Incidence among persons exposed</td>
</tr>
<tr>
<td></td>
<td>Incidence among persons not exposed</td>
</tr>
<tr>
<td>Relative risk reduction</td>
<td>$1 - \text{relative risk}$</td>
</tr>
<tr>
<td>Absolute risk reduction</td>
<td>Incidence of outcome in the treated population minus the incidence of outcome in the untreated population</td>
</tr>
<tr>
<td>Number needed to treat</td>
<td>$\frac{1}{\text{absolute risk reduction}}$</td>
</tr>
<tr>
<td>Number needed to screen</td>
<td>Number Needed to Treat</td>
</tr>
<tr>
<td></td>
<td>Disease prevalence</td>
</tr>
</tbody>
</table>

Sources: Hennekens and Buring (1987); Kendrach, Covington, McCarthy, and Harris (1997); Rembold (1998).

a specific place during a specific time period divided by the total population. In cause-specific mortality, the numerator includes only those individuals dying from a particular disorder in the total population. In case fatality, the numerator includes the individuals who died from a certain disorder and the denominator includes only those who have the disorder. When calculating the frequency of less common diseases, rates are often multiplied by 1,000, 10,000, or 100,000 so that they may be expressed as a whole number (Greenberg, 2015; Hennekens & Buring, 1987).

Measures of association compare rates in populations to estimate the probable amount of benefit or harm caused by an agent of interest. In relative risk (RR), the incidence of a studied event in an exposed population is divided by the incidence in an unexposed population. The exposure may refer to an environmental factor, disorder, diagnostic test, or treatment. A result less than one indicates lower risk (for harm or benefit) and a result greater than one indicates more risk (for harm or benefit) to the exposed population. A value of one indicates that the exposure had
little or no effect. The result may be expressed as a percent (Greenberg, 2015; Hennekens & Buring, 1987). For example, compared with a population of adults who were not exposed, a population exposed to a carcinogen over a 5-year period had an RR of 1.07 or a 7% greater risk of developing cancer than those never exposed to the carcinogen. In discussing a benefit of treatment, epidemiologists sometimes refer to relative risk reduction (RRR), which is 1-relative risk (Kendrach, Covington, McCarthy, & Harris, 1997). For instance, if 93 of 1,000 patients treated for hypertension go on to have strokes, and 100 of 1,000 untreated patients have strokes, the RR of stroke is 0.93 and the RRR is 1 minus 0.93, which is 7%. This is the percentage of baseline risk that is reduced with treatment.

Absolute risk reduction (ARR) is obtained by subtracting the incidence of an outcome among untreated persons from the incidence among treated persons. ARR, which is generally expressed as a percentage, is a meaningful way to compare the results of randomized controlled trials (RCTs) because it estimates the actual benefit of the experimental treatment to a population. Its magnitude depends on the baseline risk that occurs in the absence of the experimental treatment (Greenberg, 2015; Kendrach et al., 1997). RRR compares only the numerators of risk but ARR compares the denominators as well. RRR would be the same 7% if stroke decreased from 100 to 93 per 1,000 or from 100 to 93 per 100,000, but ARR would fall from 0.07% to 0.00007%.

A measure frequently used along with the ARR is the number needed to treat (NNT). The NNT offers health care providers and decision makers an estimate of the total sum of individuals who must receive an intervention before obtaining a beneficial (or preventing a harmful) result. The NNT (Greenberg, 2015) is the inverse of the ARR (1 divided by ARR).

**Determining the Effectiveness of Treatment**

For example, a researcher studying disease X conducts an RCT of an intervention, and finds that subjects given treatment A are less likely to have an unwanted outcome than those given no treatment. This desired result may be expressed as an RR, the incidence of the unwanted outcome among the treated patients divided by the incidence of the unwanted outcome among untreated patients. In this study, 1 out of 1,000 patients treated with medication A have the targeted unwanted outcome during the study period and 5 of 1,000 untreated patients have the unwanted outcome during the same period. The RR for the unwanted outcome is 20% (0.001/0.005) and the RRR is 80% (1 minus 20%). The researcher correctly proclaims that treatment A reduced the risk of disease X by 80%. However, before adopting the treatment a decision maker with finite resources must consider the impact of treatment for other diseases in the same population.

For example, assume that disease Y has the same unwanted outcomes but is more prevalent than disease X. In a similar study of disease Y, 100 of 1,000 patients treated with medication B have the targeted unwanted outcome during the study period, and 500 of 1,000 untreated patients have the unwanted outcome during the same period. The RR for the unwanted outcome is 20% (0.1 for persons on treatment B divided by 0.5 for persons not receiving treatment). The RRR is 80% (1 minus 20%). Both results seem equally positive, and may be reported so, but to a society considering programs based on effectiveness and cost, the impact is different because of the higher prevalence of the unwanted result in the population.
with disease Y. The risk changed from 0.005 to 0.001 in the study of disease X and 0.5 to 0.1 in the study of disease Y. The ARR for treatment of disease X is 0.004 and the ARR of disease Y is 0.4.

In the first example (disease X), only 4 patients are helped by treatment, but in the second scenario (disease Y) 400 patients benefited because the risk of the unwanted outcome in the 1,000 untreated subjects with diseases X and Y is different. Assuming that the unwanted outcomes of the diseases are similar, a decision maker with limited resources who had to choose between these competing interventions would probably elect to implement the treatment for disease Y because the impact on the population will be greater.

Another way to interpret the data is by using the NNT, which is the inverse of the ARR. In disease Y, the ARR is 0.4 and the NNT is 2.5 (1 divided by 0.4). This means that only 2.5 patients need to receive treatment B before one unwanted outcome is prevented. Practically, if 1,000 patients are treated, 400 patients will avoid the unwanted outcome. Applying the NNT with disease X (with an ARR of 0.004 and an NNT of 1/0.004 or 250), 1,000 patients need to receive treatment A in order for four patients to avoid an unwanted outcome. Thus, 250 patients need treatment to save one patient, and after 1,000 treatments only four patients benefit. A decision maker may consider not only the impact of the interventions but also the cost. If the cost of treatment is $10,000 for both disease X and disease Y, a patient with disease Y can be saved from the unwanted outcome for $25,000 but society would need to spend $2,500,000 to prevent an unwanted outcome with disease X. The price of the proposed intervention multiplied by the NNT determines the cost of the benefit but, of course, high NNT may be tolerable for inexpensive interventions.

These calculations depend on evidence-based data, usually from one or more RCTs about a disorder with a known prevalence, treated in a specified manner to avoid a specific unwanted outcome. Calculating the NNT is only useful if the source studies are well conducted and comparable (Greenberg, 2015).

**Measuring Screening Outcomes**

Screening involves testing individuals who seem healthy now, to find out if they may have a disorder that has not yet been diagnosed but can be cured or helped with treatment before symptoms develop. According to Wilson and Jungner (1968), screening can be universal (all individuals are screened) or focused (only individuals at high risk are screened). Screening is not diagnostic. There will be, depending on the screening test, a few or many persons with positive results who are later found not to have the disorder. And, depending on the test, there will be a few or many persons with negative results who are later diagnosed with the disorder when they manifest symptoms. There are specific measures that apply to all screening tests that can be used to evaluate validity and reliability. Knowing how these measures work together will help the DNP evaluate and compare screening programs.

Every screening test has four possible outcomes. A positive test result is a *true positive (TP)* if, after diagnostic follow-up, an individual is found to have the disorder. A positive test result is a *false positive (FP)* if, after diagnostic follow-up, an individual is found not to have the disorder. A negative test result is *true negative (TN)* if an individual does not go on to develop the disorder. A negative test is
false negative (FN) if, despite the negative result, an individual develops symptoms and is found to have the disorder (Greenberg, 2015; Guyatt, Sackett, & Haynes, 2006; Hennekens & Buring, 1987).

**Determining the Effectiveness of Screening**

Let us assume that there is a chronic condition called beta disorder that children develop early in life. Beta disorder occurs in an estimated 1/5,000 young children and causes seizures. There is no easy way to establish if some children are at higher risk for the disorder than others. A screening test becomes available that measures the amount of beta micrograms in the blood. If there are 75 or more micrograms, the test is positive. If there are less than 75 mcg per deciliter, the test is negative. A positive screen can be confirmed and the disorder diagnosed with additional blood tests and with magnetic resonance imaging (MRI). The diagnosed disorder is treatable with daily injections.

A universal screening program becomes available and all parents are encouraged to have their children tested before 12 months of age. As Table 4.2 indicates, out of 1,000,000 children screened, 10,196 children tested positive for the disorder, including 198 children who after follow-up with further blood tests and an MRI were diagnosed with beta disorder (TP) and 9,998 children who after further diagnostic tests were found to be healthy (FP). There were 989,804 children who tested negative for the disorder, including 989,802 children who never developed the disorder (TN) and 2 children who later developed symptoms and were subsequently diagnosed with beta disorder (FN).

Sensitivity is the chance that a child with the disorder has a positive test. It is calculated by counting the number of children who had TP tests and dividing it by the number of children tested who went on to be diagnosed with the disorder. The sensitivity of the beta screening test was 99%. Specificity is the chance that a child without the disorder had a negative test. It is calculated by counting the number of children who had TN tests and dividing it by the number of children tested who did not have the disorder. The specificity of the beta test was 99%.

**TABLE 4.2 Results of a Screening Program to Identify Children With Beta Disorder**

<table>
<thead>
<tr>
<th>Screening Test</th>
<th>Beta Disorder Present</th>
<th>Beta Disorder Absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>198</td>
<td>9,998</td>
<td>10,196</td>
</tr>
<tr>
<td>Negative</td>
<td>2</td>
<td>989,802</td>
<td>989,804</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>999,800</td>
<td>1,000,000</td>
</tr>
</tbody>
</table>

**Screening Measures**

- **Sensitivity** = \( TP/TP + FN = 198/198 + 2 \)  
  Results: 99%

- **Specificity** = \( TN/TN + FP = 989,804/989,802 + 9,998 \)  
  Results: 99%

- **Positive predictive value** = \( TP/TP + FP = 198/198 + 9,998 \)  
  Results: 1.9%

- **Negative predictive value** = \( TN/TN + FN = 989,802/989,802 + 2 \)  
  Results: 99.9%
value is the chance that a child with a positive test has the disorder. It is calculated by counting the number of children who had a TP test and dividing it by the number of all children with positive tests. The positive predictive value of the beta test was 1.9%. Negative predictive value is the chance that a child with a negative test did not have the disorder. It is calculated by counting the number of children who had a TN test and dividing it by the number of all children with negative tests. The negative predictive value of the beta test was 99.9% (Greenberg, 2015; Guyatt et al., 2006; Hennekens & Buring, 1987).

Both sensitivity and specificity should be as close to 100% as possible, although they tend to have an inverse association. The beta screening test had a high sensitivity (99%) and specificity (99%) and negative predictive value (99.9%). However, the test had a low positive predictive value of 1.9%, which means that the chance of a positive test accurately predicting the presence of the disorder was extremely small. As a consequence, the concerned parents of the 9,998 children with FP screening results may spend time and money taking their perfectly healthy children for follow-up blood tests and MRIs that were not needed. Positive predictive values tend to be small when the prevalence of the screened-for disorder in a population is low (Greenberg, 2015; Hennekens & Buring, 1987).

Recently, there has been interest in applying the concept of NNT to screening for more common adult disorders whose risks lie dormant in the population (Rembold, 1998). This introduces another layer of complexity but can yield useful public health information if done correctly. Number needed to screen (NNS) is the NNT derived from an RCT of treatment in the targeted disease divided by the prevalence of undiagnosed target cases in the population. To apply NNS, an evaluator needs a reliable estimate of the prevalence of undetected target disease in the general population as defined by the same diagnostic criteria that were used to include subjects in the reference RCT. Using the number of positive screens in the asymptomatic population is not recommended, because screens typically find many mild cases that may not have the same outcome risk that subjects in the RCT had.

Let us use disease Y, discussed in the prior section, as an example. With an ARR of 0.4 and an NNT of 2.5, treating the disease is beneficial, but there is a problem. Very few affected people know they have the disease in time for treatment to help. Assume that a blood test is developed that can identify 100% of the affected individuals (not likely, but for simplicity) at a cost of $10 per screen. A study of 1,000 randomly chosen subjects finds 21 individuals who would have qualified for the study, and that only one of these knew he had the problem. The researchers assume that screening 1,000 individuals will identify 20 new individuals (2%) who might benefit from treatment. In this example, 2% of the population has undiagnosed treatable disease Y during a specific time period. Based on the NNT of 2.5 from the treatment RCT and the 2% prevalence of undiagnosed cases, the NNS is 125 (2.5 divided by 0.02). This is the number of people from the general population who must be screened and, if positive, treated to save one person from the targeted bad outcome.

Determining the effectiveness of a screening test is complex, involving issues intrinsic to the test and to the natural history of the target disease. The results of a screening test are not definitive but must be confirmed by further study. No screen has only TPs and TNs; and even if a test has a 99% sensitivity and specificity there remains much work to be done in clearing FPs and FNs. Screening is not meant to
diagnose a disorder but only to identify individuals who may have it. Buried in the statistics of sensitivity and specificity are technical aspects of the applied test, principally expressed as validity and reliability.

Discussions of validity and reliability are about how confident we can be acting on the test data. Validity refers to the accuracy of the test method in detecting its target indicator. In the earlier pediatric screening example, this would be the test's accuracy in finding a beta level of 75 mg/dL as assessed by repeated measurement of known standards, and not by its ability to diagnose the target disease. Reliability is the ability of a specific test to reproduce the same result every time within one and repeated assays, which is of course influenced by test method and operator skill. For a test to be reliable, everyone involved must conduct the screen using the same methods. When those conducting the tests vary in their amount of training, experience, and expertise, it is hard to know how confident to be about the results. Vision screening conducted in a school setting is one example. When testing is done by volunteers, school nurses, or optometrists, each using slightly different methods, reliability may be questionable (Powell & Hatt, 2009).

Besides understanding the test methods, one must know the natural history of the target disease. In addition to identifying the treatable disease, many screening tests also pick up previously unseen mild forms of the disorder which may not need treatment, or abnormalities that will never progress to disease status. Deciding whether treatment might be beneficial or harmful can be difficult for the provider and anxiety producing for the patient in cases where the natural history is insufficiently understood (Welch, 2015). Screening may help to answer our questions about the natural history of the disease but such screening may come under the purview of research.

Another important consideration related to an understanding of the target disease is the availability of treatment. The DNP should be able to assure individuals that treatment is available for those with a positive screen and diagnosis. If treatment is unavailable, the ethical issues of screening become more complex. Screening may be done for genetic counselling to the patient's benefit or for research, and the DNP should be clear about the primary purpose of the screening test.

Further, consider whether treatment initiated at an asymptomatic stage of the disease process leads to better outcomes than treatment initiated after the disease becomes clinically apparent. Finally, if physical symptoms of a condition can be easily identified and treatment started without a screening test, then screening makes no difference (Wilson & Jungner, 1968). Every intervention has risk and screening is no exception.

**A MAZE OF MEASURES**

During the last 10 years, the U.S. health care system has made great progress in creating quality measures (Scott & Jha, 2014). In fact, thousands of measures have been described. Meyer et al. (2012) concluded that this proliferation of quality measures is a costly endeavor and threatens to dilute their influence on health care. Cassel et al. (2014) observed that while over 500 measures were used by state and regional entities, 80% of them were applied in only one program. They recommended that health care evaluators collaborate in developing a limited set of
metrics that are balanced and targeted to provide information critical to quality improvement.

One way to control the growing maze of measures is to organize them into coherent and manageable domains. Fortunately, the quality domains described by the IOM in *Crossing the Quality Chasm* (2001) appear to meet this need. Safe, timely, effective, efficient, equitable, and patient-centered health care has become the sine qua non of quality improvement. The IOM domains are widely accepted by health care facilities, providers, governmental agencies, and consumers as theoretically sound and practical (Greenberg, 2015; Sadeghi et al., 2013). They have been incorporated into the Triple Aim framework (Institute for Healthcare Improvement, 2015) and serve as a framework for research studies. Used in conjunction with Donabedian’s approaches of structure–process–outcome, they can facilitate a comprehensive and comprehensible evaluation of health care (Greenberg, 2015).

**SOURCES OF MEASURES**

There are numerous online resources that provide valid and reliable quality measures that may be used in a variety of settings. Perhaps the most inclusive resource is the NQMC (www.qualitymeasures.ahrq.gov). The NQMC is an enterprise supported by AHRQ and the U.S. Department of Health and Human Services. It provides a website with information including tutorials on such topics as establishing the validity of quality measures and deciding which outcome metrics are appropriate for specific health care settings (Sadeghi et al., 2013).

In describing the essential characteristics of quality measures, the NQMC recommended that they (a) respond to a situation that needs improvement, cover a variety of demographic groups, and be significant to stakeholders, providers, public health officials, and patients; (b) have a foundation in clinical evidence; (c) be valid, reliable, understandable, and permit case-mix adjustment; and (d) have available data sources and clear, precise methodologies. Data sources for measures include electronic health records, surveys, imaging and laboratory data, organizational protocols, provider attributes, public health, and registry information (Greenberg, 2015; NQMC, 2015a).

Risk adjustment is a critical consideration in outcome measurement. Individuals receiving health care vary by age, gender, socioeconomic status, race/ethnicity, education, and so on, and these demographic variables impact how patients respond to health care. Comorbidities must be taken into account as they make treatment more complex and sometimes less effective. In cases with a wide variety of risks, large sample sizes are needed to calculate meaningful outcome measures (Aday et al., 2004; NQMC, 2015a; Porter, 2010).

In addition to providing information, the NQMC contains an extensive database of quality measures. The database is an outgrowth of prior AHRQ programs that offered consumers, providers, hospitals, and agencies a central location for accessing valid and reliable quality metrics. The Computerized Needs-Oriented Quality Measurement Evaluation System (CONQUEST) and the Expansion of Quality of Care Measures (C-SPAN) are two of the measure sets found on the NQMC site.

The NQMC has two main groups of measures: those related to health care delivery and those related to population health. A structure–process–outcome
Underpinnings of Evaluation

approach is used in both groups of measures and the description of quality is derived from the IOM. Efficiency is incorporated as an aspect of quality measurement because in any realistic appraisal the use of resources must be considered. Measures included in the database must meet a rigorous screening process to ensure that they are current, evidence-based, valid, and reliable (NQMC, 2015a).

The site contains thousands of measures listed under a number of categories including topic, organization, domain, and endorsement by the National Quality Forum. One of the most useful tools on the site is a matrix that allows users to select among numerous domains those in which they are most interested. An example using the domains “Institute of Medicine” and “Primary Measure” is presented in Exhibit 4.2. Primary measures represent the main focus of an intervention as opposed to secondary measures, which represent a subordinate focus. The structure–process–outcome measures in the exhibit relate to palliative care for adults (NQMC, 2015b).

COMPARATIVE EFFECTIVENESS RESEARCH

RCTs may establish that an intervention is efficacious under ideal conditions, but most health care is not provided under ideal conditions. An objective of CER is to evaluate alternative interventions in order to determine the most effective means available to improve health outcomes in actual health care settings (IOM, 2009). The aim is to disseminate the evidence-based findings to practitioners, patients, consumers, stakeholders, and policy makers who will then use the knowledge to make informed decisions about health care. The methodology of CER includes clinical trials, observational studies, secondary data analysis, meta-analysis, literature synthesis, and computer modeling (Kaiser Family Foundation, 2009; Titler & Pressler, 2011).
A good example of CER is a study by Voss et al. (2011) in which the authors applied the findings of a prior RCT (Coleman, Parry, Chalmers, & Min, 2006) to their own clinical setting. The RCT had established the efficacy of care transition interventions among a sample of Medicare patients who had been admitted to a health care facility with diagnoses of cardiac or respiratory disorders. The intervention included a transition coach who visited patients at the hospital and at home, and followed up with telephone calls. The setting was an integrated hospital system. The outcome was 30-day readmission.

Voss et al. sought to determine the effectiveness of the same intervention by applying it in nonintegrated hospital settings using a quasi-experimental prospective cohort design. The study group consisted of 257 fee-for-service Medicare patients who were compared with internal controls (n = 736) and external controls (n = 14,514). Results indicated that 12.8% of patients who received the transition intervention were readmitted within 30 days compared with 18.6% in the internal and 20% in the external control group. The authors concluded that these significant findings provided good evidence that care transition interventions were effective in a real-world health care setting.

In another interesting example, Xian et al. (2015) studied the link between warfarin treatment and outcomes among elderly patients who experienced ischemic stroke and atrial fibrillation. Based on the results of RCTs, clinical guidelines support the use of warfarin, but questions remained regarding its benefit in older populations and those at increased risk for hemorrhage. The researchers used an observational design and registry data. During the entire study process, they consulted groups of patients about outcomes that were most important to them. The study was sponsored by the Patient-Centered Outcomes Research Institute (PCORI).

The study population included 12,552 persons discharged to the community from 1,487 hospitals during a 3-year period. The mean age of the 11,039 persons (88% of the study population) in the warfarin treatment group was 80.1, compared with 83.1 in the group with no anticoagulation treatment. Results indicated that warfarin significantly increased both the amount of patient time at home and time without a major cardiovascular event. There was also a significant decrease in all-cause mortality. These findings were apparent even among the most elderly patients, females, and those who had experienced devastating strokes. Providers sometimes hesitate in prescribing warfarin to these groups, thinking that the medication places them at higher risk. However, this study demonstrated in an actual community setting that warfarin was effective for these patients.

PCORI, established in 2010, continues to support CER by increasing the number and quality of patient-centered comparison studies, disseminating results, encouraging implementation of findings in a timely manner, and persuading other entities to support research that is shown to be beneficial in a real-world setting. Through September 2013, PCORI committed more than $300 million in funding to studies that addressed national priorities including the assessment of prevention, diagnosis, and treatment alternatives for health problems ($116 million), the improvement of health care systems ($77 million), the communication and dissemination of health care research ($42 million), addressing disparities in access and treatment ($52 million), and increasing the rate of patient outcome and methodological research ($28 million). In September 2014, the PCORI Board of Governors approved the recommended FY2015 budget of $463 million to support
new research studies and to sustain current programs. PCORI administrators anticipate that $1.5 billion in funding will be provided to researchers between 2014 and 2017 (Gabriel & Normand, 2012; Newhouse, Barksdale, & Miller, 2015; PCORI, 2014; Selby & Lipstein, 2014).

**CHALLENGES TO QUALITY MEASUREMENT**

Developing health care quality measures is a work in progress. The Affordable Care Act and PCORI provide enormous incentives to develop and use valid and reliable measures to improve health status, but there is a long way to go before their recommendations are realized. Some of the biggest challenges are discussed here.

First, as noted earlier, evaluators can lose their way amid the maze of measurements and methods. There are thousands of quality measures; some are valid, reliable, evidence based, and transparent, and some are not. Some measures evaluate quality improvement on the margins instead of measuring the total effect on health status (Cassel et al., 2014). Some measures overlap. At times, evaluators seem to miss the big picture of quality as a change in patient health status while focusing on the individual pixels of care (Sadeghi et al., 2013).

The sheer number of measures used internally by health care entities, and mandated by external regulatory agencies, insurance companies, and other payers, has increased the cost of quality improvement activities. It is estimated that some systems spend 1% of net patient revenue on these activities. Administrators and policy makers have called for the development of a critical but limited cluster of quality measures for submission to external agencies. This would allow health entities to spend a larger portion of their quality improvement budget on applying measures most specific to their current needs for quality improvement (Cassel et al., 2014; Meyer et al., 2012).

Second, the difficulty in linking structure–process–outcome still exists, and as measures proliferate the problem will only worsen. One cannot simply assume that a compliant administration policy or provider performance produces a good patient outcome unless the association is identified, described, and measured. If a linkage of structure–process–outcome is established and the outcome is positive, the cause of the success can clearly be recognized. Conversely, if the outcome is not positive, one can trace back to process and structure to identify the problem (Donabedian, 2003).

A third challenge is the difficulty in attributing outcomes to particular activities or persons. Outcomes are not physician outcomes or nursing outcomes; they are patient outcomes. Today, health care is delivered by a system that includes administrators, health care professionals, ancillary health personnel, families, and patients (Sadeghi et al., 2013). Weighting the contribution of each of these entities is difficult. There have been attempts to use algorithms or modeling to weight each contribution to a patient outcome (Titler, Shever, Kanak, Picone, & Qin, 2011), but even if weighting could be apportioned, the attribution might still vary among individual providers, facilities, and geographic locations. So far, national guidelines for delineating attribution are not available (Romano et al., 2010).

The challenges are great but DNPs along with policy makers, patients, and other health professionals are making good strides toward quality improvement.
And, as long as the focus remains on the destination, outcomes will continue to improve. As noted in *Crossing the Quality Chasm*, “Perfect care may be a long way off, but much better care is within our grasp” (IOM, 2001, p. 20).

**REFERENCES**


102 I Underpinnings of Evaluation


