

A RAND NOTE

**Hospital Inpatient Mortality:
Is It a Predictor of Quality?**

**Robert W. Dubois, William H. Rogers,
John H. Moxley III, David Draper,
Robert H. Brook**

RAND

The research described in this report was supported by a grant from the U.S. Department of Health and Human Services.

This Note contains an offprint of RAND research originally published in a journal or book. The text is reproduced here, with permission of the original publisher.

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of RAND do not necessarily reflect the opinions or policies of the sponsors of RAND research.

A RAND NOTE

N-3372-HHS

**Hospital Inpatient Mortality:
Is It a Predictor of Quality?**

**Robert W. Dubois, William H. Rogers,
John H. Moxley III, David Draper,
Robert H. Brook**

**Supported by the
U.S. Department of Health and Human Services**

RAND

HOSPITAL INPATIENT MORTALITY

Is It a Predictor of Quality?

Abstract Various potential measures of quality of care are being used to differentiate hospitals. Last year, on the basis of diagnostic and demographic data, the Health Care Financing Administration identified hospitals in which the actual death rate differed from the predicted rate. We have developed a similar model. To understand why there are high-outlier hospitals (in which the actual death rate is above the predicted one) and low-outlier hospitals (in which the actual death rate is below the predicted one), we reviewed 378 medical records from 12 outlier hospitals treating patients with one of three conditions: cerebrovascular accident, myocardial infarction, and pneumonia.

After adjustment for the severity of illness, the death rate in the high outliers exceeded that predicted from the severity of illness alone by 3 to 10 percent, and in the low outliers, the actual death rate fell short of the severity-adjusted predictions by 10 to 15 percent ($P < 0.01$). Reviews of the process of care using 125 criteria revealed no differences between the high and low outliers. However, detailed reviews by physicians of the records of patients who died during hospitalization revealed a higher rate of preventable deaths in the high outliers than in the low outliers. For the three conditions studied, we project that 5.7 percent of a standard cohort of patients admitted to the high-outlier hospitals would have preventable deaths, as compared with 3.2 percent of patients admitted to the low-outlier hospitals ($P < 0.05$).

A meaningful comparison of hospital death rates requires adjustment for severity of illness. Our findings indicate that high-outlier hospitals care for sicker patients. However, these same hospitals or their medical staffs may also provide poorer care. Our results need confirmation before death-rate models can be used to screen hospitals.

DEATH rates among hospitalized patients have increasingly been used to compare hospitals and set policy. Some authors have used them to argue for regionalization of procedures, citing data indicating that hospitals performing a large number of a specified procedure have lower death rates after the procedures than do hospitals performing a small number.¹⁻⁷ Others have suggested that hospital death rates may be a useful screening tool for identifying hospitals that provide inadequate care.⁸⁻¹¹

In part, hospital death rates have become popular as potential indicators of quality because they are available. Unadjusted hospital death rates, however, may chiefly reflect differences in the mix of patients admitted — not differences in the quality of care. For this reason, a variety of models have been developed to adjust a hospital's death rate for its patient mix. Drawing on information in computerized data bases, these models typically account for differences in patients' demographic and diagnostic characteristics but do not correct for differences in the severity of illness or comorbidity.^{8,10,11} These models produce an expected death rate for each hospital, that can be compared with its actual death rate.

How should we view hospitals that according to these models have death rates much higher or lower than expected? Do the hospitals with death rates that are higher than expected provide lower-quality care, or do they care for a group of patients who are sicker than average but whose sickness is not adequately reflected by the variables in the model? Are discrepancies in the coding of diagnoses or errors in the data base the source of the apparent differences in hospital mortality? The answers to these questions will determine whether death-rate models can appropriately be used as indicators of hospital quality.

Using multiple regression, we recently developed a model that predicts inpatient mortality rates.¹² It was based on 205,000 patient discharges from all 93 hospitals in one investor-owned chain. The model had four independent variables (the percentage of patients older than 70 years, percentage of admissions from the emergency department, percentage of admissions from nursing homes, and hospital case-mix index) and accounted for 64 percent of the variance in hospital death rates. The model identified 11 hospitals in which the observed death rate exceeded the expected death rate by more than 2 SD (high outliers). At nine hospitals, the expected death rate exceeded the actual death rate by a similar margin (low outliers).

In this study, we used clinical data from medical records to assess the quality of care provided by a sample of the outlier hospitals and the severity of illness of the patients admitted to them. Our purpose was to compare the results of the assessments of the medical records with the predictions based on the death-rate model and to comment on the value of using data on in-hospital patient mortality as a screening tool to identify hospitals whose medical or nursing staffs provide inadequate care.

METHODS

Overview

First, we developed a sampling design for choosing medical records for review from the selected outlier hospitals. Second, on the basis of data in the medical records, we examined whether hospitals were outliers because of differences in diagnostic coding practices. This entailed using a set of standard diagnostic criteria to determine rates of misclassification. Third, we ascertained whether death rates in outlier hospitals reflected differences in the complexity of cases at the time of hospital admission. Fourth, we determined whether death rates in outliers were due to differences in quality of care among the hospitals. In this phase, we used both structured (explicit) and subjective (implicit) reviews of quality of care.

In several of these activities (the development of diagnostic criteria, the development and weighting of structured process criteria, and the subjective review of quality), we were assisted by three panels of experts (on cerebrovascular accident, pneumonia, and myocardial infarction). The panel members were board-certified specialists or subspecialists and were selected on the basis of their reputation in the community. They represented diverse types of clinical practice, including academic and private practice, and practice in tertiary, municipal, and community hospitals.

Choice of Conditions

The death-rate model that we used to identify outliers was based on data in which the hospital, rather than the patient, was the unit of analysis. To determine whether the 20 outlier hospitals identified by this model remained outliers if patient-level data were used, we performed a patient-level logistic regression¹³ for each of three conditions: myocardial infarction (diagnosis-related groups [DRGs] 121, 122, and 123), pneumonia (DRGs 79 and 89: simple pneumonia in elderly or compromised patients and complex pneumonias); and cerebrovascular accident (DRG 14). These three conditions accounted for 36 percent of all deaths in the outlier hospitals. Surgical conditions were not examined because individually they accounted for few in-hospital deaths. The logistic regressions adjusted for age and sex.

Sampling of Medical Records

We sampled the medical records of patients with the three conditions treated in 6 of the 11 high-outlier hospitals and 6 of the 9 low-outlier hospitals. We sampled hospitals that had the largest disparity between actual and predicted death rates, while maintaining a geographic representativeness that matched the original 93 hospitals.

One of us reviewed 378 medical records from admissions in calendar year 1985 (106 patients with cerebrovascular accident, 132 with pneumonia, and 140 with myocardial infarction). We chose records according to stratified random sampling within age, sex, and discharge-status categories. The age and sex distribution in the sample matched that in its counterpart within each outlier group. We oversampled in-hospital deaths to include equal numbers of patients who died and who survived hospitalization.

After a reweighting for the oversampling of deaths, all the data from the six high-outlier hospitals were aggregated into one group and all those from the six low-outlier hospitals were aggregated into a second group.

Errors in Diagnostic Coding

We first examined the possibility that hospitals were outliers because of differences in diagnostic coding practices. This entailed determining rates of diagnostic misclassification. Standard diagnostic criteria developed by our expert panels were used. The final analysis included only patients who met the diagnostic criteria for each condition (see Appendix).

Case Complexity

We next examined whether hospitals were outliers because of differences in the complexity of cases. Using a standard form, we collected data on the severity of illness and on about 50 comorbid

conditions that were present on the day of admission. On the basis of this information, we developed a comorbidity scale for each patient that reflected the number of body systems (e.g., cardiovascular) that were affected.

The form also included the APACHE II scale.¹⁴ This severity-of-illness scale incorporates the most abnormal value within the first 24 hours after admission for 12 objective physiologic variables (e.g., heart rate, blood pressure, serum potassium level, and serum pH). The APACHE II scale is also based on age and the presence of severe chronic disease.

For myocardial infarction, we developed a simplified version of the Killip classification.¹⁵ It comprised the following classes: I (clear lung fields), II (rales present only over the lower 50 percent of the lung fields), III (rales present over more than the lower 50 percent of the lung fields), and IV (cardiogenic shock).

For patients with a cerebrovascular accident, we developed a scale based on the presence or absence of a mass effect or of blood on computerized tomographic scans of the head.

Reliability

To assess the reliability of our method for evaluating the complexity of cases, two physicians independently abstracted the same 30 charts (approximately 10 percent of all charts). We compared the scores with the Pearson correlation coefficient and found a value of 0.87.

Data Analysis

For each condition, we constructed a separate logistic-regression model¹³ (details of the models are available from us). These models used the scales relating to comorbidity and severity of illness on admission to predict each patient's probability of death during hospitalization. (See Appendix for a description of how we handled missing data.) Next, using weighted sums to correct for sampling, we reaggregated these values to determine a predicted death rate for high- and low-outlier hospitals. Finally, we developed confidence intervals around these estimates, using the bootstrap method¹⁶ of repeated random sampling from a defined population.

Review of Quality of Care

Structured Review of the Process of Care

We obtained a provisional list of criteria for quality of care in the management of the three conditions studied. Our expert panels modified these criteria, weighted each process element from -10 to +10, and grouped them in scales. Each condition had approximately 125 separate quality-of-care criteria. Many of the criteria used branching logic.

One of us who is a physician collected information that allowed all the cases to be assessed according to the structured criteria. After reweighting the data to reflect the sampling design, we compared the process scores in the high- and low-outlier hospitals, using *t*-tests for differences in means.

To assess the reliability of the structured process review, another of us who is a physician independently abstracted 30 charts (approximately 10 percent of all the charts). We used the Pearson correlation coefficient to compare the scores and found a value of 0.71.

Subjective Review of the Process of Care

One of us dictated case summaries describing the 182 patients who died during their hospitalization (50 with cerebrovascular accident, 70 with pneumonia, and 62 with myocardial infarction). These dictations followed a specific format (see Appendix).

The case summaries were evaluated by one of three expert panels (on cerebrovascular accident, pneumonia, and myocardial infarction). All the panel members underwent a 30-minute orientation session before beginning the evaluation. The dictated summaries were reviewed by each panel member, without contact or discussion with other members. For patients with cerebrovascular accident and myocardial infarction, three reviewers evaluated each case. Because of poor reliability among raters in evaluations of patients

with pneumonia ($r = 0.11$), half the cases (chosen at random) were rated by an additional three physicians (for a total of six independent reviews).

The reviewers evaluated the hospital stay and judged whether the death might have been prevented. The study staff instructed the panel members to focus only on the care received by the patient after arrival at the outlier hospital or its emergency department.

The physicians judged the preventability of death according to a scale with four levels: definitely not preventable, probably not preventable, probably preventable, or definitely preventable. For the final analysis, we collapsed the scale into a dichotomous response (preventable or not preventable) and assigned the patient to the relevant category on the basis of a majority rule.

We determined the proportion of deaths that were preventable (preventable deaths divided by all deaths). We next determined how often preventable deaths occurred among all patients admitted (preventable deaths divided by all admissions). Finally, after adjusting for differences in the severity of illness (see Appendix), we compared the estimated rate of preventable death in the high- and low-outlier hospitals.

Reliability of the Subjective Process-of-Care Review

Each patient's summary was rated by at least three physicians. The interrater reliabilities for preventable deaths were 0.55, 0.51, and 0.11, respectively, for cerebrovascular accident, myocardial infarction, and pneumonia. To assess intrarater reliability, a sample of dictations was rated twice by each physician, and there was 69 percent agreement in the 45 sets of ratings (31 of 45). Finally, in order to determine whether the dictations had insufficient or biased information, the physicians reviewed a series of photocopied medical records. We compared individual physicians' responses when they used the medical records with their responses when they used the dictated summaries of the same cases and found a rate of agreement of 84 percent (16 of 19).

RESULTS

Age- and Sex-Adjusted Death Rates

The unadjusted condition-specific death rates were higher in the high-outlier hospitals than in the low-outlier hospitals by a factor of 1.8 to 2.2. For example, the raw mortality rate among patients with cerebrovascular accident was 22 percent in the high-outlier hospitals, as compared with 10 percent in the low-outlier hospitals (Table 1). Differences in the age and sex distribution between high- and low-outlier hospitals did not explain this underlying disparity in death rates. For instance, grouping the high- and low-outlier hospitals together produced a mean observed death rate of 15.4 percent among patients with cerebrovascular accident. Adjustment for the age and sex distribution in each outlier did not significantly influence these results (high outliers had a predicted death rate of 15.7 percent, and low outliers had a rate of 15.2 percent; Table 1). Adjustments for age and sex among patients with pneumonia and myocardial infarction produced similar results.

Errors in Diagnostic Coding

The medical record did not satisfy the diagnostic criteria for the diagnosis coded on the face sheet in 20 percent of the cases (74 of 378). This error rate did not differ significantly between the high- and low-outlier hospitals (high outliers, 19 percent; low outliers, 20 percent; $P > 0.1$). In addition, among patients who died in the hospital, 16 percent (17 of 106) in the high-outlier group and 20 percent (15 of 76) in the

Table 1. Age- and Sex-Adjusted Predicted Death Rates in High- and Low-Outlier Hospitals.

CONDITION	UNADJUSTED DEATH RATE	COMBINED DEATH RATE	PREDICTED DEATH RATE*
	%	%	% (CI†)
Cerebrovascular accident			
High	22	15.4	15.7 (15.3-16.1)
Low	10		15.2 (14.9-15.5)
Pneumonia			
High	18	14.6	15.1 (14.4-15.8)
Low	9		14.2 (13.4-15.0)
Myocardial infarction			
High	21	17.8	17.3 (16.3-18.3)
Low	12		18.3 (17.3-19.3)

*Adjusted for the ages and sex of patients.

†CI denotes 95 percent confidence interval (± 1.96 SE).

low-outlier group had incorrectly coded diagnoses ($P > 0.1$). Among patients discharged alive, 22 percent (25 of 114) and 21 percent (17 of 82) in the high- and low-outlier groups, respectively ($P > 0.1$), had incorrectly coded diagnoses. These data suggest that differences in rates of coding errors did not explain the large discrepancies in mortality rates between the high- and low-outlier hospitals.

Case Complexity

Table 2 compares the average case in the high-outlier hospitals with the average case in the low-outlier hospitals using various indicators of illness severity. For all three conditions, the APACHE II score (indicating a more severely ill patient population) in the high outliers significantly exceeded the score in the low outliers.

Table 3 shows that models of disease complexity closely predicted the actual death rate in the outlier hospitals for each of the three conditions. For cerebrovascular accident, for example, the model predicted a death rate of 20.8 percent for the high-outlier group, whereas the actual death rate was 22 percent. The model performed similarly well when applied to the low outliers (actual death rate, 10 percent; predicted death rate, 11.7 percent). However, after adjustment for case complexity, there remained unexplained mortality in the high-outlier hospitals and an unexplained absence of mortality in the low outliers. For each condition, the ratio of the actual to the predicted death rate in the high outliers exceeded the ratio in the low outliers. This result was statistically significant ($P < 0.01$) across all three conditions, according to Fisher's method of combining P values.¹⁷

Quality of Care

Structured (Explicit) Review

We compared the high- and low-outlier hospitals in terms of 31 weighted process-of-care scales — 11 for pneumonia and 10 each for cerebrovascular accident and myocardial infarction. The results for high- and low-outlier hospitals did not significantly differ on 26 of the 31 scales (Table 4). The results in Table 4 are

not much different from what would have been expected with the hypothesis that there was no difference in the average process score between high- and low-outlier hospitals.

Subjective (Implicit) Review

The expert panels determined that 18 percent (9 of 50) of the deaths due to cerebrovascular accident, 24 percent (17 of 70) of those due to pneumonia, and 37 percent (23 of 62) of those due to myocardial infarction were possibly preventable. The proportion of deaths that were preventable did not differ significantly between the high- and low-outlier hospitals among patients with cerebrovascular accidents and myocardial infarctions. However, among patients with pneumonia, the expert panels found a significantly higher proportion of deaths that were possibly preventable in the six high-outlier hospitals (32 percent, or 14 of 44) than in the six low-outlier hospitals (12 percent, or 3 of 26; $P < 0.05$) (Table 5).

After adjustment for differences in severity and for the fact that we oversampled deaths, we found the following. First, among all patients admitted with myocardial infarction, the estimated rate of preventable death did not differ significantly between the high and low outliers (we estimate that 5.7 percent of a standard cohort of patients admitted to the high-outlier hospitals with a myocardial infarction would have a preventable death, as compared with 6.7 percent admitted to the low outliers; $P > 0.1$). Second, for pneumonia and cerebrovascular accident and for all three conditions combined, we estimated a significantly higher rate of preventable death in the total patient population with those conditions in the high-outlier hospitals (pneumonia — high outliers 5.5 percent, low outliers 1.0 percent; cerebrovascular accident — high

Table 2. Average Clinical Findings in Three Conditions in High- and Low-Outlier Hospitals.

CLINICAL FINDINGS	HIGH OUTLIERS	LOW OUTLIERS
Cerebrovascular accident		
APACHE II score*	14.9†	12.1
Glasgow coma score*	2.8‡	1.5
Heart rate	89	83
Age	78	75
Body-system score§	1.7	1.3
Mass effect on head CT scan* (prevalence)	0.09‡	0.02
Pneumonia		
APACHE II score*	16.2†	13.4
Heart rate	99†	87
Respiratory rate	28†	24
Age	76	75
Body-system score§	1.9	1.7
Myocardial infarction		
APACHE II score*	14.8†	11.4
Heart rate	85†	93
Respiratory rate	21	20
Age	72	71
Killip class*	1.35	1.34
Body-system score§	1.4	1.5

*A higher score indicates a greater severity of illness at the time of admission.

†Significant at $P < 0.05$.

‡Significant at $P < 0.10$.

§Number of body systems affected by the patient's comorbid conditions.

Table 3. Hospital Death Rates in High- and Low-Outlier Facilities, Adjusted for Complexity and Severity of Illness.

CONDITION	OUTLIER	ACTUAL DEATH RATE	PREDICTED DEATH RATE*	ACTUAL/PREDICTED DEATH RATE
		%	% (CI†)	
Cerebrovascular accident	High	22	20.8 (17.3–24.3)	1.05 (0.73–1.37)
	Low	10	11.7 (5.9–17.5)	0.85 (0.51–1.19)
Pneumonia	High	18	17.5 (15.9–19.1)	1.03 (0.93–1.13)
	Low	9	10 (7.7–12.3)	0.90 (0.7–1.1)
Myocardial infarction	High	21	19 (16.9–21.1)	1.10 (0.96–1.24)
	Low	12	13.9 (11.9–15.9)	0.86 (0.74–0.98)

*The predicted death rate was derived from a logistic-regression model based on the comorbidity and severity of illness of the patients.

†CI denotes confidence interval (± 1.96 SE).

outliers 5.8 percent, low outliers 1.2 percent; and all three conditions — high outliers 5.7 percent, low outliers 3.2 percent; all differences were significant at $P < 0.05$).

DISCUSSION

We have previously described a model for adjusted hospital death rates that used computerized discharge data.¹² That model identified hospitals in which the actual death rate differed greatly from the expected death rate. In this follow-up study, we found that most but not all of these differences disappeared after better adjustment for the patients' severity of illness. Using process criteria, we could not identify any systematic differences in the quality of care between the high- and low-outlier facilities. However, when we used subjective reviews by expert clinicians, we projected a significantly higher rate of "preventable" deaths in the group of hospitals with higher-than-expected death rates. This paper does not address whether it was the physicians, the nurses, or other providers of health care within the hospital that could have prevented the deaths.

Death rates have begun to receive increasing attention as potential measures of hospital performance. The Health Care Financing Administration recently compiled a list of hospitals in which the actual death rate greatly exceeded the expected one.¹⁸ This information was initially intended for use by peer-review organizations to help them locate hospitals in need of careful review. In addition, consumer groups found this list potentially valuable as a criterion with which to compare hospitals.

Studies of hospital death rates use computerized data bases and routinely collected data to build models that adjust a hospital's death rate for its own mix of patients. Unfortunately, except for diagnosis, these data bases contain few if any clinical details. They provide information about the patients' age and sex, but lack information about findings on physical examinations, laboratory results, or the results of radiologic procedures. Thus, previous reports could not really determine whether hospitals with higher-than-expected death rates merely cared for a very sick patient population.^{3-8,10,11,19}

Table 4. Relation of Hospital Outlier Status to the Quality of Care for Three Conditions, as Judged According to Explicit Process Scales.*

PROCESS SCALE	CEREBROVASCULAR ACCIDENT	PNEUMONIA	MYOCARDIAL INFARCTION
Admission history	H	L†	H
Admission physical examination	H	L	H
Admission laboratory tests	L†	H	L
Initial therapy	L	L	H
Choice of antibiotic	NA	H	NA
Management of new symptoms	H	L	H
Management of abnormal laboratory values	H	L	L
Adverse-outcome scales‡			
Laboratory intensity	H†	H	L
Treatment intensity	H	H†	L
Follow-up of abnormal laboratory results	H†	H	L
Doctor-nurse interaction	H	H	L

*H indicates that high-outlier hospitals had better process scores than low-outlier hospitals, L indicates that low-outlier hospitals had better process scores than high-outlier hospitals, and NA denotes not applicable.

†Statistically significant at P<0.05.

‡Adverse-outcome scales apply only to patients who died and reflect the process of care during the 24 hours before death.

We have tried to address this issue by reviewing patients' medical records from a set of outlier hospitals in which the actual death rate differed greatly from the rate predicted by our demographic and diagnostic model. We chose to review cases of cerebrovascular accident, pneumonia, or myocardial infarction because they accounted for a large fraction (36 percent) of the deaths in the outlier hospitals.

The medical-records review tabulated the patients' comorbid conditions and quantified the severity of their illnesses at the time of hospital admission. We used this information to build a more clinically accurate model. The predictions from this model closely matched the actual hospital death rates (Table 3). These results suggest that the hospitals in this study that had higher death rates than expected cared for a sicker group of patients than the hospitals with rates that were lower than expected (Table 2).

Table 5. Preventable Deaths in the High- and Low-Outlier Hospitals, According to Medical Condition.

OUTLIER	CONDITION			
	CEREBROVASCULAR ACCIDENT	PNEUMONIA	MYOCARDIAL INFARCTION	ALL 3 CONDITIONS
High				
Rate of preventable death among patients who died	6/25 (24%)	14/44 (32%)*	11/37 (30%)	31/106 (29%)
Estimated rate of preventable death in total patient population†	5.8%*	5.5%*	5.7%	5.7%*
Low				
Rate of preventable death among patients who died	3/25 (12%)	3/26 (12%)*	12/25 (48%)	18/76 (24%)
Estimated rate of preventable death in total patient population†	1.2%*	1.0%*	6.7%	3.2%*

*The difference between the values in the high- and low-outlier hospitals was significant at P<0.05.

†This rate reflects the probability of a preventable death among all patients admitted with this condition (adjusted for severity of illness).

However, this "clinical" model did not account for all the differences in death rates between the high- and low-outlier hospitals. For each condition, the actual death rate exceeded the rate predicted on the basis of the case complexity at the high-outlier hospitals by 3 to 10 percent and fell short of the predicted rate at the low-outlier hospitals by 10 to 15 percent (Table 3). Thus, after we used substantial clinical information to control for the severity of illness, an excess of mortality at the high-outlier hospitals and a shortfall at the low outliers of about 10 percent remained unexplained (P<0.01). We have considered four explanations for these differences: variations in diagnostic coding, statistical selection effects, unexplained severity, and differences in the quality of care.

First, hospitals could have death rates that are higher or lower than expected because of differences in the way diagnostic labels were used by either physicians or medical-records personnel. For example, a hospital would have an artificially low rate of death from pneumonia if it included in that diagnostic category patients who actually had a more benign illness (such as bronchitis). Similarly, the inclusion of patients with transient ischemic attacks in the cerebrovascular-accident category would lower the death rate.

We looked for this phenomenon by using a set of standard diagnostic criteria prepared by a panel of clinical experts. We found that the patients' primary diagnosis often did not satisfy the panel's criteria (18 to 20 percent of the time). However, there was no difference in the rate of misclassification between the high- and low-outlier hospitals, suggesting that hospitals did not become outliers merely because of systematic differences in diagnosis or coding.

Second, statistical selection effects may be responsible for some of the observed discrepancy. We chose our 11 high-outlier and 9 low-outlier hospitals not because they were representative but precisely because they had extreme total mortality rates after hospital-level adjustment for demographic characteristics and diagnoses. It is possible that some of the excess mortality in the high-outlier hospitals and shortfall of mortality in the low-outlier facilities was a consequence of this selection process. However, primarily because the selection of the outlier hospitals was based on the total mortality for all conditions rather than mortality or preventable death for the three specific conditions we studied (which represented only 36 percent of the deaths), we think that only a relatively small fraction of the disparity in mortality and preventable death rates was due to selection effects (further details are available from us).

Third, even though we used sev-

eral separate measures of severity for each condition, the observed difference in death rates may still reflect unmeasured severity. In our judgment, however, after a first adjustment for demographic characteristics and diagnoses at the hospital level and a further accounting for the severity of illness at the patient level based on detailed clinical information, any differences in severity that remain are probably small.

Fourth, the persistent differences in death rates could also reflect differences in the quality of care provided. Hospitals with higher-than-expected death rates may deliver care that meets a lower standard than other hospitals. Using process criteria, we found no systematic differences in the care provided by physicians or nurses in the two groups of outlier hospitals.

The study also used clinical experts' subjective judgments about the quality of care. We prepared dictated summaries of the hospital course of a sample of patients who died. At least three physicians who were blinded to the hospital's death rate read every summary independently and judged whether the death might have been prevented. For patients with myocardial infarction, we estimated no difference in the preventable-death rate between the two outlier groups. However, for patients with cerebrovascular accident, pneumonia, and all three conditions combined, we estimated a significantly higher rate of preventable deaths in the high-outlier hospitals. Thus, of 100 patients admitted with pneumonia to a low-outlier hospital, about 1 would be expected to have a preventable death. Alternatively, if a high-outlier hospital admitted the same 100 patients, about 5 would be expected to have a preventable death.

In summary, we found differences in quality between the high and the low outliers when we used clinicians' subjective assessments, but found no significant difference for any condition when we used structured process criteria. This discrepancy has several potential explanations.

First, the increased rate of "preventable" deaths in the high-outlier hospitals could reflect the methods used in the subjective review. The clinical experts read a dictated abstract of the hospital course prepared by the study staff. A biased dictation could have influenced the interpretation by the reviewers. However, the dictation followed a precise format. In addition, we looked for bias by comparing the reviewers' responses when they used the dictated summary with their responses when they used the actual medical record. We found adequate reliability and no overt bias.

Alternatively, the discrepancy between the structured and subjective reviews of quality may reflect differences in the emphasis of these two techniques. Structured review stresses compliance with specified criteria, whereas a subjective review allows physicians to use their judgment in an unspecified manner. Other investigators have found similar discrepancies between structured and subjective process reviews of the

same case.²⁰⁻²² Their reports make our results somewhat less surprising.

Finally, our structured approach may have lacked adequate sensitivity to detect a difference. One patient died soon after a feeding tube was inadvertently introduced into the right lung instead of the stomach. The structured review had only 125 criteria and did not have one that dealt with such a case. In other cases, the preventable deaths may have occurred in patients who became hypoxic and did not receive a timely workup or treatment. In these instances, the patients had become lethargic or confused or their complexion had become dusky. Our structured criteria defined the need for intervention on the basis of more precise clinical features, such as a rapid respiratory rate, which may not have occurred or may not have been documented. Thus, by their very nature, the structured and subjective reviews may differ in their ability to identify certain errors in performance.

Given the small scope of this study, we cannot hope to provide the definitive answer regarding the usefulness of demographic and diagnosis-adjusted death rates in comparing hospitals. The original data base had information about only 93 hospitals, and these facilities may not fully reflect the more than 5500 short-term general care hospitals in this country. In addition, we reviewed only 378 medical records of patients with three conditions treated in 12 outlier hospitals. Our findings should be viewed as preliminary. Before an issue of this magnitude can be settled, policy makers will need data from more hospitals, more conditions, and more patients.

The adjusted death-rate model identified outlier hospitals that as a group had both sicker patients and more possibly preventable deaths. However, the size of our study did not allow us to discern whether this high-outlier group really contained two distinct subgroups of hospitals — one with very ill patients and adequate quality and another with less ill patients and inadequate quality. Thus, the method for deriving adjusted death rates presented here should be viewed as a screening tool. As such, it can identify a high-outlier hospital that should undergo a detailed review. Only such a review can determine whether the hospital's medical or nursing staff provides poor care or whether the hospital has disproportionately ill patients.

The Rand Corporation
Santa Monica, CA 90406

MetaMedical, Inc.
Los Angeles, CA 90212

The Rand Corporation
Santa Monica, CA 90406

Center for Health Sciences,
University of California
at Los Angeles
Los Angeles, CA 90024

ROBERT W. DUBOIS, M.D., PH.D.
WILLIAM H. ROGERS, PH.D.

JOHN H. MOXLEY III, M.D.

DAVID DRAPER, PH.D.

ROBERT H. BROOK, M.D., Sc.D.

Supported by grants from American Medical International, Beverly Hills, Calif.; the Robert Wood Johnson Foundation, Princeton, N.J.; and the National Center for Health Services research (1 T32 HS-00007-02). The opinions, conclusions, and proposals are

those of the authors and do not necessarily represent the views of the Robert Wood Johnson Foundation or American Medical International.

We are indebted to Katherine Kahn, M.D., and Lisa Rubenstein, M.D., of the Rand Corporation, and Jacqueline Kosecoff, Ph.D., of Fink and Kosecoff, Inc., who developed the list of comorbid conditions and the preliminary list of explicit process criteria.

APPENDIX

Diagnostic Criteria

For cerebrovascular accident, the criteria included all of the following: acute onset (<12 hours), neurologic deficit lasting at least 24 hours, and a presumed vascular cause.

For myocardial infarction, the criteria were a newly documented Q wave on electrocardiography or an elevated creatine kinase level (MB >5 percent or a creatine kinase level above normal, which followed the expected progression).

For pneumonia, the criteria were an infiltrate on a chest x-ray film and at least one of the following: fever (temperature >38.3°C), dyspnea, leukocytosis (white-cell count >18,000), increased sputum, positive sputum cultures, or neutrophils on the patient's sputum Gram stain.

Handling of Missing Data

For the comorbidity scales, we assumed that a patient did not have a condition unless the chart mentioned its presence. For the severity-of-illness measures, we computed two scale scores. The first score excluded patients with missing data. The second score used the median value from the remainder of the data set for any missing elements. The basic conclusions described here were not affected by the method used. We report the results using the second approach. For the structured process criteria, we assumed noncompliance with each criterion unless the chart documented otherwise.

Dictation Format

The dictations included a verbatim transcription of the attending physician's description of the admitting history and physical examination, followed by the results of all laboratory values on admission, radiologic examinations, electrocardiograms, and any other special studies. The dictation next documented the attending physician's admission assessment and the initial therapy.

For each subsequent hospital day, the dictation summarized the vital signs, the progress notes made by physicians, and pertinent nursing notes. The summary included all new laboratory studies and changes in therapy. These daily summaries documented in detail any new changes in the patient's clinical status (abnormal vital signs, new symptoms, and abnormal laboratory values) and the response by health care providers (nurses, physicians, and other staff members). The hospital summary became more detailed during the 24 hours before the patient's death.

For patients with long stays (>10 days), the daily summary notes were different; the dictation might state, "CBC [complete blood count] and ECG [electrocardiography] performed today and the results were unchanged," rather than provide a more complete description. If any new results differed greatly from previous ones, the summary included complete data.

The summaries contained no identifying information in regard to the hospital or its geographic location, the patients, or the staff members.

Adjustment of Preventable Death Rates for Case Complexity

The complexity results showed that the high- and low-outlier hospitals cared for patients with very different degrees of severity of illness. This difference could influence the observed rate of preventable death (preventable deaths/all admissions) in the outlier hospitals. Therefore, the relevant comparison is not based on the observed rate of preventable death but on estimates of what the rate would have been in the high- and low-outlier hospitals had they cared for groups of patients with a similar severity of

illness. The adjustment process yielding these estimates proceeded in five steps.

First, we calculated a predicted probability of death for each patient in the sample (either discharged alive or died in the hospital). This calculation used data on the severity of illness on admission and the logistic regression models described under Methods. Second, we reweighted the data to reflect the original sampling method. Third, for each outlier (high and low) and according to each condition (cerebrovascular accident, pneumonia, and myocardial infarction), we divided the patients into severity-of-illness strata on the basis of their probability of death at the time of admission (probability of death <0.4, 0.4–0.6, or >0.6). Fourth, we determined the preventable death rate within each of these 18 strata (outlier, condition, and severity; $2 \times 3 \times 3 = 18$). Fifth, using indirect standardization, we collapsed the information from the strata first into rates of preventable death for each condition (according to outlier) and then into rates of preventable death for each group of outlier hospitals.

Statistical comparisons used standard errors based on a Poisson approximation to the binomial distribution, adjusted for the sampling method. A more detailed description of this adjustment process is available from the authors.

REFERENCES

1. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part I: the relation between hospital volume and outcomes for selected diagnostic categories. *Med Care* 1984; 22:98-114.
2. *Idem*. Does practice make perfect? Part II: the relation between volume and outcomes and other hospital characteristics. *Med Care* 1984; 22:115-25.
3. Maerki SC, Luft HS, Hunt SS. Selecting categories of patients for regionalization: implications of the relationship between volume and outcome. *Med Care* 1986; 24:148-58.
4. Luft HS, Hunt SS. Evaluating individual hospital quality through outcome statistics. *JAMA* 1986; 255:2780-4.
5. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. *N Engl J Med* 1979; 301:1364-9.
6. Luft HS. The relation between surgical volume and mortality: an exploration of causal factors and alternative models. *Med Care* 1980; 18:940-59.
7. Showstack JA, Rosenfeld KE, Garnick DW, Luft HS, Schaffarick RW, Fowles J. Association of volume with outcome of coronary artery bypass graft surgery: scheduled vs nonscheduled operations. *JAMA* 1987; 257:785-9.
8. Wennberg JE, Roos N, Sola L, Schori A, Jaffe R. Use of claims data systems to evaluate health care outcomes: mortality and reoperation following prostatectomy. *JAMA* 1987; 257:933-6.
9. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med* 1986; 104:410-8.
10. Roemer MI, Moustafa AT, Hopkins CE. A proposed hospital quality index: hospital death rates adjusted for case severity. *Health Serv Res* 1968; 3:96-118.
11. Duckett SJ, Kristofferson SM. An index of hospital performance. *Med Care* 1978; 16:400-7.
12. Dubois RW, Brook RH, Rogers WH. Adjusted hospital death rates: a potential screen for quality of medical care. *Am J Public Health* 1987; 77:1162-6.
13. Harrell F. The LOGIST procedure. In: SUGI supplemental library user's guide. Cary, N.C.: SAS Institute, 1983.
14. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system for severely ill patients. *Crit Care Med* 1985; 13:818-29.
15. Braunwald E. Heart disease: a textbook of cardiovascular medicine. 2nd ed. Philadelphia: W.B. Saunders, 1984.
16. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat* 1979; 7:1-26.
17. Wallis WA. Compounding probabilities from independent significance tests. *Econometrica* 1942; 10:229-48.
18. Brinkley J. U.S. releasing lists of hospitals with abnormal mortality rates. *New York Times*. March 12, 1986:1.
19. Moses LE, Mosteller F. Institutional differences in postoperative death rates: commentary on some findings of the National Halothane Study. *JAMA* 1968; 203:492-4.
20. Hulka BS, Romm FJ, Parkerson GR Jr, Russell IT, Clapp NE, Johnson FS. Peer review in ambulatory care: use of explicit criteria and implicit judgments. *Med Care* 1979; 17:Suppl 3:1-73.
21. Kane RL, Rubinstein LZ, Brook RH, et al. Utilization review in nursing homes: making implicit level-of-care judgments explicit. *Med Care* 1981; 19:3-13.
22. Brook RH, Appel FA. Quality-of-care assessment: choosing a method for peer review. *N Engl J Med* 1973; 288:1323-9.

RAND/N-3372-HHS