

# The Outcome Model of Quality

Susan I. DesHarnais

*“Quality is never an accident. It is always the result of intelligent effort.”*

—John Ruskin

A critical question facing most health care quality improvement efforts is how to evaluate clinical performance. The objectives of this chapter are to:

- Present a conceptual framework for measuring the quality of health care
- Provide a definition of quality that focuses on the outcomes of care
- Present a brief historical overview of outcome measurement in the United States
- Examine the data requirements and risk-adjustment techniques for comparing health outcomes across providers and/or over time

## A CONCEPTUAL FRAMEWORK AND DEFINITIONS OF QUALITY

---

Quality may be defined in many ways and from many perspectives. Dr. Avedis Donabedian (1980, 1982, 1986) observed that definitions of quality ordinarily reflect the values and goals of the current medical care system, as well as those of the larger society of which it is part. In 1980, Donabedian presented his model for categorizing the different

ways that one might measure the quality of health care in a given setting. His model has provided an excellent framework for conceptualizing quality in a broad manner and then classifying the measures that one can use to assess different aspects of the quality of care.

Donabedian began by differentiating three aspects of care:

- *Structure*: The resources available to provide adequate health care. Resources include facilities, equipment, and trained personnel.
- *Process*: The activities of giving and receiving care (the patient's activities in seeking care as well as the practitioner's activities).
- *Outcomes*: Primarily, changes in the patient's condition following treatment; outcomes also include patient knowledge and satisfaction.

In addition, Donabedian broadened the definition of quality to include not just the technical management of the patient but also the management of interpersonal relationships, as well as access to care and continuity of care. The conceptual framework shown in **Table 5-1** allows us to appreciate the complexity of defining and measuring the quality of health care and provides guidance in what aspects of care we might wish to measure. One could then fill in this matrix to apply to a particular setting. For example, in **Table 5-2**, the matrix is applied to the care provided at a cancer center. This approach is important in that it gives a broad definition to quality of care that goes well beyond simply looking at technical management.

In 1988, the U.S. Office of Technology Assessment (OTA) defined quality of care as "the degree to which the process of care increases the

**TABLE 5-1** Donabedian's Matrix for the Classification of Quality Measures

	Structure	Process	Outcome
Accessibility			
Technical management			
Management of interpersonal relationships			
Continuity			

*Source*: Donabedian, A. 1980. The definition of quality and approaches to its assessment. In *Explorations in Quality Assessment and Monitoring* (Vol. 1, pp. 95-99). Ann Arbor, MI: Health Administration Press.

**TABLE 5-2 Donabedian's Matrix for the Classification of Quality Measures Applied to Cancer Care**

	<b>Structure</b>	<b>Process</b>	<b>Outcome</b>
<b>Accessibility</b>	Hours of operation of mammogram facility	Waiting time for mammogram appointment	Satisfaction with various aspects of accessibility
<b>Technical management</b>	Certification of nurses in oncology nursing; availability of various pieces of up-to-date radiation equipment	Systematic use of evidence-based practices	5-year survival rates for stage 1 breast cancer patients ages 50-70 at the time of diagnosis
<b>Management of interpersonal relationships</b>	Physicians and nurses trained in cultural competency techniques	Involving the patient in treatment decisions	Patient satisfaction with whether they were able to participate in treatment decisions
<b>Continuity</b>	Presence of a trained nurse navigator	Number of contacts per patient with the nurse navigator	Patient satisfaction with continuity of care

probability of outcomes desired by the patient, and reduces the probability of undesired outcomes, given the state of medical knowledge.” This is a useful definition because it emphasizes the patient’s role and perspective in choosing among possible treatments. This definition also makes an explicit connection between the processes of treatment that are used and the resulting outcomes, thus demanding that evidence-based medicine be the standard of care. One is therefore forced to focus on evidence of the effectiveness of various treatments from the patient’s point of view. This definition also implies that there are no meaningful or useful measures of quality if there is no effective treatment known for a given condition. Thus, one can use quality measures only for those conditions where the technology is reasonably effective and also acceptable to the patient.

More recently, the Institute of Medicine (IOM) discussed quality of care in a series of reports. *To Err Is Human: Building a Safer Health System* was released in 2000, and *Crossing the Quality Chasm: A New Health System for the 21st Century* was released in 2001. These two reports documented the scope of quality and safety problems in the United States and offered an analysis of these problems. These committees

stressed that quality health care must be all of the following (IOM, 2001, pp. 5–6):

- *Safe*—avoiding injuries to patients from the care that is intended to help them
- *Effective*—providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit
- *Patient centered*—providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions
- *Timely*—reducing waits and sometimes harmful delays for both those who receive and those who give care
- *Efficient*—avoiding waste, including waste of equipment, supplies, ideas, and energy
- *Equitable*—providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status

This definition of quality broadens the earlier definitions of quality to recognize that high-quality care must not only focus on the processes of care (timeliness), patient outcomes (safety and effectiveness), and the patient’s perspective (patient centered), but must also focus on some of the broader requirements of the social and economic system within which health care is provided (efficiency and equity). While recognizing this broader perspective, this chapter will concentrate on the measurement of health outcomes, a difficult enough task in itself.

Why might one choose to use outcome measures, when it is much easier to measure or monitor structure or processes of care? Structure measures are relatively simple to use. In many cases, one can simply do an “inventory” of structural measures by using a checklist of those resources that are thought to be necessary to ensure the capacity for providing a given type of care. The Joint Commission (TJC) took this approach in its early days because there was some agreement that certain structural elements were needed as minimal standards to ensure an environment in which good care was possible. However, it should be evident that adequate inputs alone do not ensure good outcomes. All the structural measures can do is indicate whether a facility has the capacity to provide good care.

Then why not focus on process measures, which take into account professional performance? It is often easier to measure provider performance than it is to measure patient outcomes. Processes of care are generally documented in patient records, and also in billing or claims data sets, since the procedures that are done usually determine the payment that the professional receives. However, there are several problems with using process measures to look at the quality of care. For a process measure to be valid, there must be good evidence regarding what a professional should do under defined circumstances. This means that a particular process must be strongly linked to better patient outcomes, compared with alternative processes.

While it is sometimes possible to use evidence from clinical trials and published studies, and to translate these studies into treatment guidelines, often this is not possible. Clinical trials are often done on carefully selected people/subjects, and compliance is carefully monitored. Once the treatment goes into general use, it does not work in the same way. The people who actually get the treatment may be older, may have comorbid conditions, and may be noncompliant with various aspects of the treatment protocol. Therefore the evidence from clinical trials may not be generalizable to the population for which the treatment is intended.

Because of this problem, it is often necessary for a group (or groups) of experts to translate evidence from clinical trials into treatment guidelines. It is very difficult to develop consensus among relevant professional groups on treatment guidelines and then to develop explicit process criteria that state under what circumstances one should or should not follow the guidelines, due to certain combinations of comorbid conditions, the advanced age of the patient, patient preferences, or other valid reasons. Another difficulty of using process measures is that the provider may do the “right thing in the right way,” but the patient may be dissatisfied, may be noncompliant, or may respond poorly to the treatment. The process, though done correctly, may not always produce the desired outcome.

Using a process measure, rather than an outcome measure, to evaluate the quality of care is valid if and only if there is solid evidence that supports doing so. This means that there is strong evidence that there is a very high correlation between “doing the right thing in the right way” and getting good outcomes. This criterion will be met for some conditions, but not for all. For example, if a certain type of treatment is very effective and has few side effects, then the process of doing that treatment can be

used as a valid and useful quality measure, rather than trying to monitor the effects of the treatment on patients' health status. Unfortunately, not many treatments fall into this category.

### *Outcome Measures*

Outcome measures are what we really would like to use, since the whole point of treating the patient is to increase the probability of outcomes desired by the patient and reduce the probability of undesired outcomes, given the state of medical knowledge, according to the OTA definition of quality previously cited. Outcome measures are, in effect, the "gold standard" for measuring the quality of care.

However, it is much more difficult to gather and analyze outcome data than it is to measure structure or process. Ideally one would like to have data on each patient's health status before and after treatment for a large national sample of patients treated for each common condition. Instead, the only information available in most of our databases is information on what procedures were done and, to some extent, what adverse events occurred. Data on patient outcomes are usually missing.

There are many reasons why useful health status information is often lacking. In most cases, there is a time delay until one can really assess the effect of a treatment on a patient. One must wait until the patient has recovered from the treatment. It is expensive to try to follow up on patients once they have completed treatment and recovery, and it is difficult to systematically measure the health status of each patient after treatment. Moreover, health status following treatment is often not a direct result of the care provided, since outcomes are not determined solely by professional performance. Other patient-related factors, such as comorbid conditions, patient age, patient compliance, and financial resources, also enter into the equation. Unless one can adequately account for these factors, one cannot validly compare the performance of different providers by looking at patient outcomes. Outcomes attained by a provider treating higher risk patients cannot really be compared with outcomes attained by a provider treating lower risk patients unless one can adequately adjust for the impacts of the risks when comparing the providers.

Because of these difficulties in measuring outcomes, we are often forced to measure negative outcomes rather than positive outcomes. Since the purpose of care is to produce the positive outcomes while minimizing

the negative outcomes, this is a real problem. Some examples of the type of positive outcomes we would like to measure include the proportion of patients who have the following outcomes:

- A better score on a depression scale 3 months after a specific drug treatment
- A given level of improvement in range of movement of a joint 1 year following joint replacement
- Greater time between hospitalizations for acute episodes for patients with a chronic disease, such as diabetes or alcohol/drug problems
- Return to work within 60 days after a given type of heart surgery
- A given level of improvement in quality of life after back surgery

Instead, we often end up using available data, and thus measuring negative outcomes, such as the proportion of patients who have the following outcomes:

- Death during their hospital stay
- Unscheduled readmission to the hospital within 30 days of discharge
- Complications of surgery during their hospital stay
- Preventable adverse events, including medication errors, wrong site surgery, and so on, during their hospital stay

While the information on negative outcomes is useful, it is only part of the picture when we are measuring patient outcomes. Instead, we would like to measure quality using data on both positive and negative outcomes of care.

What do we do when we find unacceptably high rates of negative outcomes? The general approach is to go back and see what went wrong with the processes of care the patient received. Sometimes a good process has been described, but the health care professionals are not using it. In that case, we need to understand why they are not willing to use the process. More often, we will discover that we may have to redesign the processes to attain better outcomes. As described in *Crossing the Quality Chasm*:

Health care has safety and quality problems because it relies on outmoded systems of work. Poor designs set the workforce up to fail, regardless of how hard they try. If we want safer, higher-quality care, we will need to have redesigned systems of care, including the use of information technology to support clinical and administrative processes. (IOM, 2001, p. 4)

Risk adjustment is crucial in accurately evaluating providers. In terms of quality, we want to take into account what health outcomes we could reasonably expect from a provider, given the technology available, the severity of the disease treated, and other risk factors of the provider's patients. It is therefore essential to risk-adjust outcome variables to allow for valid comparisons of these outcomes across hospitals.

## INFORMATION TECHNOLOGY/DATA AVAILABILITY CHANGES

---

In the second half of the 20th century, computers and large databases made it much easier to benchmark and monitor the outcomes of hospital care (see Chapter 12). Also, researchers began to develop more sophisticated techniques for modeling risk factors affecting the outcomes of care. The increased availability of data on the use, cost, and outcomes of medical services also enabled consumers, insurance companies, and regulatory agencies to independently analyze trends in the use and costs of health care services and to draw their own conclusions.

In the mid-1980s, the Health Care Financing Administration (HCFA), which is presently known as the Centers for Medicare and Medicaid Services (CMS), began releasing information on hospital mortality rates to the public. Because the methods HCFA used to derive these rates had major flaws, in many cases the findings were invalid. Hospitals needed to defend themselves against such data releases. In some communities, hospitals received negative publicity for having high mortality rates when, in fact, their mortality rates were better than what would have been expected, given the severity and complexity of the cases they treated.

By the late 1980s, several states began to gather mortality data for various types of cardiac surgery. In 1980, the New York State Department of Health and its Cardiac Advisory Committee began an effort to reduce mortality from coronary artery bypass grafts by collecting clinical data on all patients undergoing that procedure. In 1990, the department made public the data on mortality rates, both crude and risk-adjusted. Surgeon-specific data on mortality were released after a lawsuit by a newspaper. Subsequently, other data releases were made, some of which were likely misleading and superficial (Chassin et al., 1996). Understandably, many surgeons and hospitals had unfavorable reactions to these releases. There were concerns with the accuracy of the data, as well as the methods of



risk adjustment. Many of these problems have been resolved, and public releases of high-quality mortality data have become more common.

Pennsylvania has had a similar program of reporting hospital performance. In 1986, the Pennsylvania Health Care Cost Containment Council was established by the General Assembly and the state governor to help improve the quality, and restrain the costs, of health care. This council developed a series of "Hospital Performance Reports," covering 28 different conditions that are reasons for hospitalization. Reports are divided into regions and are hospital-specific. These reports have been made available on the Internet for several years.

In addition, various sites on the Internet have had an influence on public awareness of outcome measures, including mortality rates. An example is <http://www.HealthGrades.com>, which has been publishing hospital ratings since 1999, as well as other Web sites that have focused a great deal of attention on the quality of health care.

By the end of the 1990s and the early 2000s, another type of information about health care quality was put before the public. As mentioned earlier, several important reports were issued by the IOM, which brought serious quality problems to the public eye. These included the Committee on the Quality of Health Care in America IOM report *To Err Is Human: Building a Safer Health System* (2000), which focused on patient safety issues, and the 2001 report *Crossing the Quality Chasm: A New Health System for the 21st Century*, which focused on how the health care delivery system can be designed to improve the quality of care. In addition, the IOM Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care published *Unequal Treatment Confronting Racial and Ethnic Disparities in Health Care* in 2003. This report focused on the clinical encounter that minority patients experience and the processes of care that have resulted in poor care for minority patients. Since these reports have been made public, a variety of other books, research reports, and broadcasts have focused on these quality problems.

Data availability has increased further in the 21st century and has been characterized by greater information access by individuals and organizations to complex information sources via the Internet. Employers, unions, consumers, and insurance companies began to demand access to data. This change in data availability was significant, making it possible for both professionals and others to compare the performance of various providers. Organizations that have the mission and the capability to analyze and interpret secondary data sources, such as Quality Improvement

Organizations (QIOs), began to focus more on available outcome data to make recommendations for health improvements based on CMS and other large national databases. A further discussion of the role of QIOs is presented later in this text (Chapter 15) and can be found in associated CMS Web sites (e.g., <http://www.cms.gov/QualityImprovementOrgs/>).

Interest in evaluating the quality of care had clearly moved from the professional domain to the public domain. Many physicians felt that the medical profession was under attack from the outside as government and consumers sought to measure and evaluate quality. In addition, governmental, consumer, and industry groups were attempting to measure the value received for their money, to evaluate the relative effectiveness of various treatments, and to compare the quality of care provided by different hospitals and physicians. This interest led to, or paralleled, the development of more sophisticated, complex, and useful models of medical decision making, including computerized decision-making systems, complex treatment protocols for various diseases, and risk-adjusted measures of hospital performance (DesHarnais et al., 1988). As a result, there was an increase in the demand for information about the quality of care, and particularly about the outcomes of care. This interest was manifested in many different ways.

### *Consumers Take a More Active Role*

Consumers began to take a much more active role in their own health care. The women's movement in the 1960s and 1970s was a force that was critical of many medical practices. Consumers began to independently analyze trends in the use and costs of health care services. Various consumer interest groups question effectiveness of various practices. Individual consumers, more knowledgeable about health care, get second opinions, review data on providers, and make decisions concerning treatment options. They have become interested in obtaining accurate and useful data on costs in relationship to the outcomes of care.

### *Hospitals Become Interested in Outcomes*

Hospitals became much more interested in measuring patient outcomes as a defense against public release of mortality data. Hospitals also need information on physician performance for appointment and reappointment decisions. Hospitals often lacked the ability to compare physician performance in terms of outcomes produced or resources utilized. As cost-containment pressures increased alongside concerns for quality,

many hospitals wanted objective information on physician performance as part of decision making on privileges.

Hospitals also want information on both quality and costs for planning and marketing. Many facilities are developing integrated management information systems that provide data on both inputs and outcomes. These information systems can integrate medical records, risk management, quality management, and financial management systems.

In 2009 and 2010, CMS took several actions to reduce hospital payments for hospitalizations that include various complications and for hospitalizations that are unplanned readmissions. These changes in reimbursement policies make it even more important for hospitals to track such problems, due to the negative financial impact that these events will have on hospital revenue.

### *Professional Societies Seek Information on Outcomes*

Specialty societies and certifying boards for various specialties would like information on the outcomes of care for relevant procedures for several reasons. First, such information could promulgate standards for better practice of medicine within their specialty. Outcome data could be used to help evaluate the relative effectiveness of various ways of treating patients, when different treatments are possible and there are wide variations in practices. Second, the information could help set standards for certifying specialty physicians. Information on outcomes could be analyzed and used in designing certification examinations.

While data on patient outcomes could be useful for these endeavors, the various specialty societies are just beginning to understand that this is a very difficult proposition. To gather and use outcome data effectively, it is necessary to do the following:

1. Standardize data reporting requirements
2. Mandate reporting of outcomes data
3. Develop and maintain a patient registry
4. Risk-adjust the data
5. Develop a reporting and benchmarking mechanism

The American College of Surgeons has, in fact, been able to develop such a registry, the National Cancer Data Base, because cancer is reportable

by law and because they have had the authority to certify cancer programs. They have made data reporting one of the requirements of their approval. Most other specialty societies are not currently in a position to take such actions to gather outcomes data from their members.

### *Insurance Companies Want Outcome Measures*

HMOs and preferred provider organizations (PPOs) were prevalent in the 1970s and 1980s. These types of organizations demanded data on costs, use patterns, and practice patterns because such information was crucial in managing care in these systems. It was also essential to evaluate the costs and quality of care given by the providers with whom these insurance organizations contracted. PPO contracts required the contracting agency to exercise care when designing preferred providers. If these providers were producing poor outcomes, marketing of the plan would be impossible, and the PPO could face legal problems. Insurance companies also need such information to market their products successfully in a more competitive environment.

### *Regulators Seek Data on Outcomes*

It also became clear that federal and state programs were paying large amounts of money for treatments and for procedures that might not be the most effective means of caring for patients. By the 1980s, the federal government began to allocate research dollars for “effectiveness research” to learn more about the most effective treatments in areas where great variations in medical practice were discovered. Some outcomes of this federal initiative are as follows:

- Regulatory agencies began independent analyses of trends in the use and costs of health care services.
- Federal initiatives, including some at the Veterans Administration, focused increased attention on quality measurement and improvement, including outcomes.
- Federal regulators became involved. Through the use of peer review organizations, now known as QIOs, the HCFA began to find new uses for data on cost and outcomes of medical care. The federal government used the information for developing changes in payment systems, both for hospitals (Diagnosis Regulated Groups) and for professionals (relative value scales).

- As mentioned earlier, CMS recently has taken several actions that will reduce hospital payments for hospitalizations that include various complications and for hospitalizations that are unplanned readmissions.
- TJC began to examine the possibility of using outcome measurement as part of its accreditation process.
- To provide standardized data sets on costs and outcomes, insurance commissioners and state legislators in many parts of the United States (California, Florida, Iowa, Maine, Massachusetts, New Hampshire, New York, Vermont, Washington, West Virginia, and others) mandated that hospitals report specific data. Several states prescribed the data elements that were required. In many cases, new data elements were mandated beyond the common data set used for billing purposes, at considerable cost to hospitals.

The National Quality Forum (NQF), which functions as a quasi-regulatory agency for certifying quality measures, would like to eventually be able to develop and use patient outcome measures as part of its repertoire. However, due to the difficulty of doing this type of work on developing and certifying patient outcome measures, NQF has tended to focus on process measures.

New requirements for providers to use electronic medical records may eventually make it easier for regulators to gather some of the data they need to measure patient outcomes. If the required data sets include patient assessments and enough patient information to risk-adjust the outcome data, then it may become possible to incorporate outcome measures into the data used by regulators. If this comes about, we may see that the regulators and insurance companies really can begin to “pay for performance.”

### *Individual Employers Want Data on Quality of Care*

A broader concern with quality measurement has developed in industry. Many industries in the United States became highly concerned with methods of measuring and controlling the quality of the products and services they produced. There was a growing focus on using scientific methods and harnessing the energy and creativity of all levels of personnel in an organization. Total quality management (TQM) principles were adopted by many U.S. industries. In many communities, industries using

TQM were represented on hospital boards as well. TQM concepts were introduced into hospital management and eventually began to change the way certain hospitals approach quality.

In addition, unions and industry demanded information on cost, quality, and outcomes as they negotiated contracts. As new benefits were added, it was necessary to analyze whether they were worth what they cost. In some cases, it was necessary to evaluate the performance of providers to decide whether to offer certain plans. Companies that self-insured needed to develop information on users, costs, and outcomes in order to better manage their insurance plans. Local providers that used excessive resources or had consistently poor outcomes could pose a real problem for such plans.

### *Business Groups/Coalitions Become Interested in Outcomes*

Several business coalitions also organized to consider ways to improve health care quality and to control costs. Two examples are the Pacific Business Group on Health and the Leapfrog Group. The Pacific Business Group on Health was founded in 1989 and represents more than 50 large purchasers of health care, with coverage for more than 3 million employees. The coalition identifies health care and business trends, assesses the impact of those trends, and recommends practical steps to advance a common agenda. It works closely with payers, providers, researchers, and others to achieve the highest quality and most cost-effective health care. The Pacific Business Group on Health also works collaboratively with all purchasers in California and with other business coalitions throughout the United States (see <http://pbgh.org>).

The Leapfrog Group is another example. It represents employers, with more than 34 million covered health care consumers in all 50 states. Composed of more than 150 public and private organizations that provide health care benefits, the Leapfrog Group works with medical experts throughout the United States to identify problems and propose solutions that it believes will improve hospital systems that could break down and harm patients. Leapfrog provides important information and solutions for consumers and health care providers (see <http://www.leapfroggroup.org>).

**Table 5-3** illustrates likely uses of performance measures by various stakeholders.

**TABLE 5-3 Performance Measures for Improving Quality**

Consumers	Using performance as selection criteria for providers and plans Using guidelines to evaluate ongoing care Taking a more meaningful role in managing own care
Purchasers	Using quality as selection criteria for providers and plans Displaying quality information to employees and families Devising incentives to get employees to choose quality Developing incentive payment systems to reward provider quality
Health plans	Selecting networks based on quality measures Showing quality results to enrollees and physicians Developing incentive payment systems to reward provider quality Submitting quality measures for review by public
Regulators	Using evidence-based data to develop regulations Assessing quality impact of proposed regulations
Clinicians	Practicing evidence-based medicine Choosing colleagues and services for referrals Submitting quality measures for review by public Using quality methods to improve safety and outcomes
Care delivery systems	Making quality a strategic factor Developing capacity for quality improvement Developing information systems to support evidence-based practice and quality improvement efforts Enabling a culture and systems to support quality and safety

Source: Galvin, R. S., and McGlynn, E. A. 2003. Using performance measurement to drive improvement: A roadmap for change. *Med Care*, 41(1): 148-160.

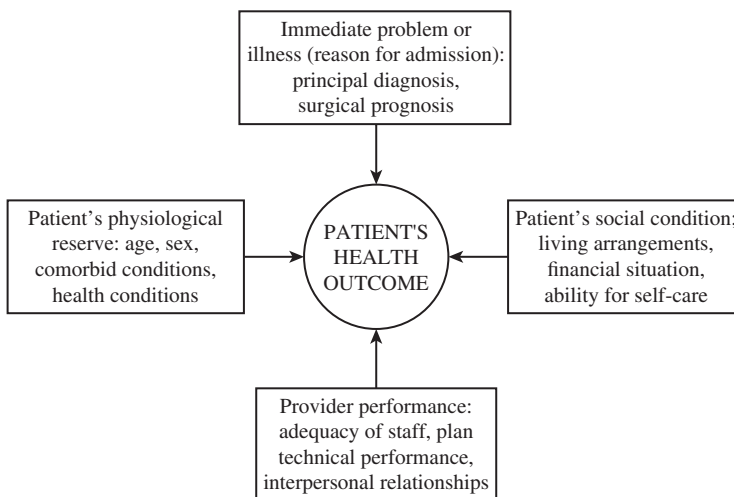
Although comparisons of mortality rates and measures of adverse events across institutions are potentially useful to providers and patients as one way to measure quality of care, such information might be misleading and potentially damaging if misused. This is particularly important when considering how such “report cards” can be used by the government or the public. Such information must be compiled and interpreted correctly. Several studies have demonstrated that raw death rates, without adjustment for differences in case mix and case complexity, lead to misleading comparisons among hospitals, with those hospitals that treat higher risk patients appearing to provide poorer care (Knaus et al., 1986; Moses and Mosteller, 1968; Pollack et al., 1987; Wagner et al., 1986). Death rates must be risk-adjusted and interpreted carefully along with other indicators of quality.

## RISK ADJUSTMENT AND BENCHMARKING OF OUTCOME DATA: DATA REQUIREMENTS AND TECHNIQUES

As explained earlier, differences in outcomes across hospitals (patients' responses to treatment) can be viewed as a result of many different factors that may influence health outcomes (see **Figure 5-1**).

To measure the effect of provider performance on outcomes with accuracy, it is necessary to control for all the other factors. This is clearly not possible, given the existing data sets and measurement tools. However, because "report cards" on providers are going to be produced, it is essential to try to develop as valid an approach as possible for risk adjustment by accounting for as much of the variation that is due to patient characteristics as possible.

Historically, two different approaches have been used to perform risk adjustment of hospital mortality data: hospital-level variables to adjust crude death rates and indirect standardization of patient-level data.



**FIGURE 5-1** Schematic Diagram of Some Factors Related to Health Outcomes

*Source:* Reprinted from DesHarnais, S., Chesney, J. D., Wroblewski, R. T., et al. 1988. The risk-adjusted mortality index: A new measure of hospital performance. *Med Care*, 26(12): 1129-1148, with permission of J.B. Lippincott, © 1988.



Hospital-level data were used in several early studies. In a 1968 study by Roemer et al., hospital-level aggregate measures of patient characteristics (e.g., average age, percentage nonwhite, and percentage of cancer deaths) along with hospital characteristics (e.g., control, occupancy rate, and technology level) were modeled in an attempt to understand whether these proxies for case mix and case complexity were related to the observed differences in crude death rates among hospitals. The authors reasoned that if these hospital-level proxy measures were related to the crude death rates, they could be used to adjust the rates to represent more accurately each hospital's performance.

This early risk adjustment, as the authors acknowledged, was rather crude. They justified the approach by pointing out that detailed patient-level data on diagnosis and severity of illness were not yet available. They acknowledged hospital-level proxy measures to be an indirect approach to estimating case mix and case complexity. The authors stated:

Ideally, one would like to examine the exact diagnosis of each patient admitted and classify it according to a scale of gravity, which might be based on case fatality rates derived from a general literature of clinical investigation. . . . But it is obvious that such a task of calculating average case severity by such an analytic process could present formidable problems of data collection. (Roemer et al., 1968, p. 98)

It certainly would have been difficult in the 1960s, given the limited availability of computers, to model the risks of death for all types of hospital patients using large data sets, even if such information had been available.

### *Using Patient-Level Data for Risk Adjustment*

Because hospital-level data are of limited use as proxies for differences in case mix and case complexity across hospitals, there is no apparent justification for using such data for risk adjustment today. Discharge-level data are now available and are much more sensitive for measuring differences in case mix and case complexity across hospitals. The techniques of using adjusted discharge-level outcome data are documented in early studies such as the National Halothane Study in the 1960s (Moses and Mosteller, 1968), the Stanford Institutional Differences Study in the 1970s (Flood et al., 1982), and work by Luft and Hunt (1986) on the relationship of surgical volume to mortality.

### *Risk Adjustment vs. Severity Adjustment*

Risk adjustment is an empirical approach, using condition-specific risk factors and outcome-specific models. Severity adjustment is quite different insofar as it makes use of one of any number of standardized indexes to assign a severity score to each case. That score is then used as one of the predictors of the outcome of interest, along with other patient characteristics such as age, gender, and so on. The reason severity adjustment is usually inappropriate for adjusting patient health outcomes is really quite simple: Most of the severity systems are the results of models designed to predict resource use rather than patient outcomes. Since there is little or no correlation between resource use and patient outcomes, it is not helpful to use severity measures for risk adjustment. Instead, risk adjustment should be done when looking at health outcomes. This means using an empirical approach with condition-specific risk factors and outcome-specific models. Severity adjustment can be relevant, however, when adjusting resource use data such as costs for comparison purposes.

### *What Procedures Are Used for Performing Risk Adjustment?*

In an article summarizing many of the methodological issues in the risk adjustment of outcome data, Blumberg (1986) described indirect standardization, the principal technique used for risk adjustment of discharge-level data:

Indirect standardization is the method most widely used for risk-adjusted outcome studies. It requires estimates of the expected outcome in a study population, based on the outcome experience of a standard population. To estimate expected outcome, the numbers of cases in the study population with risk-related attributes are multiplied by the probability of the outcome in a standard population with matching attributes. These expected outcomes in the study population are then compared with the observed number having that outcome in the same study population. . . . The first step involves the development and testing of a risk-prediction model, while the second step is a study of the residuals of the observed less the expected outcomes in the study population. (p. 384)

Risk-prediction models can be developed using regression methods (see Chapter 3) that allow control for factors, other than provider performance, that may affect patient outcomes (DesHarnais et al., 1991).

## USES OF RISK-ADJUSTED DATA: WHAT IS BENCHMARKING, AND WHY MIGHT WE WANT TO DO IT?

---

Benchmarking is simply the use of external comparisons to understand how one is doing compared to one's peers and/or one's competitors. Usually one benchmarks outcomes at the service level, or even the diagnosis or DRG level, not at the hospital level. To make meaningful comparisons, the data must first be risk-adjusted, since hospitals differ in the "riskiness" of the patients they treat.

External comparisons allow one to identify areas of strength and weakness. These external comparisons are useful when trying to understand how to prioritize problems within one's own hospital (i.e., to decide which quality issues to address first). Benchmarking of risk-adjusted data can also be used to do self-comparisons over time to see if quality improvement efforts are successful.

### *What Standard Should We Use?*

Benchmarking requires a decision regarding the type of standard that should be used when comparing outcomes across facilities or within a facility over time. Such standards may be developed in three different ways:

1. *Absolute (normative)*: In this approach, results are determined by clinical trials and/or consensus conferences. Standards developed in this manner by academic health centers reflect the ideal practice of medicine, or the best possible outcomes that can be achieved under optimal circumstances (i.e., the most skilled surgeon, the best possible equipment, and the best trained team assisting). Although it is useful to know the theoretical "efficacy" of a treatment, or the best possible result one could achieve, such standards may not be realistic under ordinary circumstances of practice. That is why they are often called "best practices." Clinical trials are the basis of "evidence-based" medicine, but they may be better executed than normally because extra resources are put into execution and control. "Consensus conferences" rest on leading expert opinion but still result from a process that one of our colleagues calls GOBSAT, which stands for Good Old Boys Sitting Around Talking.

2. *Empirical*: In this approach, results are assessed relative to other institutions treating similar patients. Standards developed by comparing oneself to other institutions treating similar patients may be useful to help identify problem areas. If, for example, a hospital is experiencing 20% more unanticipated readmissions than other hospitals when treating a specific type of patient, that could be a signal that some correction is needed. On the other hand, it is possible that the “average” care in the community is poor. Such comparisons are only relative to the level of quality in the institutions used for comparison.
3. *Institutional*: In this approach, results are based on self-comparisons over time. Such standards are often used in conjunction with both quality assurance and CQI. One collects observations of the same phenomenon over time to determine if a process is in control (small random variations) or out of control (major fluctuations). This information uses the institution as its own “control” and can be coupled with the goal of continuously raising standards in the institution. Although this approach is useful, some external comparisons are required to understand how to prioritize problems. One needs such external comparisons (benchmarks) to decide which processes to address first.

### *How to Benchmark Outcomes*

To benchmark outcomes, the following steps should be followed:

1. Using the risk-adjustment models, assign the predicted probability of each relevant adverse event to each case. Consider the following examples:
  - An 82-year-old woman is admitted to the hospital with pneumonia, with secondary diagnoses of cancer of the pancreas and type 2 diabetes. Her probability of death is .591. If discharged alive, her probability of readmission within 30 days is .307.
  - A 36-year-old woman is admitted to the hospital with pneumonia with no secondary diagnoses. Her probability of death is .008. If discharged alive, her probability of readmission within 30 days is .001.

**TABLE 5-4 Predicted and Actual Mortality and Readmissions, and Ratios, for Hospital A**

	Predicted Mortality	Actual Mortality	Ratio (P:A)	Predicted Readmissions	Actual Readmissions	Ratio (P:A)
Pneumonia	23.8	35	0.68	46.9	42	1.12*
All respiratory diseases	70.2	87	0.81	123.3	116	1.06

\*Indicates statistical significance at 0.001

2. Add all of the predicted probabilities for each hospital product line; also add all of the actual events for the same product line. Use these numbers to develop reports for each hospital, comparing predicted frequencies for each category of adverse event to the observed frequencies (see **Table 5-4**).

Note that in this example the hospital has mortality that is significantly higher than predicted, both for pneumonia and for all respiratory diseases, given the risk factors of the patients treated. Readmissions within 30 days of discharge, however, are significantly fewer than predicted for the pneumonia patients and lower than predicted (but not significantly) for all respiratory diseases.

3. Perform statistical tests on the differences between predicted and observed frequencies to determine whether the differences are statistically significant or might merely represent random variations.
4. Develop systems profiles, comparing hospitals using these multiple risk-adjusted measures, similar to the example in **Table 5-5**.

We can use these profiles for a “first cut.” Hospitals with unusually poor (significant) patterns of adverse occurrences should examine medical records and perform peer reviews to determine whether there are problems with the process of care and, if so, whether administrative actions may be required at a system level. In the preceding example, Hospital A might want to examine why its mortality rates for pneumonia and other respiratory diseases are relatively high; Hospital C might want to look at its readmission rates for respiratory diseases other than pneumonia.

**TABLE 5-5** Ratios of Predicted to Actual Values, by Hospital and by Outcome, for Pneumonia and for All Respiratory Diseases in Three Hospitals

	Hospital A	Hospital B	Hospital C
<b>Mortality</b>			
Pneumonia	0.68*	1.09	1.32*
All respiratory diseases	0.81*	0.98	1.03
<b>Readmissions</b>			
Pneumonia	1.12*	1.01	0.99
All respiratory diseases	1.06	1.31*	0.87

\*Indicates statistical significance at .001

### *Recognizing the Limitations of Outcome Measures*

Outcome measures derived from discharge abstracts and billing data have inherent limitations because they lack the context provided by the relevant in-depth clinical information. For example, detailed clinical information allows us to determine time sequences; in the preceding example, did pneumonia or another upper respiratory infection develop while the patient was in the hospital, or was it already present at the time of admission? Patient compliance is an obvious factor for predicting readmissions but is not included in billing data. Furthermore, we cannot assume that data quality is good or uniform across hospitals. Problems with data quality will definitely affect hospital scores on these measures. Poor coding of comorbidities can make a hospital look worse; good coding of complications can also make a hospital look worse. There is no evidence that a hospital that does well on one measure is necessarily doing well on the other measures.

Many of these same symptoms are evident in attempting comparisons across cities and countries. Marshall et al. (2003) report that indicators compare reasonably well between the United Kingdom and the United States, but that some caution is needed because of differing practice cultures. Hussey et al. (2004) compared five industrialized countries on the basis of 21 indicators and found that each country performs best and worst in at least one area of care and that all could show improvement.

## *Problems With the Aggregation of Different Measures of Adverse Events*

### **Are Different Measures Correlated With One Another?**

A valid index of hospital performance must encompass the multiple aspects of hospital care. It may not be possible, either conceptually or technically, to construct a single, all-inclusive index of the quality of hospital care. It is possible, however, to construct several indexes that validly measure important aspects of quality and then to examine the relationships among the various measures to see if they are correlated. If the various indicators are highly correlated, we eventually may be able to construct an overall (unidimensional) quality measure. If they are not correlated, we can conclude that the various components measure distinct dimensions of quality and that the separate measures are all necessary in obtaining a valid impression of a hospital's performance.

For example, a 1991 study analyzed the relationships among three measures that seem to be “intrinsically valid,” in that they clearly are outcomes to be avoided. The three indicators—mortality, unscheduled readmissions, and complications—were adjusted for some of the clinical factors that are predictive of the occurrence of deaths, readmissions, and complications. Risk factors were established empirically within each disease category for each index. The authors demonstrated that hospitals' rankings on the three indexes were not correlated. This result provides some evidence that these different indexes appear to be measuring different dimensions of hospital performance. Thus the three indexes should not be combined into a unidimensional measure of quality, at least not at the hospital level of analysis. Neither should any one measure be used to represent all three aspects of quality (DesHarnais et al., 1991).

One cannot simply choose one hospital-wide measure such as a “death rate” to validly represent a hospital's performance. Neither can one simply add up occurrences of different types of adverse events and then claim to have a unidimensional measure of hospital performance. Those hospitals that rank well in terms of mortality rates do not necessarily do well on the other measures and may have excessive readmissions or complications.

### **Can Different Measures of Adverse Events Be Weighted to Create a Unidimensional Index?**

Can these different types of adverse events be weighted in a meaningful way so that they can be combined and used as a tool to rank hospitals? Probably not. Even after careful risk adjustment and data quality control, one is still left with the problem of how to weight a death in importance relative to a return surgery or an unscheduled readmission. Clearly, they are not of the same importance, and it would not make sense to treat them as if they were.

## **CONCLUSIONS**

---

Quality is something that all health care providers favor, but it is not, as many would like to believe, something that happens without planning and conscientious effort. The outside world is demanding that health care organizations provide care of the highest quality at a reasonable price. Information with which to make assessments of outcome performance in health care is increasingly available. Providers can fight to maintain professional autonomy by trying to push the lay assessors back, or they can take the lead by becoming experts on quality assessment and applying their newfound skills to ongoing operations. They can then educate the public in how to interpret the impact of age, comorbidity, and other risk factors on outcome measures.

The medical profession can educate its members in how to participate in the process of quality improvement, to cooperate with other disciplines and professional groups, to lead the way in analysis and process improvement, and to help develop consensus about what is currently known and what warrants further study. It can go much further in empowering all of its constituents to follow the scientific method at a pragmatic level in all aspects of medicine and in all settings, to the benefit of its consumers. It can move from being on the defensive about consumer-oriented quality and how it is measured toward being its primary advocate.

### **Cross-References to the Companion Casebook**

(McLaughlin, C. P., Johnson, J. K., and Sollecito, W. A. [Eds.]. 2012. *Implementing Continuous Quality Improvement in Health Care: A Global Casebook*. Sudbury, MA: Jones & Bartlett Learning.)



Case Study Number	Case Study Title	Case Study Authors
2	<b>Holtz Children's Hospital: Reducing Central Line Infections</b>	Gwenn E. McLaughlin
13	<b>The Folic Acid Fortification Decision Bounces Around the World</b>	Curtis P. McLaughlin and Craig D. McLaughlin

## REFERENCES

- Blumberg, M. 1986. Risk-adjusting health care outcomes: A methodological review. *Med Care Rev*, 43: 351–393.
- Chassin, M. R., Hannan, E. L., and DeBuono, B. A. 1996. Benefits and hazards of reporting medical outcomes publicly. *N Eng J Med*, 334: 394–398.
- DesHarnais, S., Chesney, J. D., Wroblewski, R. T., et al. 1988. The risk-adjusted mortality index: A new measure of hospital performance. *Med Care*, 26: 1129–1148.
- DesHarnais S., McMahon, L. F., Jr., and Wroblewski, R. 1991. Measuring outcomes of hospital care using multiple risk adjusted indexes. *Health Serv Res*, 26: 425–445.
- Donabedian, A. 1980. The definition of quality and approaches to its assessment. In *Explorations in Quality Assessment and Monitoring* (Vol. 1, pp 95–99). Ann Arbor, MI: Health Administration Press.
- Donabedian, A. 1982. *The Criteria and Standards of Quality*. Ann Arbor, MI: Health Administration Press.
- Donabedian, A. 1986. Criteria and standards for quality assessment and monitoring. *Qual Rev Bull*, 14(3): 99–108.
- Flood, A., Scott, W. R., Ewy, W., et al. 1982. Effectiveness in professional organizations: The impact of surgeons and surgical staff organizations on the quality of care in hospitals. *Health Serv Res*, 17: 341–366.
- Galvin, R. S., and McGlynn, E. A. 2003. Using performance measurement to drive improvement: A roadmap for change. *Med Care*, 41(1): I48–I60.
- Hussey, P. S., Anderson, G. F., Osborn, R., et al. 2004. How does the quality of care compare in five countries? *Health Aff*, 23(3): 89–99.
- Institute of Medicine. 2000. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academies Press.
- Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press.
- Institute of Medicine. 2003. *Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care: Unequal Treatment Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press.

- Knaus, W., Draper, E. A., Wagner, D. P., et al. 1986. An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med*, 104: 410–418.
- Luft, H., and Hunt, S. 1986. Evaluating individual hospital quality through outcome statistics. *JAMA*, 255: 2780–2786.
- Marshall, M. N., Shekelle, P. G., McGlynn, E. A., et al. 2003. Can health care indicators be transferred between countries? *Qual Safety Health Care*, 12: 8–12.
- Moses, L. E., and Mosteller, F. 1968. Institutional differences in postoperative death rates: Commentary on some of the findings of the National Halothane Study. *JAMA*, 202: 492–494.
- Pollack, M., Ruttimann, U. E., Getson, P. R. 1987. Accurate prediction of the outcome of pediatric intensive care: A new quantitative method. *NEJM*, 316: 134–139.
- Roemer, A., Moustafa, T., and Hopkins, C. E. 1968. A proposed hospital quality index: Hospital death rates adjusted for case severity. *Health Serv Res*, 3(2): 96–118.
- U.S. Office of Technology Assessment (OTA). 1988. The quality of medical care: Information for consumers. OTA-H-386. Washington, DC: U.S. Government Printing Office.
- Wagner, D. P., Knaus, W. A., and Draper, E. A. 1986. The case for adjusting hospital death rates for severity of illness. *Health Aff*, 5(2): 148–153.