Interpreting Hospital Mortality Data

How Can We Proceed?

Katherine L. Kahn, MD; Robert H. Brook, MD, ScD; David Draper, PhD; Emmett B. Keeler, PhD;
Lisa V. Rubenstein, MD, MSPH; William H. Rogers, PhD; Jacqueline Kosecoff, PhD

DATA about whether patients live or die following short-term-care hospitalization can serve at least two purposes: (1) to determine if hospital performance has improved or deteriorated over time and (2) to determine if performance differs across hospitals at one given time. Hospital mortality figures have been used in this manner primarily because such data are readily available, easy to understand, and potentially important to the public. However, if mortality data are to be used for these purposes, they must reflect actual differences in the quality of care, not other factors, such as patient sickness at admission.

We have been studying hospital mortality and other outcomes as part of a clinically detailed national study to assess the impact of the diagnosis related group (DRG)-based prospective payment system (PPS) on the quality of care for hospitalized Medicare patients. In this report we document and justify some of the design decisions we made in planning the study. We believe these decisions are relevant to understanding how hospital mortality varies both over time and across hospitals at the same point in time. In addition, justification for the design decisions provides a framework for understanding the steps that need to be taken to assure that the outcome data are useful.

To understand how mortality can be used as a meaningful measure of quality of care, at least three basic issues must be addressed:

1. Should we look at disease-specific or across-disease outcomes and mortality rates?
2. How should we adjust outcomes for sickness at admission?
3. How should we sample patients and hospitals?

USING DISEASE-SPECIFIC MORTALITY RATES

Mortality following short-term hospitalization is a meaningful measure of quality only if lower rates are associated with better medical care. For patients with treatable diseases (eg, pneumonias, congestive heart failure), a reduced mortality rate could reflect better care. By contrast, for patients with terminal diseases (eg, end-stage cancer), death may be a sought-after and inevitable outcome, and a higher mortality rate thus may not reflect poor quality of care. For this latter group of diseases, the functional status and quality of the dying experience may be more appropriate measures of quality of care.

To develop a valid measure of the quality of hospital care based on mortality data, we must combine evidence in the literature about the effectiveness of care with clinical judgment, and we must select diseases for which medical care is efficacious in postponing death. Studying the mortality associated with highly prevalent diseases is also useful, because what is learned will be applicable to a large number of patients.

Selecting diseases with relatively high mortality rates helps make it possible to do meaningful statistical analyses. Four medical diseases account for approximately 15% of Medicare admissions to short-term-care general hospitals: congestive heart failure (6.6%), pneumonia (3.6%), cerebrovascular accident (2.9%), and acute myocardial infarction (2.7%). These same conditions account for approximately 30% of in-hospital Medicare deaths and are associated with significant mortality rates within 30 days after hospital admission: congestive heart failure (15%), pneumonia (19%), cerebrovascular accident (20%), and acute myocardial infarction (26%). In addition, for each of these conditions, mortality rates vary substantially among hospitals. For example, among hospitals with 25 or more pneumonia admissions in fiscal year 1986, unadjusted 30-day postadmission mortality rates varied from 0% to 60%.

In addition, each of these four medical conditions can be diagnosed by means of a reasonably set of diagnostic criteria. Studies of hospital mortality differences among Medicare patients should begin with consideration of these diseases. Focusing on prevalent conditions for which differences in hospital mortality data exist results in the exclusion of all surgical and psychiatric conditions. We chose to include a surgical and a psychi-
atrie condition. We chose hip fracture for the surgical condition because it has high frequency, is predominately treated with surgery, and has a higher 30-day postadmission mortality rate than other prevalent surgical conditions. We chose depression for the psychiatric condition because it is the most prevalent psychiatric reason for admission to a short-term-care hospital. If diseases with low mortality rates, such as hip fracture or depression, are included in quality-of-care studies, it is important that outcomes other than mortality are also studied.

To study the mortality associated with a specific disease, patients with that disease need to be distinguished from other patients. Diagnosis related groups or the International Classification of Disease (ICD-9-CM) codes can be used for that purpose. The DRGs were developed to group together patients with similar expenditures for hospital care who also had related clinical problems. In contrast, the ICD codes were developed specifically to group patients who had similar clinical diagnoses and surgical procedures. Defining patient groups using the ICD codes rather than DRGs is more likely to result in patient groups whose clinical outcomes are unknown.

Even after patients are selected by ICD-9-CM codes, there will be clinical heterogeneity in the sample, because the coded diagnoses may not describe the patient’s admitting diagnoses in sufficient clinical detail. One way to increase the sample’s homogeneity is to apply clinically based inclusion and exclusion criteria. The clinical criteria could verify, for example, that among patients assigned an acute myocardial infarction code (ICD-9-CM code 410), all patients had evidence of an acute infarction according to a specified combination of historical, electrocardiographic, and enzyme evidence. Those without such evidence would be excluded. Furthermore, patients who had infarction as a complication rather than as the cause of the hospital admission would also be excluded. We selected patients for each disease by using ICD-9-CM codes and then applied clinically detailed inclusion and exclusion criteria.

Supplementing Mortality Outcome Data With Other Outcome Data

Even when we select diseases for study because of their relatively high mortality rates, the number of patients who die during or immediately following hospitalization in any time period or within any hospital will be small. This is particularly true if diseases with lower mortality rates than those mentioned here are studied or if a younger, non-Medicare cohort of patients is included. Even in the most favorable circumstances, a large sample size is required to identify statistically significant differences across time periods or hospitals. For example, approximately 1400 patients are required in both the pre-PPS and post-PPS study samples to achieve at least an 80% chance of detecting a change of at least 3.5 percentage points in mortality from a pre-PPS mortality rate of 12%.

One possible way to increase power without increasing sample size is to include outcomes other than death (eg, in-hospital complications, changes in functional status). For example, for patients with acute myocardial infarction, one could collect information about the occurrence of complications (eg, coma, shock, or pneumonia), periods of clinical instability (eg, unstable vital signs or new onset of altered mental status), or of life-threatening laboratory values (eg, serum potassium level <2.5 mmol/L).

Ignoring nonfatal adverse in-hospital outcomes and focusing only on whether the patient is discharged alive or dead is equivalent to assigning patients a score of either 0 or 100 on an outcome scale. It is not only reasonable but probably important to assign a score between 0 and 100 to patients who suffer adverse in-hospital outcomes but who do not die during the hospitalization or in the immediate posthospital period; if these patients were followed up for a longer time period, a large number of them would die, and others would have reduced functioning. We included nonfatal in-hospital complications. Further work is required to determine how much benefit is obtained from this additional information about the quality of care balanced against the additional costs of data collection and analysis.

Defining the Interval for Studying Outcomes

Regardless of the outcome to be studied, choosing the time period during which patients are observed is important. For example, one might study in-hospital outcomes, outcomes for an interval after admission (eg, 30 days after admission), or outcomes for an interval following discharge. To avoid bias, it is important to select the same interval for all patients studied. For instance, if we studied only in-hospital outcomes, hospitals with longer average lengths of stay would have a greater number of days during which the outcome could be observed than hospitals with shorter lengths of stay. Alternatively, if we considered a fixed time interval after admission, all patients would be observed for that time period regardless of whether they remained hospitalized or had been discharged.

Selecting a fixed time interval from admission has theoretical advantages but also creates a problem; the data on which the analysis is based must include information from both the in-hospital and postdischarge periods. Collecting such information can be difficult. This is true both for the Medicare population, for whom Social Security databases exist, and for the nonelderly population, for whom the National Death Index (a registry of death certificates maintained by the National Center for Health Statistics) may be used. We gathered in-hospital data from the medical record and supplemented it with postdischarge data obtained from Health Care Financing Administration files and fiscal intermediary databases.

ADJUSTING OUTCOMES

Patient Complexity

To use mortality rates as a meaningful proxy for hospital quality of care, we must know the complexity of the patient’s condition (sickness at admission) prior to the delivery of in-hospital care. Mortality rates can then be adjusted for complexity. We used literature reviews and consensus panels to identify variables that have been found in clinical studies to predict outcomes. Patient characteristics that predict outcomes and are sufficiently frequent to be worth collecting can be abstracted from the medical record and used to see if they help to explain the variability in outcomes across time periods or hospitals.

Patient complexity measures should be multidimensional. Information about morbid conditions that relate to the reason for hospital admission (eg, history of valvular heart disease in patients with congestive heart failure or history of coma in patients with cerebrovascular accidents) should be collected. In addition, comorbid conditions that describe clinical problems other than those related to the reason for hospitalization should be included. For example, depression has been demonstrated to be a predictor of outcomes for patients with acute myocardial infarction. Functional status prior to hospitalization constitutes another variable that has been demonstrated to be an important determinant of outcome. For example, ambulatory status prior to a hip fracture or cerebrovascular accident is an important determinant of whether a patient lives or dies, while incontinence is an important determinant of relocation from home to nursing home.
influence outcome, attention should be paid to collecting information about both acute and chronic conditions.

When studying the effect of complexity on outcomes, it is important to use a clinical model based on principles of pathophysiology and clinical judgment. It may be that acute complexity variables (e.g., fever, acute acid-base disturbances, and meningitis) have important effects on short-term outcomes (in-hospital or 30-day postadmission mortality rates), while chronic complexity variables (e.g., cancer, malnutrition, dementia, and depression) have important effects mostly on long-term outcomes (e.g., six-month postadmission mortality rate and functional status). If a particular complexity variable is not important in the prediction of short-term outcomes, it should not be discarded until its effect on long-term outcomes is also studied.

**Collecting Complexity Data**

To be useful, complexity data must be accurate. Complexity data can be obtained from routine billing information, through prospective data collection, and from the medical record. Age, sex, and diagnostic categories are available from billing data, but this information may not be sufficient to provide an adequate adjustment for the effect of complexity on outcome. Prospective data collection can make use of a patient interview or a checklist completed by a physician or nurse. Unfortunately, such techniques cannot be used to describe past hospital performance, and they are not currently feasible for routine use.

Medical record review is currently the most feasible system for collecting complexity data across time periods and hospitals. However, there are challenges associated with extracting clinically meaningful complexity data. After most feasibility variables have been precisely defined, multiple data sources within the medical record need to be reviewed (e.g., physicians' notes, nurses' notes, x-ray reports, laboratory results, electrocardiograms, and pathologic findings). For example, to recognize whether the patient is immunocompromised, the patient's history should be examined for use of the word immunocompromised, for diagnostic equivalents (e.g., metastatic cancer), and for the use of immunocompromising drugs (e.g., chemotherapy or steroids). Review of the physical examination results provides evidence about cachexia, debilitation, or thernar wasting, and review of laboratory reports provides information about white blood cell counts and albumin values.

Two. One has to avoid both overestimating the complexity level of patients who have multiple descriptions of complexity and understating the complexity level of patients with only a single description. For example, some records describe patients as sick according to a number of factors (e.g., history, diagnoses, use of medications, physical examination results, and/or laboratory data), while other records only use one term to describe an equally sick patient. In collecting and analyzing complexity data across time periods, institutions, or places, reasonable criteria must be developed to identify patients with similar complexity levels, even if there are varying amounts of complexity information in the record.

3. Careful attention should be paid to differences in medical language and recording practices. For example, if the study uses data from different time periods, one might expect differences in recording, since incentives for recording have changed over time. Physicians in different settings, such as teaching vs nonteaching hospitals or different sections of the country, may use different vocabularies to describe the same level of complexity.

4. Special care must be taken when using reports of studies rather than the raw data. This is especially true when using reports of roentgenographic studies. Review of the medical record allows access to the radiologists' (and/or clinicians') reports rather than to the roentgenograms themselves. Because roentgenograms may provide valuable information about patient sickness at admission (e.g., the presence of congestive heart failure or pulmonary edema for patients hospitalized with acute myocardial infarction or the presence of infarct, hemorrhage, or increased intracranial pressure for patients hospitalized with cerebrovascular accident), it is important to develop a valid system to grade the severity of roentgenographic findings.

5. An approach must be developed for handling unrecorded or missing data. For example, in the admission history and physical examination, physicians customarily report all important positive historical and physical findings but only pertinent negative findings. This means that abstraction from the medical record of many clinical characteristics present at the time of admission can identify the presence of a clinical problem but cannot distinguish the absence of the condition from a situation in which no data were found in the record about the presence or absence of the problem. Under these circumstances a bias can result if complexity variables are not carefully selected, because a hospital with detailed recording practices might appear to have sicker patients than a hospital with less-comprehensive practices. To reduce the likelihood of bias, complexity measures should include only items from the medical record that generally would be recorded if the patient had the disease in question. If this rule is followed, it will be possible to assign normal values or no comorbidity if specific data are absent. For example, if there is no report about abdominal pain on examination, the patient can be considered to have no pain. These decisions can be tested by comparing whether patients with missing data are more or less likely to die or suffer some other adverse outcome than are patients who have known negative or normal findings.

Preliminary results indicate that patients with missing laboratory data were no more likely to die than those with normal laboratory values, so that inferring normal values was reasonable. Because the degree of uncertainty increases with the amount of data inferred, caution is required in interpreting the findings of studies in which large amounts of the data were inferred, particularly if one time period or group of hospitals has more missing data than others.

6. Just as a choice needs to be made between disease-specific and nonspecific sampling, a choice needs to be made between using a generic measure of complexity and a disease-specific measure. Although APACHE II (Acute Physiologic and Chronic Health Evaluation), which uses the same items and the same weights to determine complexity across diseases, has been shown to account for a noticeable amount of the variability in hospital outcomes, clinical principles suggest that such a generic measure should be supplemented with various disease-specific item sets and weights. Our preliminary work and the findings reported elsewhere in this issue of THE JOURNAL suggest that the disease-specific approach produces models that better predict the outcome. We do not yet know whether this improvement is outweighed by the increased cost of gathering more data from the medical record.

**PATIENT AND HOSPITAL SAMPLING**

Collecting information on complexity is costly. This is true whether one is trying to study changes in outcomes over time or differences in hospital mortality at any point in time. To produce the most cost-effective sample for studying the effect of a PPS on the quality of care, we used cluster sampling based on five states representative of major geographic regions of the coun-
try, and we used four to six geographic areas in each of the states. We then randomly sampled approximately 12 hospitals in each area and approximately 60 patient records per hospital. Such a sample, which can be constructed to be representative of the US hospital and patient population, is also relevant to any effort at a national level to examine, at one point in time, differences in publicly released hospital mortality rates. If it is desirable (and this appears to be the case) to adjust publicly released hospital mortality data for patient complexity, at least two sampling strategies are possible.

1. The first option makes use of the above strategy and produces for a nationally representative sample information on specific conditions on the relationship of selected complexity variables to mortality rates, a mean value for each of the complexity variables, and complexity-adjusted mortality rates for the hospital sample. Using such data, a hospital could adjust its publicly released disease-specific, age-adjusted, and sex-adjusted mortality rates. Hospitals could calculate an expected and observed number of deaths for each disease and compare them statistically with the national sample. Only hospitals in the higher region of the raw mortality distribution for a given disease might wish to collect the additional data and perform the appropriate analysis.

2. The second option involves collecting complexity data for patients with specific diseases from all hospitals. This option would permit the public release of complexity-adjusted mortality data by hospital. This option may be preferred once it is clear which complexity variables should be included in such a national effort. Before this is attempted, however, we need to acquire more knowledge both about the effects on outcome of complexity adjustment at the hospital level and about the costs (in dollars, morale, and time) associated with a compulsory national activity.

**USING ADJUSTED OUTCOME DATA AS A MEASURE OF QUALITY**

Once data about complexity and outcomes are available for patients in a representative group of hospitals, we can use logistic and other regression methods to determine, for each disease and even across diseases, which variables suggested by clinical experience and judgment are important predictors of outcomes. After the important complexity variables are used to adjust outcome data across hospitals, the next step is to assess whether the adjustments are sufficient so that differences in complexity-adjusted outcomes can serve as proxies for quality. This judgment can best be made through an independent assessment of hospital performance based on study of the process of care. Once data about the process of care (i.e., what health professionals do to or for patients) are also collected, after clinical models are specified and outcomes are adjusted for complexity, we can determine whether a better process is associated with a better outcome. This work is currently under way as part of the RAND Proactive Payment Study. When such relationships are found, we can increase our confidence that differences in outcomes either over time or across hospitals truly reflect differences in the quality of care.

With the recent changes in financial incentives for hospitals and the evidence of large variations in age-, sex-, and disease-specific mortality rates across hospitals, it seems worthwhile to try to understand what outcome measures can tell us about the quality of care. The steps described here summarize some of the tedious, costly, and intellectually challenging issues that need to be considered in trying to relate differences in outcomes to differences in the quality of care. Research on the subject is not likely to occur without the use of hospital mortality rates, but users must realize that more research is needed to increase the validity of this information.

This study was supported by cooperative agreement 18-C-88833/9-03 with the Health Care Financing Administration, US Department of Health and Human Services.

We thank Joseph E. Newhouse, PhD, and Albert P. Williams, PhD, for their contributions. We also thank Florence McIlvy for assistance in preparing the manuscript.

**References**
