Interpreting Hospital Mortality Data
The Role of Clinical Risk Adjustment

Stephen F. Jencks, MD, MPH; Jennifer Daley, MD; David Draper, PhD; Neal Thomas, PhD; Gregory Lenhart, MS; Janice Walker, BSN, MBA

This study uses national Medicare data as well as data that were abstracted to calibrate the Medicare Mortality Predictor System to assess the usefulness of a risk adjustment system in interpreting hospital mortality rates. The majority of variation in annual hospital death rates for the four conditions studied (stroke, pneumonia, myocardial infarction, and congestive heart failure) is chance variability that results from the relatively small numbers of patients treated in most hospitals in a year. For hospitals in the highest and lowest quartiles of observed death rates, the difference between observed rates and those predicted by the Medicare Mortality Predictor System is not quite one third smaller than the difference between observed rates and unadjusted national rates. Risk adjustment methods do not show whether the unexplained difference in mortality rates results from differences in effectiveness of care or unmeasured differences in patient risk at the time of admission. Risk-adjusted mortality rates, therefore, should be supplemented by review of the actual care rendered before conclusions are drawn regarding effectiveness of care.

See also pp 3617 and 3625.

THE PURPOSE of this article is to help physicians and hospital managers assess what hospital death rates reveal about effectiveness of care. In a companion article we describe the Medicare Mortality Predictor System (MMPS), a microcomputer-based system that was created to help hospitals interpret death rates. The MMPS uses clinical data abstracted from medical records to estimate the likelihood of death within 30 days of hospital admission for patients older than 64 years of age who are admitted with stroke, pneumonia, myocardial infarction, and congestive heart failure. In this article we provide a set of analyses intended to help users of hospital mortality data decide when and how a system such as the MMPS might be useful in interpreting death rates.

Figure 1 shows an excerpt from Medicare Hospital Mortality Information 1986 from the Health Care Financing Administration (HCFA). This publication provides death rates for 17 diagnostic categories for every hospital that treated Medicare patients. It also provides, for each reported rate, a predicted range that is defined so that about 5% of hospitals are expected to fall outside the range by "chance." Others may fall outside the range because of a systematic tendency to treat unusual patients or to provide unusually effective or ineffective care. The publication's users must decide whether the relation of the observed death rate to the predicted range raises sufficient concern about effectiveness of care to warrant further study.

The ranges are adjusted for patient risk using Medicare administrative data, but these data contain no clinical information beyond diagnoses and procedures. The MMPS was created to let hospitals compare actual death rates with predictions based on such clinical admission data as vital signs, laboratory findings, and ability to walk. Table 1 shows a report produced by the MMPS for one condition in a simulated hospital. The report gives the hospital's actual mortality rate, a predicted rate based on clinical data, the difference, and the probability that a difference this large or larger would occur by chance. The goal of this article is to help the reader understand what a report like that in Table 1 means and when it would be useful. To do this we address four questions: (1) What is the role of unpredictable or chance variations in hospital death rates? (2) Are death rates easier to interpret when averaged over patients with different conditions? (3) How much of the observed variations in death rates among hospitals does the MMPS predict? (4) Does MMPS provide a clinically reasonable way to screen cases for clinical review?

The common theme of these questions is that variation in hospital mortality rates has three sources: the underlying average risk the hospital's patients carry on admission; the underlying average effectiveness of the care the hospital renders; and unpredictable or chance variation in the risk that individual patients carry, the effectiveness of the care they receive, and the outcomes they experience. These sources of variation interact in complex ways; eg, patients with similar illnesses may respond differently to the same treatment, and physicians might choose different treatments for reasons that are difficult to measure. Because mortality rates are measurable only through the survival or death of individual patients, a large number of patients is necessary to determine a hospital's underlying death rate accurately. Although a hospital might have a relatively stable underlying death rate for a condition, the observed annual rate may fluctuate substantially because of chance variation. If a hospital's underlying mortality rate was 20% and the hospital treated 100 patients each year for 100 years, we would expect by chance alone that the mortality rate would be greater than 24% or less than 12% in about five years.

METHODS

Database

We used two sources of data for this analysis. The first (the national database) was a file of all discharges in the
Medicare Provider Analysis Record for Oct 1, 1985, to Sept 30, 1986, that met International Classification of Diseases, Ninth Revision—Clinical Modification criteria as cases of stroke, pneumonia, myocardial infarction, or congestive heart failure. The diagnostic code criteria are described in the accompanying article. We used the national file for the analyses of the variation in death rates that results from having small numbers of cases from individual hospitals and for the analysis of the correlation among mortality rates for the four conditions.

The second data source (the MMPS database) was the calibration data collected for the HCFA mortality prediction project and described in the accompanying article. This file comprises 5998 cases that are distributed equally across the four conditions and between patients who died within 30 days of admission and those who survived. Each record includes the estimated risk of death for that patient generated by the MMPS. All analyses on the MMPS data were performed using the patient as the unit of analysis rather than the hospital. Except as noted, all analyses were performed separately for each condition. For analysis, MMPS data were weighted to represent Medicare admissions of patients older than 64 years of age in the seven states from which the MMPS database was drawn; Daley et al describe this weighting. Because cases selected for the MMPS database met clinical inclusion criteria that could not be applied to the Medicare Provider Analysis Record file, the MMPS database is more clinically homogeneous than the national file.

**Definition of Hospital Mortality**

We defined mortality as death within 30 days of admission, which is the definition used both in the MMPS and in Medicare Hospital Mortality Information 1986. Both databases contain the date of death from death certificates. Calculated hospital mortality rates for the MMPS database are slightly lower than those for the national database because identifiable transfer admissions were ineligible for MMPS sampling. Death rates for individual hospitals in the MMPS database are calculated from outcomes for patients eligible for sampling into that database.

**The Risk Adjustment System**

We used the MMPS for risk adjustment in the MMPS database. We did not evaluate or use the risk adjustments computed by the HCFA for its mortality information release because our categories do not match those used in the 1987 release.

**Chance Variation**

We analyzed the national database to determine the 30-day mortality rate for each of the four MMPS conditions in each of the nation's 5576 short-term hospitals. We then calculated characteristics of the distribution of mortality rates using the hospital as the unit of analysis (Table 2). Next, we determined the distributions of mortality rates that would be expected on the basis of chance (binomial) sampling variation if all hospitals had the same underlying mortality rate and differed only in the number of discharges in each condition, which we set to the observed annual values from the national database. We further calculated the difference between these distributions by subtraction. Finally, we employed a β-binomial model, which is more realistic than the simple binomial model. Each hospital in this model has its own underlying mortality rate, with its annual observed number of deaths varying binomially according to this rate. The distribution of underlying rates in this model is assumed to be a β-distribution, whose parameters we estimated by maximum likelihood using standard methods. The result is the estimated distribution of underlying rates that would be most likely to lead to the observed distribution of hospital death rates. These calculations are similar in logic and detail to those used in the estimation of underlying cancer mortality rates in small geographic areas and in other medical and nonmedical settings.

We performed these calculations for all four conditions in all hospitals (Table 2) and plotted the distributions for one condition, stroke (Fig 2). To assess how the relative magnitude of total variability...
Correlation of Death Rates Among Conditions

We used the national data set to calculate, for each condition, the simple correlation between the death rate for the condition in a hospital and the composite death rate for the other three conditions in the same hospital (Table 4). We performed these calculations for the same strata used for Table 3.

Explanatory Power of the MMPS

Because the MMPS database has an average of only 11.3 cases per hospital, it cannot be used to estimate the typical size of the MMPS risk adjustment at individual hospitals. To obtain an idea of the hospital-level effect of the MMPS, we ranked all MMPS cases according to the mortality rate for the given condition in the hospital that provided the care and divided the cases into quartiles on the basis of this ranking. We then calculated observed mortality rates and average estimated risk for the top and bottom quartiles for each of the four conditions (Table 5). Finally, we divided the difference between the predicted mortality rates for the highest and lowest mortality groups by the difference in the observed mortality for these groups. The resulting fraction is a rough estimate of the percentage reduction in the unadjusted difference between observed and national average death rates that could be expected by a user of the MMPS whose death rates are in the outer quartiles.

Thresholds for Clinical Reviews of Individual Cases

The simplest selection rule for clinical review would be to choose patients who died despite a low estimated risk on admission. To estimate the number of cases that would be selected for clinical review by this rule, we computed the cumulative frequencies of estimated risk for those who died (Table 6).

RESULTS

Chance Variation

Figure 2, top left, shows the distribution of raw annual hospital death rates for Medicare patients admitted to the nation’s hospitals with a diagnosis of stroke in federal fiscal year 1986. Figure 2, top right, shows the distribution of death rates for stroke that would be expected, given the number of patients treated in each hospital during the year, if all hospitals had the same underlying death rate. The variation in Fig 2, top right, is entirely chance variation resulting from the limited number of cases treated at each hospital during the year; the importance of this effect is emphasized by the fact that the majority of hospitals have fewer than 50 stroke discharges a year and an annual combined total of fewer than 200 discharges with the four MMPS conditions.

Figure 2, bottom left, shows the difference between the top left and top right. A mortality rate that occurs in more hospitals than would be expected from chance variation alone shows as a bar above the line; a mortality rate that occurs in fewer hospitals than expected from chance alone shows as a bar below the line. Compared with the distribution of death rates that would be expected by chance, there are somewhat fewer hospitals with rates near the average and somewhat more with rates substantially above or below average.

Figure 2, bottom right, shows a theoretical distribution of underlying hospital death rates for stroke that is sufficient, when combined with chance variation at the patient level, to generate the distribution of observed death rates shown in Fig 2, top left. This un-
underlying distribution of death rates is substantially more concentrated around the average than the observed variation in annual death rates, but there is more variation than might be expected from the similarities of Fig 2, top left and top right. Corresponding plots for pneumonia, acute myocardial infarction, and congestive heart failure have somewhat different shapes, but the overall conclusions are similar.

Table 2 presents a numerical summary of the data in Fig 2 and of the corresponding figures for the other three conditions. This table illustrates that chance variation alone is sufficient to account for a substantial amount of the observed variability in death rates and that the variability of the underlying death rates across all hospitals is far less than the observed variability in annual rates for all four conditions.

Table 3 presents the SDs of observed hospital death rates and the SDs that would be expected from chance variation alone. It shows the steady decrease in variability with increasing number of discharges. It also shows that chance variability is a much larger fraction of the observed variability in annual rates for smaller hospitals than for larger hospitals and for individual conditions than for all four conditions. Pooling the four conditions reduces chance variation about as much as having four times as many discharges for one condition.

**Correlation of Death Rates Among Conditions**

Table 4 shows that the correlation among death rates for the four MMPS conditions is not strong, although the strength of the association does rise sharply in larger hospitals. There is no case in which the death rate for one of the four conditions correlates more strongly than $r = .5$ with the composite death rate for the other three. This correlation is equivalent to saying that if we know the death rate for one condition, we will be able to predict only 25% of the variation in the composite death rate for the other three conditions even in the largest hospitals. Few hospitals can expect correlations even this strong because the highest correlations are in hospitals with more than 1000 annual discharges in the four conditions, and only about 1% of all hospitals have so many.

**Explanatory Power of the MMPS**

The MMPS predicts about 20% of variation in outcomes for individual patients, but it might be either more or less powerful in predicting variation among hospitals. Table 5 shows the average observed and predicted mortality rates for hospitals in the MMPS database lying in the highest and lowest quartiles of mortality rates. The numbers provide crude estimates of how the difference between a hospital's death rate and an MMPS prediction might be expected to compare with the difference between a hospital's death rate and the national average rate for all hospitals. The decreases in these differences after MMPS adjustment for risk range roughly from one third to one fifth. The decreases can be thought of as the fraction of the difference between observed and average mortality in hospitals in the outer quartiles that is accounted for by the MMPS, but they are not analogous to the percentage of explained variance expressed by $R^2$. Nor should a hospital expect decreases this large from the differences published in the HCFA's Medicare Hospital Mortality Information 1986 because those differences already have received a limited risk adjustment.

**Thresholds for Clinical Reviews of Individual Cases**

Table 6 shows percentile ranks for various estimated likelihoods of dying for nonsurviving patients in each of the four conditions. Half of patients who die have predicted likelihoods of death below .32 for stroke, below .26 for pneumonia, below .29 for acute myocardial infarction, and below .20 for congestive heart failure. A larger fraction of patients usually would have values above these levels in most hospitals with high mortality rates, but a hospital that reviewed the half of its deaths with lowest predicted likelihood of death for each condition would be focusing on cases for which death would be unlikely if the mortality rate of the hospital was like that of the rest of the nation.

**COMMENT**

**Chance Variation**

Our analysis of chance and underlying variation in death rates might be limited by inaccuracies in the diagnostic data on...
Table 5.—Observed and Medicare Mortality Predictor System (MMPS)—Adjusted Mean Mortality Rates

<table>
<thead>
<tr>
<th>Hospital Type</th>
<th>Condition</th>
<th>MMPS Adjusted</th>
<th>MMPS Adjusted</th>
<th>MMPS Adjusted</th>
<th>MMPS Adjusted</th>
<th>MMPS Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>All hospitals</td>
<td>Stroke</td>
<td>212</td>
<td>207</td>
<td>180</td>
<td>185</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td>Pneumonia</td>
<td>246</td>
<td>213</td>
<td>210</td>
<td>206</td>
<td>262</td>
</tr>
<tr>
<td></td>
<td>Acute Myocardial Infarction</td>
<td>128</td>
<td>187</td>
<td>109</td>
<td>156</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>Congestive Heart Failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*From the MMPS calibration sample.

Table 6.—Distribution of Predicted Death Probabilities for Patients Who Died

<table>
<thead>
<tr>
<th>Percentile of Nonsurvivors, %</th>
<th>Congestive Heart Failure</th>
<th>Stroke</th>
<th>Acute Myocardial Infarction</th>
<th>Pneumonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.027</td>
<td>0.040</td>
<td>0.070</td>
<td>0.042</td>
</tr>
<tr>
<td>5</td>
<td>0.047</td>
<td>0.067</td>
<td>0.098</td>
<td>0.081</td>
</tr>
<tr>
<td>10</td>
<td>0.062</td>
<td>0.090</td>
<td>0.12</td>
<td>0.080</td>
</tr>
<tr>
<td>25</td>
<td>0.10</td>
<td>0.15</td>
<td>0.18</td>
<td>0.13</td>
</tr>
<tr>
<td>50</td>
<td>0.20</td>
<td>0.22</td>
<td>0.29</td>
<td>0.26</td>
</tr>
<tr>
<td>75</td>
<td>0.40</td>
<td>0.41</td>
<td>0.47</td>
<td>0.51</td>
</tr>
<tr>
<td>90</td>
<td>0.64</td>
<td>0.63</td>
<td>0.70</td>
<td>0.74</td>
</tr>
<tr>
<td>95</td>
<td>0.75</td>
<td>0.89</td>
<td>0.82</td>
<td>0.84</td>
</tr>
<tr>
<td>99</td>
<td>0.95</td>
<td>0.95</td>
<td>0.93</td>
<td>0.92</td>
</tr>
</tbody>
</table>

*From the Medicare Mortality Predictor System calibration data.

which the condition-specific death rates depend. However, in developing the MMPS we did apply clinical diagnostic criteria to cases from this file, and these criteria did not substantially change the death rates. On the whole, we believe we have presented a fair picture of the importance of chance variation and one that reflects both clinical and administrative issues.

Clinicians are well aware that the outcomes for most patients are not predictable and that survival is determined by many factors we cannot measure as well as some we can. The fact that substantial variations in death rates will occur from month to month despite a stable underlying death rate reflects the clinical reality that patients differ from one another and that outcomes differ for apparently similar patients.

The most straightforward way of reducing chance variation is to increase the number of cases in a year by aggregating discharges across conditions that are treated by similar groups of physicians and hospital staff. Aggregation increases the number of patients, which decreases the relative importance of chance variation, but aggregation may obscure high or low death rates for individual conditions. The four MMPS conditions generally are treated by family practitioners, internists, or internal medicine subspecialists; except acute myocardial infarction, they generally are treated on the same medical nursing units. Nevertheless, we find that the correlation among mortality rates for the four MMPS conditions is only moderate, which suggests that aggregation carries a considerable risk of obscuring problems in individual conditions. Averaging less-related conditions seems to present greater problems. A second way to increase the number of cases is to average performance over periods longer than a year, an approach recently recommended by the General Accounting Office. Examining data from multiple years is useful, but going back more than a modest period threatens to produce results of largely historical interest. For example, in the 1987 mortality data that the HCFA will publish in December 1988, the average case was discharged in mid-1987; averaging over three years would include cases from early 1985. The concept of a stable underlying mortality rate may be largely hypothetical because of changes in hospital, physician, and patient characteristics. Because of chance variation and these other influences, mortality data are not very stable over time. In 1980, a hospital whose overall mortality rate was above the HCFA’s predicted range in one year had a 52% chance of being within the range in the next year and a 56% chance of being within the range two years later.

Administators and physicians should realize that aggregating experience from several years or from several conditions may create an erroneous appearance of stability that does not exist. Failure to understand chance variability can lead to overreactions to that variation. Leaders in industrial quality control emphasize the futility and danger in trying to reward or punish chance variation. They point out that understanding the pattern of chance variation is a critical prerequisite for efforts to improve practices or identify ineffective workers. As hospitals seek to understand outcomes such as mortality, they will need to understand the nature of chance variation and to learn how to distinguish chance variation from underlying patterns. Making these distinctions is an art, not a clean science, and a risk adjustment tool such as the MMPS cannot totally adjust away chance variation. In addition, aggregation may obscure variations that are so unlikely to occur by chance that they should be investigated promptly. For these reasons it is important to examine outcomes for both individual conditions and individual years and also for aggregated outcomes and trends.

Explanatory Power of the MMPS

For hospitals in the lowest and highest quartiles of observed mortality rates, the difference between observed death rates and those predicted by the MMPS is one third to one fifth less than the difference between observed death rates and national averages, depending on the condition studied. Iezzoni et al. studied 12 hospitals whose death rates deviated from the average rate by an average of 9.5%; after adjustment using MedisGroups scores, the rates deviated from average by 8.2%. This reduction of about 1.3 percentage points, or 14%, is consistent with our results with the MMPS, given the slightly lower predictive power of MedisGroups.

On the other hand, DuBois et al. used adjustment methods similar to ours to study stroke, pneumonia, and myocardial infarction. They found that risk adjustment explained more than two thirds of the difference in death rates between hospitals, and they found significantly more “avoidable” deaths in hospitals with high mortality rates. Their sample makes results difficult to generalize because it was drawn exclusively from hospitals with outlier death rates; risk adjustment may account for more of death rate variation in hospitals with extreme rates, but we cannot test this hypothesis with our data. Also, the low reliability of their identification of avoidable deaths (r = .10 to .50) makes their conclusions regarding avoidable deaths difficult to assess.

What clinicians and a great many others would like from a risk adjustment instrument like the MMPS is a powerful diagnostic tool that will show whether a hospital suffers from a disease we might
call “ineffective care.” Our analysis suggests that MMPS is closer to being a screening tool than an instrument for definitive diagnosis. Available evidence suggests that the other two generally available risk adjustment instruments, MedisGroups and APACHE II, are unlikely to be better (the accompanying article focuses on the relative performance of MMPS, MedisGroups, and APACHE II in making predictions for individual cases).

How should the user interpret the MMPS report in Table 1? First, the user should understand that the MMPS is not simply a refinement of the HCFA release but a different tool. The definition of conditions is slightly different: there is an 80% to 90% overlap between the MMPS diagnosis categories and the corresponding categories in the HCFA release, but MMPS also uses clinical inclusion criteria. The statistical methods are different: the HCFA release provides a predicted range while the MMPS estimates the likelihood that a difference as large as that between observed and predicted rates would have occurred by chance. The HCFA publication examines death rates for each patient’s last admission in the year while the MMPS uses all admissions.

The MMPS leaves more than two thirds of the differences in death rates among the outer quartiles of hospitals unaccounted for. While this unexplained variation may result entirely from variation in effectiveness of care, there is no data to support so radical an interpretation. Given our limited ability to adjust for clinical risk, we suggest a three-step approach to interpreting mortality data:

1. Examine patterns across conditions and over years. The predicted range in the HCFA publication (a 95% confidence interval) is a familiar concept, but it is a somewhat arbitrary guideline for helping hospitals decide whether they have a problem. With more than 6000 hospitals in the country, about 300 will fall outside the range by chance. A hospital that falls at the high end of this range for several years or for several conditions has more cause for concern than a hospital that falls slightly outside the range one year and in the middle the next year. Nevertheless, because correlations among death rates for different conditions and across years are only moderate, hospitals should not discount a seriously aberrant rate for one condition or time period solely because other rates are unremarkable. Hospitals should use clinical and managerial knowledge of the hospital’s programs to assess death rates. An outlier death rate in a unit with high staff turnover and questionable physician skills is more alarming than a similar death rate in a stable, high-morale unit with skilled, involved clinicians.

2. The MMPS is a useful way to test, for the four conditions, whether a high death rate results from patients being at higher risk on admission. If the difference between an observed rate and an MMPS prediction might well have occurred by chance, a hospital should be more comfortable. Our results have lead us to expect that, for hospitals with high rates, the MMPS will predict a death rate higher than the national average but lower than the observed rate.

3. Remaining concerns about effectiveness of care require assessment of clinical practices and staff communication on units with high death rates and review of individual medical records. The MMPS is a practical tool for selecting individual cases for clinical review; reviewing deaths of patients who had lower-than-average predicted risk on admission seems to be clinically reasonable and administratively feasible. For technical reasons, when there are fewer than 20 discharges the MMPS does not estimate the likelihood that chance variation would have produced a difference between observed and expected death rates as large as the difference that occurred. In these situations, review of unexpected deaths assumes greater importance. Systems other than MMPS can be used for this step, but cannot be used for step 2 because they are not calibrated on national data. While this strategy is reasonable, only limited evidence indicates that it is a powerful way to identify cases in which care was ineffective, and future research on this question is important.

Our strategy is conservative. We see mortality data as a screening tool of unproved specificity and sensitivity and believe that risk adjustment can improve the accuracy of screening only moderately at the present state of the art. Our view is appreciably more conservative than that of proponents of such risk adjustment systems as APACHE II and MedisGroups. We also are more optimistic than the state legislatures that have required collection of data to measure "quality of care." We are uncertain that the current state of the art can meet such a requirement.

Prospects

Two developments may improve our ability to relate hospital mortality rates to effectiveness of care:

1. Instruments that predict death more accurately. The most dramatic improvements would result, we believe, from greater ability to measure chronic health status (ie, nutritional status, severity of chronic diseases, and psychoendocrine status). However, we do not yet have scientific consensus on how to measure these variables, and the necessary data are certainly not in most medical records. We also believe that prediction would be enhanced by better measures of the importance attached to keeping the patient alive. By contrast, we doubt that more-detailed measurement of acute physiological variables can produce major improvements; we found, in constructing the MMPS, that even variables that are independent predictors of death made only modest additions to the predictive equation. The most important obstacles to more accurate prediction are the limited data in medical records and our limited scientific understanding of why some patients survive and others die.

2. Empirical studies of whether discernibly ineffective care is substantially more common among patients who die despite a low predicted likelihood of death. These studies will be difficult to carry out because measurements of effectiveness of care tend to be unreliable and of uncertain validity, but they are critical if more realistic improvements are likely to produce gains, and to the agenda that state legislatures have set in requiring collection of data that will measure quality of care.

We offer two final comments. (1) We should be interested in identifying and understanding hospitals that provide excellent care as well as identifying and improving hospitals with problems. (2) The difficulties we have identified in interpreting hospital death rates also apply to other outcomes. Although using risk-adjusted outcomes to assess effectiveness of care is an exciting approach to improving medical care, it will require all of the caution and judgment needed in other kinds of medical innovation.

References