

GETTING STARTED AND DEFINING TERMS

Lisa I. Iezzoni

The first step in risk adjustment is defining terms. Many clinicians erroneously assume that other health care professionals define terms such as *risk* and *case mix* the same as they do. In the 1980s, non-clinicians in the health care delivery system and health policy arenas adopted similar terms, prompted by the introduction of Medicare's inpatient prospective payment system based on Diagnosis-Related Groups (DRGs). *Risk adjustment* joined other poorly defined but oft-used words and phrases, such as *complexity*, *severity*, *intensity*, and *health status*, used not only by clinicians and researchers but also by others involved in health care delivery. Researchers, payers, policymakers, managers, regulators, quality measurement and improvement experts, performance profilers, and health insurance actuaries assign fundamentally different meanings to these terms. Different definitions of these words complicate discussions about critical health care issues.

Throughout this book, we use *risk adjustment* broadly to mean accounting for patient-related factors before examining outcomes of care, regardless of the context. To define and devise appropriate risk-adjustment strategies, however, we must be specific. We start with four major questions:

1. Risk of what outcome?
2. Over what time frame?
3. For what population?
4. For what purpose?

Other questions soon follow, such as: Considering what risk factors? Using which data source? Employing which analytic methods? This chapter sketches answers to these questions, and the remaining chapters fill in more details.

We cannot exhaustively review risk adjustment methods; the field is too large and grows continuously. Furthermore, details of risk adjusters change as developers revise and update their methods. Some risk adjustment methods are widely used for highly public purposes, such as establishing hospital reimbursement levels, setting capitated payments to health plans, or producing provider performance profiles posted on websites. Many more have been developed for disease-specific research projects or quality measurement initiatives. Numerous

commercial methods, primarily intended for managerial or administrative purposes, are also available for use; their complete logic is rarely open to external scrutiny. Therefore, we use examples from seminal studies, as well as our own research, to examine risk adjustment methods that may or may not be the most up-to-date versions but nevertheless illustrate critical points that will always be relevant. As described in Chapter 8, development of risk adjusters *de novo* is complicated and often frustrating. We generally recommend taking methods “off the shelf” if their attributes match a project’s goals (or the policy context) reasonably well.

Most of our examples involve risk adjusters that have been reported in the peer-reviewed literature or are described on public websites (e.g., methods the Centers for Medicare & Medicaid Services [CMS] uses to risk-adjust hospital mortality rates for the Hospital Compare performance measures). The glossary at the end of this book includes a selection of major risk adjustment methods (and their acronyms). Readers interested in using specific methods should not rely on descriptions provided here, as they may be out of date. When choosing a risk adjustment method, users should seek one with transparent logic—i.e., information about the method’s inner workings is completely open to outside examination. Only with complete access to the algorithm can users judge whether a particular risk adjustment method is fully appropriate for the intended purpose and understand completely the results it generates.

Risk of What?

As stated in Chapter 1, the notion of risk permeates daily life. The word portends negative consequences; the *American Heritage Dictionary* (2000) defines it as “the possibility of suffering harm or loss; danger” and “a factor, thing, element, or course involving uncertain danger; a hazard.” However, the term *risk adjustment* is meaningless without identifying the outcome being risked. Answers fall broadly into three camps:

- Clinical outcomes of care, such as deaths, complications, physical functional status, and mental health
- Resources used or required, such as costs for a hospitalization, a year of care, or lengths of hospital stays
- Patient-centered outcomes, such as patients’ reports that care met their preferences and expectations

Two brief vignettes demonstrate differences among definitions of risk. An adenocarcinoma of the lung was detected on a routine chest radiograph of Mr. A taken during an employment physical examination. Radiographically,

the tumor looked like an isolated lung nodule. A needle biopsy of the nodule identified lung cancer. Mr. A underwent further diagnostic testing, including a positron emission tomography (PET) scan and a magnetic resonance imaging (MRI) scan of his head, to determine whether the cancer had metastasized elsewhere. Finding no metastases, Mr. A's oncologist performed a lobectomy, a major operation to remove the tumor; Mr. A experienced no complications. Mr. A, a nonsmoker, is otherwise healthy, and his physicians believe he has a high likelihood of a surgical cure.

In contrast, Mr. B had a widely metastatic adenocarcinoma of the lung. He had exhausted current aggressive therapies and desired to be kept comfortable as he neared death. He requested "comfort measures only," declining even routine blood tests. Mr. B also wanted "do not resuscitate" (DNR) and "do not intubate" (DNI) status: If his respirations or heart stopped, clinicians would not intervene. At home under hospice care, he was placed on round-the-clock pain medications and other drugs to control uncomfortable symptoms; most of the drugs were delivered through patches placed on his skin. With family and friends at his bedside, Mr. B died without evident pain or distress.

Both scenarios involve an adenocarcinoma of the lung. Mr. A had a high risk of incurring high costs (for extensive diagnostic workup and major surgery), but he had a low risk of imminent death. In contrast, Mr. B's care was relatively inexpensive (home-based pain control), but he had a high risk of dying soon. Mr. A received aggressive treatment aiming for a cure; Mr. B desired to maintain comfort without intensive intervention. Mr. A and Mr. B chose clinicians and institutions that best met their personal goals and clinical needs. Mr. A sought care at a major academic medical center, whereas Mr. B obtained care in his community, close to family and friends. Thus, comparisons of cost and mortality outcomes for lung cancer patients across different hospitals must account for differences in their patient mixes: Some institutions treat more Mr. As, whereas others see more Mr. Bs. Mr. As have high risks of incurring costs and low risks of dying, whereas Mr. Bs present the opposite scenario.

Existing risk adjustment methods typically assess risks differently (i.e., use different risk factors or weight the same risk factors differently) depending on the targeted outcome (i.e., clinical outcome, resource consumption, or patient-centered outcome). Some risk adjustment methods come in multiple versions, each version calibrated to predict a specific outcome. For example, some vendors of hospital-based risk adjustment methods have different versions for predicting hospital costs versus predicting in-hospital mortality. When using risk adjusters from a family of methods (with these different versions), users must choose the method designed for their specific outcome of interest. As suggested by the scenarios of Mr. A and Mr. B, risk adjusters designed to predict costs generally do less well at predicting deaths than do methods derived specifically for mortality analyses.

Over What Time Frame?

Risks must be framed within specific time windows. As an extreme example, calculating the risk of death is moot if the time window involves lifetimes; everybody faces a 100 percent risk of dying. Mr. A may not die for decades, whereas Mr. B's death occurred in a few weeks. Similarly, costs are typically measured within explicit time frames, such as a hospitalization or a year of care. Time frames thus clarify the outcome of interest and suggest which risk factors are most important. Chapter 4 addresses in detail the issues raised by different time frames.

The time frame generally determines the data sources for risk factors and vice versa. For example, as described in Chapter 5, many studies rely on computerized hospital discharge abstracts; their diagnosis and procedure codes represent the entire hospitalization, although the recent addition of “present on admission” flags may narrow this time frame. Risk adjusters modeling hospital outcomes to identify quality shortfalls aim to capture risk factors that predate care. Otherwise, serious clinical findings (potential risk factors) could become confounded or confused with substandard care. For quality assessment, the time window for extracting risk factors helps determine the *attributional validity* of the risk-adjusted outcomes information—the likelihood that poor risk-adjusted outcomes reflect poor care rather than high patient risks (see Chapter 9). The attributional validity of narrower pretreatment time windows is presumably superior to that of wider windows.

Risk adjusters predicting costs over a year, such as those intended to set capitated payment levels for managed care health plans, often have two versions, both typically derived from computerized claims or encounter records. *Concurrent* models use data from a particular year to predict costs for that same year, whereas *prospective* models predict costs for the following year. Obviously, predicting future costs is more difficult than retrospectively modeling concurrent costs. The preferred approach depends on the purpose.

Perceptions of outcomes can change substantially with even small shifts in the window of observation. An excellent example is the short-lived Cleveland Health Quality Choice (CHQC) program, a voluntary coalition of businesses, hospitals, and physicians, which involved gathering detailed clinical data from medical records to risk-adjust hospital outcomes for general medical, surgical, and obstetrical patients. Initial CHQC data were released privately to hospitals for internal improvement activities; the first public report was issued in April 1993. The program gathered data from 1991 through 1997, during which Cleveland's absolute, risk-adjusted in-hospital mortality rates declined by up to 4.8 percent (Baker et al. 2002a). Had quality of care actually improved that significantly? Probably not. When Baker and colleagues (2002a) looked instead at mortality 30 days after hospital admission (a fixed time window), the rates did not change significantly. Between 1991 and 1997, deaths had shifted from Cleveland hospitals to other settings soon after discharge.¹

For What Population?

The US population is remarkably diverse. Many of us are exquisitely aware of this diversity on a daily basis because of the effect our visible attributes have on our interactions with others. From the earliest moments of self-recognition, we learn our age, sex, skin color, and language and perceive the immediate world around us. As we grow, our interpretations of these basic dimensions modulate, and their meanings expand. Age becomes generation; sex becomes gender (and gender identification, with all its complexities); skin color becomes race; language becomes ethnicity; and the immediate world, with its myriad complexities, becomes culture. We develop other, sometimes shifting associations with economic class, educational attainment, religion, occupational group, disability, political ideology, and other identities.

These many dimensions, alone or in combination, help delineate populations or subpopulations that have different risks for various health-related outcomes (see Chapter 3). Some distinctions are self-evident: Children, on average, face lower risks of imminent death than do persons in extreme old age. Women and men have different risks for certain diseases. As described in Chapter 3, troubling risks arise not from intrinsic individual factors (i.e., not from biological or physiological differences), but from disparities in the way people are treated in our health care system or society at large because of their characteristics.

The population of interest helps determine the range of risk factors required for assessing the specified outcome within the pertinent time frame. For example, when examining intensive care unit (ICU) mortality rates, the relevant physiologic parameters vary among neonates, children, and adults, although immediate acute findings are particularly relevant. Depending on how populations are defined, some outcomes are more pertinent than others. This book contains four chapters on specific populations: children (Chapter 13); persons with mental health conditions (Chapter 14); persons with disabilities (Chapter 15); and individuals receiving long-term care in institutional and home-based settings (Chapter 16). Especially relevant outcomes vary across these populations. Although children experience similar life-and-death outcomes as adults, albeit at different rates, specific functional outcomes differ (e.g., school performance and developmental milestones for children; productive employment for working-age adults). Important outcomes for persons with psychiatric disorders emphasize mental and emotional health and ability to perform routine social roles. For persons with disabilities and long-term care populations, functional abilities and performance of daily activities are key outcomes.

The target population reflects the underlying purpose of the developers of the risk adjustment method. For example, the designers of the Pediatric Risk of Mortality Score (PRISM) were explicitly interested in children treated in ICUs (Pollack, Ruttimann, and Getson 1987), while the developers of the

Acute Physiology and Chronic Health Evaluation (APACHE) focused on adult ICU patients (Knaus et al. 1981). The purpose generally determines the population and thus typically indicates the relevant data source. The developers of the Chronic Illness and Disability Payment System (CDPS), for example, wanted to create a method of capitating payments specifically for Medicaid recipients (Kronick, Zhou, and Dreyfus 1995). By using Medicaid databases, they created a risk adjuster calibrated to impoverished persons, primarily women and their children and persons with disabilities.

For What Purpose?

Answers to the three previous questions (risk of what, over what time frame, and for what population) are driven by the purpose of risk adjustment. As described in Chapter 1, the underlying motivation of risk adjustment is comparison: contrasting outcomes or performance for individual patients, groups of patients, or populations to those of their counterparts. The following are examples of potential purposes:

- To set payment levels for individual patients (e.g., DRGs for acute care hospitalizations) or health plan enrollees (e.g., capitation payments)
- To encourage providers or health plans to treat or accept high-cost or potentially high-risk patients
- To compare efficiency and costs of care across providers or health plans
- To compare clinical or patient-centered outcomes across providers or health plans
- To produce public report cards about performance of individual providers, as on CMS's Hospital Compare website
- To compare patient outcomes across physicians or services in an individual practice or institutional setting to guide and monitor quality improvement

The purpose dictates how well the risk adjuster must perform to succeed (i.e., to produce valid comparisons; see Chapter 9). For example, methods designed to predict costs over one year rely on administrative data, which are often messy and contain limited clinical information (see Chapter 5); these risk adjusters typically explain less than 25 percent of the variation in future costs. Nevertheless, this performance is far superior to adjustments using only demographic information (e.g., age, sex), and it meets the needs of important purchasers such as Medicare and Medicaid.

Another purpose for risk adjustment is to motivate quality improvement. Without this adjustment, clinicians or institutions with poor outcomes

could argue that they are treated unfairly: “Our patients are sicker; that’s why our results are worse.” Most clinicians will not believe and act on results if they do not consider the risk adjuster to be clinically credible. For a risk adjuster to be perceived as such, additional data collection and in-depth review of the clinical logic underlying the risk adjustment model by the participating clinicians may be required (see Chapter 8).

No risk adjuster is perfect. As described in Chapter 3, adjusting for all patient characteristics is neither necessary nor possible. Therefore, efforts shift to identifying risk factors that are sufficiently valid for the explicit purpose. Statistical measures of model performance alone (e.g., percentage of variation explained; see Chapter 10) do not determine validity. Such measures reveal little about whether systematic errors in predictions occur for selected sub-populations or whether important risk factors are included appropriately.

For some purposes, ethical concerns raise questions about whether and how to risk-adjust. Such situations arise when persons with certain attributes (e.g., gender, race, socioeconomic status) that might be potential risk factors for a given outcome simultaneously face the likelihood of receiving substandard care because of those attributes. One example involves performance reports that compare rates of routine screening tests or preventive services for enrollees of different health plans. Outcomes (here, technically, processes of care²) that depend on patients’ actions (e.g., having a mammogram, having an infant immunized) raise special concerns. Education, motivation, wherewithal (e.g., transportation, child care, time off from work), care and outcome preferences, cultural concerns, and a host of other factors affect whether patients take these actions. Different health plans and providers see different mixes of patients along these critical dimensions. Therefore, from a purist’s perspective, risk adjustment is indicated. However, evidence suggests that racial and ethnic minorities and persons with low socioeconomic status obtain preventive services at lower rates, probably because of a complex mix of factors but also potentially as a result of discriminatory attitudes. As Romano (2000, 978) observed:

Before instituting case-mix adjustment of health plan or provider performance measures, we must consider both the hidden assumptions and the potential consequences. One assumption is that persons of lower socioeconomic status *inherently* use preventive services less than persons of higher socioeconomic status. If culturally sensitive, readily accessible systems of care can eliminate or substantially reduce sociodemographic disparities . . . then adjusting for case mix would implicitly “excuse” health plans for failing to implement disparity-reducing innovations. . . . [Plans might also find that] it is easier to boost [their] scores by focusing on better educated, easier-to-reach members. A related implication is that we should accept lower performance, or set lower performance targets, for plans that enroll diverse populations.

Therefore, risk adjusting for these sociodemographic attributes seems inappropriate given the ultimate purpose of using outcomes data to motivate improvement for *all* patients. Risk stratification, described in Chapter 3, is a simple solution that could also yield useful insight about how different sub-populations fare.

Even when applied to the same data set, different risk adjusters can produce different answers about the outcome of interest, and it is sometimes impossible to determine which method's answers are "right" (i.e., which method produces results that best represent the underlying truth). This difficulty could complicate decisions about which risk adjuster best meets the purpose. For example, in 2006, Massachusetts established its Health Care Quality and Cost Council, aiming to set strategies for improving health care quality while controlling costs and eliminating racial and ethnic disparities in care. To motivate Massachusetts hospitals to improve their quality of care, the Council issued a proposal to develop and publicly report overall hospital mortality rates (i.e., mortality rates reflecting care across the entire hospital rather than mortality rates for specific diagnoses or procedures).

Recognizing the analytic complexity of risk adjusting and producing these hospital-level mortality rates, the Council asked the Massachusetts Division of Health Care Finance and Policy (DHCFP) to evaluate methods for this purpose. In November 2008, DHCFP issued a call for methods of producing hospital-wide mortality measures that could be used for quality improvement and public reporting. Five commercial vendors responded: 3M Health Information Systems (3M), using its All Patient Refined Diagnosis-Related Groups (APR-DRGs); the Dr. Foster Unit at Imperial College London (Dr. Foster); Thomson Reuters; University HealthSystem Consortium (UHC); and Premier. The latter two, UHC and Premier, decided to collaborate to develop a new UHC-Premier method. All vendors received identical standard abstract information on 2,528,624 discharges from Massachusetts acute care hospitals from October 1, 2004, through September 30, 2007. The vendors applied their risk adjustment algorithms to predict probabilities of in-hospital death for each discharge and hospital level observed and expected mortality rates (Shahian et al. 2010).

Despite using the same data, the four risk adjustment methods produced different results (Shahian et al. 2010). One explanation may be their use of different subsets of discharges, despite the ostensible purpose of looking at hospital-wide mortality. As shown in Exhibit 2.1, 3M considered 95 percent of the total discharges, while UHC-Premier analyzed only 28 percent of discharges. All four risk adjustment algorithms were applied to only 22 percent of the hospital discharges. The methods also considered patients with differing characteristics (Exhibit 2.1), including variations in average age and type of case (e.g., childbirth, neonates, mental health conditions). Each method calculated

Characteristic	Risk adjustment method				All discharges	Discharges used by all four methods
	3M	Dr. Foster	Thomson Reuters	UHC-Premier		
Hospitals						
<i>No. included</i>	83	82	83	81	83	81
<i>% of total</i>	100	99	100	98	100	98
Discharges						
<i>No. included</i>	2,406,881	1,072,918	2,048,377	716,315	2,528,624	567,784
<i>Median no. per hospital</i>	21,926	10,140	18,747	7,114	23,428	5,870
<i>% of total</i>	95	42	81	28	100	22
Mean length of stay (days)	5.8	6.1	5.7	7.0	5.8	6.5
Inpatient mortality rate (%)	2.0	4.0	2.4	5.9	2.1	6.2
Mean age (years)	50 (28)*	52 (33)	56 (24)	66 (23)	51 (28)	69 (20)
Selected diagnoses (%)**						
<i>Respiratory</i>	9.7	17.1	11.2	25.0	9.9	26.2
<i>Circulatory</i>	15.1	25.5	17.1	25.9	15.5	30.4
<i>Pregnancy or childbirth</i>	10.4	< 0.1	12.2	< 0.1	10.0	< 0.1
<i>Newborn or neonate</i>	10.0	21.7	< 0.1	3.3	9.8	0
<i>Mental disorders</i>	4.7	< 0.1	5.1	0	4.6	0

*Mean age in years (standard deviation)

**Five conditions from among the top ten major diagnostic categories that accounted for more than 5 percent of discharges included by at least one risk adjustment method

Source: Adapted from Shahian et al. (2010, 2533).

a probability of death for each discharge it analyzed. Exhibit 2.2 shows the Pearson correlation coefficients from 2007 data comparing these predicted probabilities of death between pairs of methods for the roughly 22 percent of discharges considered by all four methods.³

Given that a primary purpose of these data was public reporting, differences across the methods with regard to hospitals' mortality rates (higher or lower than expected)—after accounting for their populations' risk factors—were especially sobering. For each of the four risk adjustment methods and for each of the three years, Exhibit 2.3 shows the number of hospitals considered to have either higher- or lower-than-expected mortality rates. As shown, the

EXHIBIT 2.1
Hospitals and Patient Characteristics by Risk Adjustment Method and All Hospital Discharges in Massachusetts 2005–2007

EXHIBIT 2.2
Correlation
Coefficients
for Agreement
Among
Discharge-Level
Estimates of
In-Hospital Risk
of Death*

Method	Correlation coefficient			
	3M	Dr. Foster	Thomson Reuters	UHC-Premier
3M	1.00	0.61	0.58	0.67
Dr. Foster		1.00	0.48	0.59
Thomson Reuters			1.00	0.59

*Correlation coefficients from 2007 Massachusetts hospital data

Source: Adapted from Shahian et al. (2010, 2534).

methods differed substantially in the numbers of hospitals considered to be outliers, especially in 2005. According to Shahian et al. (2010, 2534):

Kappa statistics [see Chapter 9] indicated poor-to-substantial agreement between methods in classifying hospital mortality performance, depending on the year and method pairs. In fiscal year 2007, kappa statistics for agreement between methods in designating higher-than-expected outliers ranged from -20.04 (UHC-Premier and 3M) to 0.39 (Thomson Reuters and Dr. Foster). For some individual hospitals, categorizations varied widely and in some cases were completely discordant. For instance, in fiscal year 2006, of 28 hospitals designated as having higher-than-expected hospital-wide mortality by one method, 12 were simultaneously classified as having lower-than-expected mortality by other methods (6 by one method, 3 by two methods, and 3 by three methods).

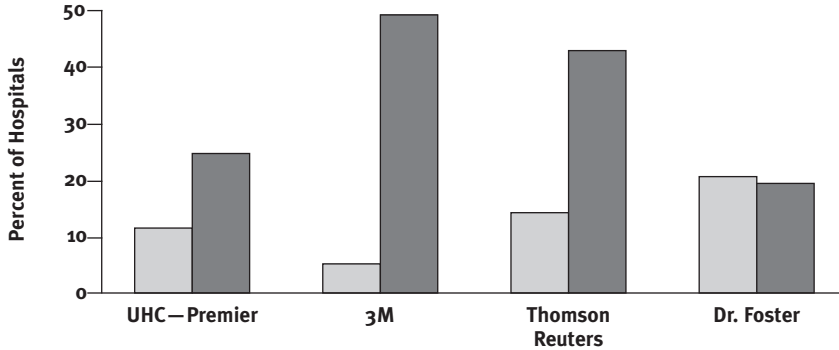
Exhibit 2.4 shows plots, by pairs of risk adjustment methods, of hospitals' standardized mortality ratios produced by each method and multiplied by 100 (see Shahian et al. [2010, 2536] for details about computations). Especially striking were the differences between the rates produced by the Thomson Reuters method and those produced by the other vendors' methods for a facility labeled Hospital C, likely because the Thomson Reuters method used only 3 percent of Hospital C's discharges to calculate the standardized mortality ratios while the other methods used at least 30 percent. The 3 percent used by Thomson Reuters had a much larger number of high-mortality diagnoses (e.g., respiratory diseases and cancers) and a lower number of low-mortality diagnoses (e.g., childbirth) than did discharges overall. Thus, the 3 percent included in Thomson Reuter's calculations for Hospital C had a mortality rate of 59.8 percent, while Hospital C's overall mortality rate was just 2.2 percent.

Findings from this Massachusetts project prompted Shahian and colleagues (2010) to question which method's results were right or the truest reflection of hospital quality (as proxied by hospital-wide mortality rates). Given that the purpose of producing these figures was to provide insight into hospital quality, their question was central to the study, but one that they could not answer. They did not have an independent, "gold standard" indicator of

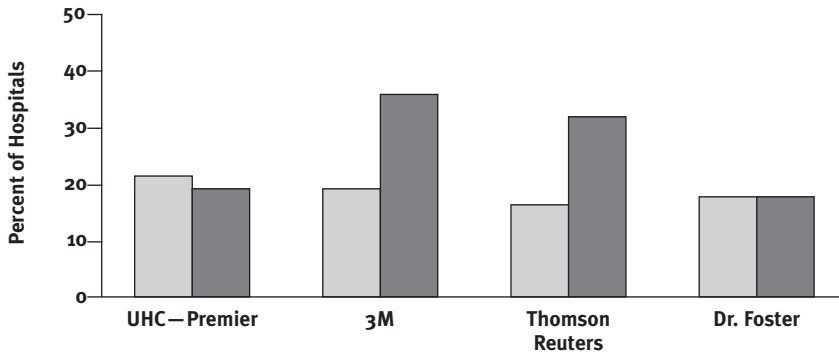
EXHIBIT 2.3
 Percentages of Massachusetts Hospitals with Mortality Rates Higher or Lower than Expected According to Four Risk Adjustment Methods: 2005–2007

□ Mortality rates lower than expected ■ Mortality rates higher than expected

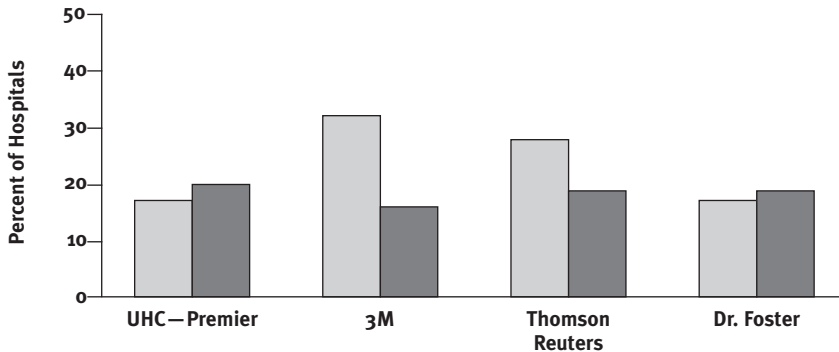
A Fiscal Year 2005 (77 hospitals)



B Fiscal Year 2006 (78 hospitals)

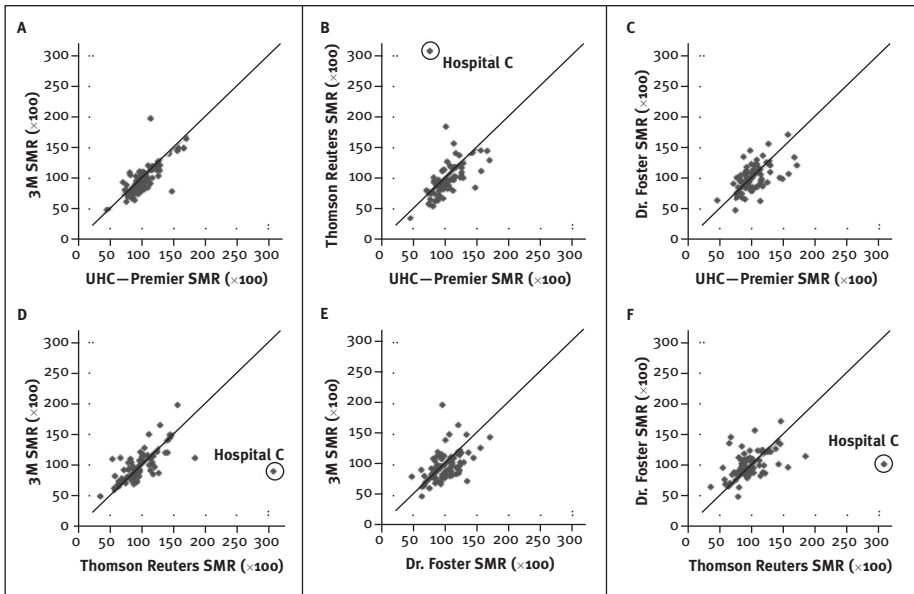


C Fiscal Year 2007 (75 hospitals)



Source: Reprinted from Shahian et al. (2010, 2537); copyright held by Massachusetts Medical Society. Used with permission.

EXHIBIT 2.4
Comparison of
Hospital-Level
Standardized
Mortality Rates
According
to Four Risk
Adjustment
Methods: 2007



Source: Reprinted from Shahian et al. (2010, 2536); copyright held by Massachusetts Medical Society. Used with permission.

hospital quality against which to compare the findings of the risk adjustment study. The researchers concluded (Shahian et al. 2010, 2534–35):

The goal of assessing hospital-wide mortality rates is to make inferences about the relative quality of care among hospitals. Proponents believe that hospital-wide mortality metrics provide useful warning flags about problems with the quality of inpatient care, aid consumers in choosing a hospital, and help provide a focus for hospital quality-improvement activities. . . .

The four commercially available methods for assessing hospital-wide mortality that we studied are marketed to hospitals to support internal quality-improvement activities. However, their implications are even more important, and the corresponding need for methodological accuracy is greater, when such measures are used for broader initiatives, such as public reporting or performance-based purchasing. We found that estimates of hospital-wide mortality could vary, sometimes widely, among methods, which consequently leads to different inferences regarding the quality of hospital care.

...

Differences in categorizing performance on the basis of hospital-wide mortality rates raise the inevitable question of which method best identifies potential quality problems. Our study could not address that question, since an observable benchmark for overall hospital quality does not exist. . . . This disagreement suggests that all methods are not reflecting the same underlying construct, although it is possible that one method might perform better than the others in estimating the quality of hospital care.

Additional Considerations

Answering the four major questions outlined earlier is simply the beginning; many important issues remain. The most crucial practical consideration is the data source. Will the risk adjuster rely on standard, coded administrative data (see Chapter 5); clinical information from medical records, which should become increasingly electronic in coming years (see Chapter 6); or direct responses from patients (see Chapter 7)? The nature of the database shapes the design of the risk adjustment method: With large data sets, analysts can develop and test risk adjusters empirically, whereas without such data, measures must rely on clinical judgment. As described in Chapter 8, the most statistically and conceptually robust risk adjusters generally result from interactions between clinicians and statistical modeling.

The data also delimit the range of candidate risk factors (see Chapter 3). One important distinction, especially when the purpose of the analysis is to predict costs, is whether the risk adjuster considers procedure use. Methods designed to predict short-term costs (e.g., DRGs) rely heavily on procedural information, particularly the presence and type of major surgery: The costs of operations generally overwhelm the costs of medical or recuperative care. Not surprisingly, DRGs overall are poor risk adjusters for hospital mortality, although they perform slightly better within surgical DRGs (Hofer and Hayward 1996). Because the use of many procedures is highly discretionary, risk adjusters targeting clinical outcomes generally eschew procedures in rating risk. As pharmacy data become increasingly available, some risk adjusters are using them to proxy disease burden.

Finally, selection of appropriate and reasonable analytic techniques raises important questions. As described in Chapter 11, risk adjustment is often used to examine the results of observational studies in which patients are not randomly assigned to various treatments or care plans. In many cases, sample sizes are small and there is no single right way to analyze the data. Nevertheless, analytic choices can carry important implications, as suggested in chapters 10 through 12. Consideration of these implications is essential to meaningful interpretation of risk-adjusted outcomes information.

Even after risk adjustment, questions remain about what comparative outcomes information really means. For example, even with optimal risk adjustment, do risk-adjusted mortality rates provide meaningful clues about hospital quality (Pitches, Mohammed, and Lilford 2007; Lilford and Pronovost 2010; Black 2010)? Answering this question in a meaningful fashion may prove even more vexing than designing the risk adjustment methodology. Nonetheless, as Nightingale and Codman said (see Chapter 1), this question about quality highlights the ultimate purpose of gathering and analyzing the data—to motivate and guide improvements.

Notes

1. CHQC disbanded in July 1999 because hospitals affiliated with Cleveland Clinic refused to voluntarily submit their data. Since the outset, Cleveland Clinic officials complained that the CHQC risk adjustment ignored special characteristics of their patient population (Vogel and Topol 1996). They also protested against paying \$2 million annually for data collection. As further justification for its withdrawal, the Cleveland Clinic noted that the CHQC data were not being used.
2. Performance measures are often sorted into two types: outcomes and processes. The rationale for risk adjustment of outcomes (i.e., how patients fare) is generally clear. Process measures (i.e., what is done for patients) may also warrant risk adjustment, although the conceptualization of the measures dictates the extent of adjustment required. Some observers increasingly blur the semantics distinguishing outcome from process measures. For example, is obtaining a mammogram an outcome or process of care? Many quintessential process measures build in explicit information about patient characteristics that are essentially risk factors for obtaining the service. For example, the use of beta-blockers after a heart attack is a widely accepted process measure, but with the stipulation that patients not have certain contraindications to receiving the beta-blockers. Risk adjustment for comparing beta-blocker use across providers does not need to control for contraindications to those drugs because persons with those conditions are eliminated from consideration.
3. Exhibit 2.2 shows Pearson correlation coefficients derived from 2007 data. Figures from 2005 and 2006 data were comparable (Shahian et al. 2010, 2534).