A RAND NOTE

Explaining Variations in Hospital Death Rates: Randomness, Severity of Illness, Quality of Care

Rolla Edward Park, Robert H. Brook,
Jacqueline Kosecoff, Joan Keesey,
Lisa Rubenstein, Emmett Keeler,
Katherine L. Kahn, William H. Rogers,
Mark R. Chassin

July 1990

RAND

The research described in this report was sponsored by the Health Care Financing Administration, U.S. Department of Health and Human Services under Cooperative Agreement No. 17-C-98684/9-04.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

A RAND NOTE

N-3196-HCFA

Explaining Variations in Hospital Death Rates: Randomness, Severity of Illness, Quality of Care

Rolla Edward Park, Robert H. Brook,
Jacqueline Kosecoff, Joan Keesey,
Lisa Rubenstein, Emmett Keeler,
Katherine L. Kahn, William H. Rogers,
Mark R. Chassin

July 1990

Prepared for the
Health Care Financing Administration,
U.S. Department of Health and Human Services

RAND

RAND/UCLA/HARVARD
Center for Health Care
Financing Policy Research

Explaining Variations in Hospital Death Rates

Randomness, Severity of Illness, Quality of Care

Rolla Edward Park, PhD, MBA; Robert H. Brook, MD, ScD; Jacqueline Kosecoff, PhD; Joan Keesey; Lisa Rubenstein, MD, MPH; Emmett Keeler, PhD; Katherine L. Kahn, MD; William H. Rogers, PhD; Mark R. Chassin, MD, MPP, MPH

We used administrative (Part A Medicare) data to identify a representative sample of 1126 patients with congestive heart failure and 1150 with acute myocardial infarction in hospitals with significant unexpectedly high inpatient, age-sex-race-disease–specific death rates ("targeted") vs all other ("untargeted") hospitals in four states. Although death rates in targeted hospitals were 5.0 to 10.9 higher per 100 admissions than in untargeted hospitals, 56% to 82% of the excess could result from purely random variation. Differences in the quality of the process of care (based on a medical record review) could not explain the remaining statistically significant differences in mortality. Comparing targeted hospitals with subsets of untargeted ones, eg, those with lower than expected death rates, did not affect this conclusion. Severity of illness explained up to 2.8 excess deaths per 100 admissions for patients with myocardial infarction. Identifying hospitals that provide poor-quality care based on administrative data and single-year death rates is unlikely; targeting based on time periods greater than 1 year may be better.

(*JAMA*. 1990;264:484-490)

IT WOULD be convenient if hospitals with higher death rates, identified by using easily collected administrative data (age, sex, previous hospitalization, diagnosis), turned out to be providing lower quality of care. It is easy to use administrative data to identify hospitals with high death rates. If a high death rate were a marker for bad care, then health care consumers would know to avoid those hospitals, and professional organizations and the hospitals themselves could work to correct the quality problems.

Apparently in the hope or belief that high death rates and low quality of care are associated, the Health Care Financing Administration (HCFA) has, annually since 1986, released, at an individual hospital level, increasingly sophisticated analyses of hospital death rates for Medicare patients (*New York Times*, March 12, 1986:1).[1-3] The release of the analyses has been criticized,[4-6] but HCFA has taken many of the criticisms into account. Even HCFA's critics seem to share the hope that sufficiently sophisticated analyses will succeed in targeting hospitals that provide substandard care.

Variations in hospital death rates have been studied for a long time, but such studies are few in number and, until the release of data by HCFA, have not been performed to identify individual hospitals as possibly providing poor quality of care.[7] Death rates have been shown to vary by specific hospital characteristics[8-10] and by experience (ie, volume),[8-13] but we do not know a lot about how much of the variation results from differences in severity of illness, how much from differences in quality of care, and how much from purely random or selection effects.

Some evidence has accumulated, from studies in limited numbers of hospitals, that some of the differences in death rates among hospitals may be due to differences in severity of illness or level of comorbidity. This was true for patients with pneumonia, myocardial infarction, or stroke in a single large hospital chain,[14] for patients with cancer in seven hospitals,[4] and for patients in nine pediatric intensive care units.[15]

One recent study of four common medical conditions in a Medicare population found that chance variation accounts for a major part of the differences in hospital death rates, but that severity measures based on data obtained from a medical record review also helped to explain the differences.[16] Another study of five common conditions in 13 hospitals found that severity measured from the medical record added substantially to the explanatory power of HCFA's 1988 model and reduced instances of higher than expected mortality of chance levels.[17]

Only one study has shown some connection between high death rates and quality of care. Using implicit peer review of quality of care in a single hospital chain, that study showed that patients with pneumonia, stroke, or myocardial infarction were twice as likely to suffer a possibly preventable death in high-death outlying hospitals than were patients in hospitals that were not statistical outliers.[14]

We previously found for all US acute-care hospitals that age-sex-race-disease–specific death rates were significantly different (both clinically and statistically) by hospital for 22 of 48 specific conditions or diagnoses.[18] Because our previous study used data from hospital claims only, we were unable to address the question of how to explain the remaining systematic variation. This study attempts to shed some additional light on the relationships among hospital death rates, severity of illness, and quality of care by using data from medical records. We chose two medical conditions, congestive heart failure (CHF) and acute myocardial infarction (AMI), for more detailed clinical investigation because their death rates varied significantly by hospital and they accounted for 7.5% of all Medicare admissions and 17.5% of all Medicare hospital deaths.

Our primary objectives for this study were to determine in a representative sample of acute-care hospitals (1) whether hospitals with high age-sex-race-disease–specific death rates provide lower quality of care or treat more

From the Health Sciences Program at The RAND Corporation (Drs Park, Brook, Rubenstein, Keeler, Kahn, and Rogers and Ms Keesey) and Value Health Sciences (Drs Kosecoff and Chassin), Santa Monica, Calif; and the Departments of Medicine (Drs Brook, Rubenstein, and Kahn) and Public Health (Drs Brook and Kosecoff), University of California at Los Angeles.

The opinions, conclusions, and proposals herein are those of the authors and do not necessarily represent the views of the Health Care Financing Administration, The RAND Corporation, or the University of California.

Reprint requests to The RAND Corporation, 1700 Main St, Box 2138, Santa Monica, CA 90406 (Dr Park).

severely ill patients than do hospitals with lower death rates, and (2) how the probability of death is related to severity of illness and quality of care. Answers to these questions could result in better policy decisions regarding whether the public identification of poor-quality hospitals requires collecting more data (eg, severity of illness at time of admission) than those available on a discharge abstract. We summarize our methods and results herein; additional details are available elsewhere.[19]

METHODS

Administrative Data

We obtained information on all hospital stays for Medicare beneficiaries from HCFA's Bill Record File for all admissions occurring between October 1, 1983, and September 30, 1984. To make the data as comparable as possible across hospitals, we (1) excluded all Medicare beneficiaries under the age of 65 years (those eligible to receive Medicare benefits solely because of various disabilities, including chronic renal disease); (2) excluded data from long-term–care hospitals, psychiatric facilities, hospices, and rehabilitation hospitals; (3) excluded interim bills; (4) edited the data to include only one complete record for each hospital stay; and (5) counted transfers from one acute-care hospital to another as live discharges from the first hospital and separate admissions to the second.

We obtained additional information on hospital characteristics from HCFA's Provider of Service File and information on out-of-hospital deaths from HCFA's Health Insurance Master File. We used only the principal diagnosis and defined CHF according to *International Classification of Diseases, Ninth Revision*, code 398.91, 402.11, 402.91, 428.0, 428.1, 428.9, or 785.51 and AMI according to codes 410.0 through 410.9.

Targeting

For each acute-care US hospital, we calculated the death rate it would have experienced if its patients with CHF or AMI had died at nationwide average rates for each condition for each of 20 age-sex-race cells. We then calculated the binomial probability that a hospital would have as many deaths as it actually did. Hospitals with less than a .05 probability of having as many deaths as they did were called "targeted"; all others were "untargeted." In simple terms, we targeted using a one-sided test at the .05 level. We targeted separately for each condition (CHF and AMI) and separately for inpatient deaths and for deaths within 30 days of admission.

Sampling

For logistic reasons, we confined the sample to four states (California, Illinois, Minnesota, and New York), which together had 20% of US hospitals and 22% of Medicare hospitalizations. Power calculations showed that a sample of 350 patients in each of the four targeted/untargeted dead/alive cells for each of the two conditions would be adequate. For a quality of care measure with an SD of 1.0, we could expect to detect a 0.15 point difference in average quality between targeted and untargeted hospitals, or between dead and alive discharges, in 80% of repeated samples, using a one-tailed test at the .05 significance level.

We drew a systematic random sample of discharges in the following manner. Our sample frame consisted of Medicare claims records arranged into eight lists. There was a separate list for each condition for each of its four cells: targeted or untargeted hospitals based on inpatient deaths and dead or alive patients at time of discharge; eg, one cell was patients with heart failure discharged dead from inpatient targeted hospitals. We sorted each list by state and hospital; within each hospital, we listed patients in random order and sampled systematically from that list.

We sampled based on inpatient deaths because that is what HCFA was using for its mortality data release at the time we drew the sample. Subsequently, HCFA shifted to analyzing 30-day deaths. Is our sample useful for analyzing 30-day deaths as well? The answer is "yes." Analyzed using appropriate population weights, our sample yields unbiased estimates of differences between 30-day targeted and untargeted hospitals, albeit with slightly higher variance than the inpatient estimates.[19]

Simulation

To determine how much of the variation in the hospitals' death rates for patients with CHF and AMI could be attributed to random variation, we simulated hospital deaths in the four study states on the null hypothesis that the probability of death for each hospital was the age-sex-race–standardized value described under "Targeting" above. We calculated the binomial probability of having as many deaths as simulated and ranked hospitals in order of that probability. We counted the hospitals with the lowest probability values as targeted in the simulations, using the number of hospitals that were targeted when using actual data as the cutoff number. This number differed for CHF and AMI and for inpatient or 30-day targeting. We calculated death rates in the simulated targeted hospitals as a group and in simulated untargeted hospitals as a group. We repeated the process 100 times and averaged the simulated death rates over the 100 repetitions.

Medical Records Data

To collect detailed data on severity of illness and quality of care, we developed separate detailed abstraction forms for the two conditions[20,21] and contracted with local peer review organizations in the four states to do the abstraction. We trained nurses and medical records abstractors in the use of the abstraction forms. The peer review organizations asked the hospitals to send them complete photocopies of the sampled records. The records were abstracted at each peer review organization and the completed abstraction forms were sent to RAND, where selected items were reviewed first by a nonphysician to ensure completeness, legibility, and internal consistency of the abstracted data. All of the abstracts were then reviewed by the physician principal investigator (M.R.C.), who determined for each case whether the principal diagnosis of CHF or AMI was accurately coded and verified other exclusionary criteria. We excluded patients if surgery occurred during the hospital stay, if the patient had metastatic cancer or cancer under active treatment with radiation or chemotherapy, or if the patient had been transferred from another acute-care hospital. We also excluded patients for which the principal diagnosis of CHF or AMI was coded incorrectly.

We used the abstracted data to calculate disease-specific measures of severity of illness and quality of care. The measures are those developed by the RAND prospective payment system study.[22] The severity measure is a weighted sum of APACHE[23] and other items, including rescaled systolic blood pressure, the results of laboratory tests, and an inventory of chronic morbid and comorbid disease markers. It has been shown to predict 12% of the variance in deaths for patients with CHF and 22% for those with AMI.[24] The quality score measures the process of care based on an explicit set of processes that should be done, including physician and nurse examination and history taking and the use of diagnostic, therapeutic, and intensive services. It is branched, ie, different criteria apply to different patients. It is disease specific and standardized to reflect differing levels in difficulty of complying with a criterion. Most important, it has been shown to be valid at the patient level, ie,

increased scores on this process scale indicate lower probability of death.[25]

Analysis

We used Cox proportional hazards regression to explore the effect of differences in severity and quality on probability of death. We estimated separate equations for the two conditions and for inpatient and 30-day deaths. We report unweighted (ie, not corrected for oversampling deaths and targeted hospitals) Cox proportional hazards regressions herein, but only after comparing separate logistic regressions for targeted and untargeted hospitals and confirming that there are no substantial differences in the estimated coefficients.[19] We compared mean values of quality of care, severity of illness, and "do not resuscitate" status (DNR) in targeted vs untargeted hospitals; these means were reweighted to reflect population values. We used the estimated Cox regressions to calculate the differences in average death rates corresponding to the observed differences in quality, severity, and DNR.

Retargeting

To understand whether our results are sensitive to the targeting method chosen (ie, hospitals with $P<.05$ of having as many deaths as observed vs all others), we changed how we classified hospitals and thereby tested other comparisons. First, we compared only the "best" and "worst" hospitals in our original sample. We defined the best hospitals as those that had lower than expected death rates (not necessarily significantly lower), thus excluding small hospitals with high death rates but too few patients to achieve statistical significance, as well as larger hospitals with moderately high death rates. We defined the worst hospitals as those with $P<.01$ of having as many deaths as observed. We retargeted best and worst hospitals, using both inpatient and 30-day death rates.

We also took advantage of HCFA's 1988 analysis of 1986 hospital data, which, although also based on administrative data, includes slightly more adjustment for severity than does the one that we used.[2] We reclassified our hospitals (1984 data) as targeted if they had $P<.05$ of having as many deaths as observed in HCFA's analysis of 30-day deaths during 1986.

Third, we developed an ad hoc 3-year targeting method in an attempt to take advantage of random effects averaging out over time. Specifically, we multiplied together the probabilities that a hospital would have as many deaths as it did from our 1984 thirty-day death analysis and HCFA's 1988 analyses of 1986 and 1987 data. We then ranked hospitals by the result of that computation and counted as targeted the same number of hospitals from the top of the list that our 30-day method targeted in 1984.

Table 1.—Actual and Simulated Death Rates for 1137 Hospitals Treating CHF Patients and 1121 Hospitals Treating AMI Patients*

	CHF Patients		AMI Patients	
	Inpatient Deaths	30-d Deaths	Inpatient Deaths	30-d Deaths
Actual deaths per 100 patients in targeted hospitals	15.4	17.6	30.2	34.1
Actual deaths per 100 patients in untargeted hospitals	7.9	12.6	20.0	23.2
Targeted minus untargeted actual deaths per 100 patients	7.4	5.0	10.2	10.9
Targeted minus untargeted simulated deaths per 100 patients†	5.1 (0.2)	4.1 (0.3)	6.3 (0.5)	6.1 (0.6)
% of actual difference in death rates between targeted and untargeted hospitals due to random variation	69	82	62	56

*Death rates are for hospitals in four states, October 1983 through September 1984. CHF indicates congestive heart failure; and AMI, acute myocardial infarction.
†Simulated values are means from 100 trials, with SDs in parentheses.

RESULTS

Hospital Targeting With Administrative Data

In our four study states, 1137 hospitals had at least one Medicare admission for a patient with CHF and 1121 hospitals had at least one admission for a patient with an AMI. Using our original targeting method (ie, $P<.05$), 13% of hospitals were targeted using inpatient deaths and 7% using deaths within 30 days of admission for patients with CHF; 9% of hospitals were targeted using inpatient deaths and 6% using 30-day deaths for patients with an AMI. Of the hospitals targeted for one condition using inpatient deaths, 22% were also targeted for the other condition. Using 30-day deaths, the overlap was about 17%.

Table 1 shows that death rates in targeted hospitals are substantially higher than those in untargeted hospitals, ranging from 40% higher for 30-day deaths among patients with CHF to almost 100% higher for inpatient deaths among CHF patients. For patients with AMI, targeted hospitals have about 50% higher actual death rates, regardless of whether deaths are counted in the hospital or within 30 days of admission.

The simulation results show how much of the difference in death rates can be attributed solely to the targeting method. Even though differences in death rates in the targeted and untargeted hospitals are statistically significant, random variation and the selection of targeted hospitals account for a large share, between 56% and 82%, of the difference. The remaining nonrandom components of the death rate differences between targeted and untargeted hospitals are both clinically important and highly significant statistically.[18] For example, at 30 days following admission an additional 4.8 deaths per 100 patients admitted with AMI are unexplained after allowing for the way targeted hospitals were selected; the corresponding figure for patients with CHF is 0.9 deaths per 100 patients admitted.

Validating Severity of Illness and Quality of Care Measures

Of the 3200 sampled patients, 32 were from hospitals that we could not identify from HCFA data. We obtained 97% participation by sampled hospitals. Hospital refusals resulted in 2% of patients with CHF and 5% of those with AMI being excluded from our sample (Table 2). On examining the medical record, we found that 248 (16%) of sampled patients with CHF and 231 (14%) of those with AMI had to be excluded because of coding errors; that is, the intended condition was not the true principal diagnosis.

In addition, 92 patients with CHF and 63 with AMI were excluded for other reasons, such as death in the emergency department or claims data errors. Finally, 85 patients with CHF and 64 patients with AMI were excluded because the hospital was unable to locate the sampled admission. We thus obtained complete data on 1126 (90%) of the 1246

Table 2.—Population and Sample Counts by Sampling Category After Sample Attrition*

	CHF Patients		AMI Patients	
	Untargeted Hospitals	Targeted Hospitals	Untargeted Hospitals	Targeted Hospitals
Hospitals				
4 States	992	145	1017	104
Sampled	533	141	525	104
Participating (%)	516 (97)	137 (97)	511 (97)	100 (96)
Patients				
4 States	65 702	16 465	74 844	7322
Sampled	800	800	800	800
Less				
Hospital not identifiable†	0	14	0	18
Hospital refused	23	12	15	59
Coding errors†	124	124	116	115
Claims data errors and other exclusions†	40	52	26	37
Failure to obtain usable copy of sampled record	30	55	21	43
Usable data (% of true records‡)	583 (92)	543 (89)	622 (95)	528 (84)

*CHF indicates congestive heart failure; and AMI, acute myocardial infarction.
†Unavoidable attrition.
‡After excluding unavoidable attrition.

Table 3.—Estimating the Relative Probability of Dying on Any Day Following Hospitalization*

	CHF Patients				AMI Patients			
	Inpatient Deaths		30-d Deaths		Inpatient Deaths		30-d Deaths	
Explanatory Variable	Coeff	t Stat	Coeff	t Stat	Coeff	t Stat	Coeff	t Stat
Severity of illness	5.47	12.1	5.83	12.7	4.58	18.3	4.59	18.5
DNR at admission	0.87	5.9	0.94	6.4	0.34	2.0	0.42	2.4
Quality of care	−0.12	−2.7	−0.10	−2.2	−0.16	−3.4	−0.13	−2.9
After discharge	. . .		−2.62	−11.7	. . .		−2.26	−7.4
No. of observations	1126		1714		1149		1727	

*Model used is Cox proportional hazard model. CHF indicates congestive heart failure; AMI, acute myocardial infarction; DNR, "do not resuscitate"; Coeff, estimated coefficients; and *t* stat, estimated statistics. The higher number of observations for the 30-day estimates is an artifact of the estimation method, which requires replicating observations for patients discharged alive fewer than 30 days after admission to create one observation before and one after discharge.

patients with CHF who were eligible for the study after excluding those ineligible because of claims or coding errors and on 1150 patients (89% of eligibles) with an AMI.

Table 3 summarizes our patient level results on the effect of severity of illness and quality of care on probability of death and helps to establish the validity of the severity and quality measures. Greater severity of illness or DNR status independently increases the risk of death; better quality of care lowers it. For example, for the 30-day CHF model, patients at the 25th percentile of severity and at the median for DNR and quality of care have a predicted 30-day death rate of 9.0 per 100 admissions, while those at the 75th percentile of severity have 17.4 predicted deaths. Correspondingly, patients at the 25th percentile of quality have 13.4 predicted deaths, while those at the 75th percentile of quality have 11.9 predicted deaths.

Comparing Targeted and Untargeted Hospitals

Table 4 compares targeted and untargeted hospitals in terms of average severity of illness, quality of care, and proportion of patients who had DNR orders written at admission. The comparisons are presented for both inpatients and 30-day targeting. There are separate comparisons for dead and alive patients (at discharge for inpatient targeting, at 30 days postadmission for 30-day targeting), as well as, after reweighting for the sampling strategy, averages over all patients.

Significant differences between targeted and untargeted hospitals are designated as going in the expected direction (eg, targeted hospitals have lower quality of care or more severely ill patients) or going in the unexpected direction. There are only spotty differences, and they go in the unexpected direction as often as not. Patients with CHF who died within 30 days of admission received significantly worse care in 30-day targeted hospitals than in untargeted hospitals, but in all the other quality comparisons for both CHF and AMI patients, targeted hospitals were as good as or better than untargeted hospitals, though never significantly so. Patients with AMI in 30-day targeted hospitals were significantly sicker overall than those in untargeted hospitals, but that is the only significant severity comparison; for patients with CHF the nonsignificant trends are in the unexpected direction.

Average quality scores and, especially, average severity scores appear to be estimated quite precisely in Table 4, and the estimated differences between targeted and untargeted hospitals appear to be quite small. Still, it is worthwhile to investigate explicitly the importance of the estimated differences, and the importance of the uncertainty in the estimated differences, in terms of their implied effects on death rates. The effects of estimated differences in quality are small, and what differences there are tend to favor targeted hospitals (Table 5). That is, targeted hospitals from Table 4 have *better* estimated average quality (except for patients with AMI in inpatient targeted hospitals). Moreover, even at the lower bound of the confidence intervals for inpatients with an AMI, where quality is worse in targeted than in untargeted hospitals, poorer quality would contribute, if that result were true, just 0.28 percentage points to excess deaths in targeted hospitals.

The estimated differences in severity also have fairly small effects on death rates for patients with CHF. For patients with AMI, however, higher average severity in targeted hospitals has a substantial effect on death rates. For example, Table 4 shows that patients with AMI in 30-day targeted hospitals averaged 0.03 higher severity scores than those in untargeted hospitals. If all such patients had 0.03 lower severity scores, the death rate in targeted hospitals predicted by the Cox estimates would decrease by 2.8 deaths per 100 admissions. At the upper bound of the confidence interval, the patients in targeted hospitals had 0.05 higher severity scores. If all such patients had severity scores 0.05 lower than observed, that would have lowered targeted hospital death rates by 4.8 percentage points.

Explaining Differences in Death Rates Between Targeted and Untargeted Hospitals

In Table 6, we pull together several possible explanations for higher death

Table 4.—Differences in Severity of Illness, DNR Status, and Quality of Care Between Targeted and Untargeted Hospitals*

	CHF Patients		AMI Patients	
	Untargeted Hospital	Targeted Hospital	Untargeted Hospital	Targeted Hospital
Inpatient Targeting				
Patients in sample				
Alive	318	290	311	285
Dead	265	253	311	243
Total	**583**	**543**	**622**	**528**
Severity score				
Alive	0.32 (0.00)†	0.31 (0.00)	0.22 (0.01)	0.21 (0.01)
Dead	0.41 (0.01)	0.40‡(0.01)	0.38 (0.01)	0.39 (0.01)
Weighted average§	0.33 (0.00)	0.32 (0.00)	0.25 (0.01)	0.26 (0.01)
DNR status at admission				
Alive	0.02 (0.01)	0.00‡(0.00)	0.01 (0.01)	0.00 (0.00)
Dead	0.17 (0.02)	0.05‖(0.01)	0.08 (0.02)	0.07 (0.02)
Weighted average§	0.03 (0.01)	0.01‖(0.00)	0.02 (0.01)	0.02 (0.01)
Quality of process¶				
Alive	0.10 (0.05)	0.22 (0.05)	0.31 (0.04)	0.35 (0.05)
Dead	−0.27 (0.06)	−0.14 (0.06)	0.05 (0.06)	0.05 (0.06)
Weighted average§	0.07 (0.04)	0.16 (0.04)	0.26 (0.03)	0.26 (0.04)
30-d Targeting				
Patients in sample				
Alive	526	109	464	129
Dead	402	89	429	128
Total	**928**	**198**	**893**	**257**
Severity score				
Alive	0.32 (0.00)	0.31 (0.01)	0.22 (0.01)	0.23 (0.01)
Dead	0.40 (0.00)	0.40 (0.01)	0.36 (0.01)	0.39 (0.02)
Weighted average§	0.33 (0.00)	0.33 (0.01)	0.25 (0.00)	0.28#(0.01)
DNR status at admission				
Alive	0.01 (0.00)	0.00‡(0.00)	0.01 (0.00)	0.00‡ (0.00)
Dead	0.13 (0.02)	0.23**(0.05)	0.07 (0.01)	0.10 (0.03)
Weighted average§	0.03 (0.01)	0.04 (0.01)	0.02 (0.01)	0.03 (0.01)
Quality of process¶				
Alive	0.12 (0.04)	0.23 (0.08)	0.30 (0.03)	0.41 (0.07)
Dead	−0.13 (0.05)	−0.42**(0.13)	0.12 (0.05)	0.15 (0.08)
Weighted average§	0.09 (0.03)	0.11 (0.07)	0.26 (0.03)	0.32 (0.05)

*CHF indicates congestive heart failure; AMI, acute myocardial infarction; and DNR, "do not resuscitate."
†SEs are in parentheses.
‡$P<.05$, unexpected direction.
§Number was reweighted to reflect the fact that dead patients were oversampled relative to alive patients.
‖$P<.01$, unexpected direction.
¶Higher score is better care; see text and reference 24.
#$P<.01$, expected direction.
**$P<.05$, expected direction.

rates in targeted hospitals. Consider first the 30-day targeting method for patients with AMI. From Table 1, we know that targeted hospitals have a 30-day death rate that is 10.9 percentage points higher than in untargeted hospitals, and that 6.1 percentage points of that difference could result from random variation and the way hospitals were targeted. From the calculations in Table 5, we know that the estimated differences in average severity, DNR status at admission, and quality could account for a death rate an additional 2.8, 0.1, and −0.1 percentage points higher in targeted hospitals. That leaves a 2.0–percentage point gap that is not accounted for by either the way targeted hospitals were selected or measured differences between targeted and untargeted hospitals.

There is sufficient uncertainty in our estimate of the difference in average severity between targeted and untargeted hospitals so that, if that difference were at the upper limit of the 95% confidence interval, severity could account for an additional 2.0 percentage points of difference in death rates (4.85 − 2.81 = 2.04 from Table 5). Uncertainty in our estimate of the difference in average quality could account for 0.3 percentage points. Thus if both severity differences and quality differences were at their 95% confidence bounds, they would be sufficient to close the unexplained gap.

The situation is similar for 30-day targeting for patients with CHF, although here both the unexplained gap and the effect of a possibly underestimated severity difference or overestimated quality difference are smaller.

In contrast, a substantial unexplained difference remains between targeted and untargeted hospitals' inpatient death rates, even at the confidence bounds for differences in severity and quality (2.0 percentage points for CHF and 0.9 percentage points for AMI).

Retargeting

Table 7 compares average quality of care received by patients in alternatively targeted vs untargeted hospitals. Comparing only the "best" (lower than expected deaths) and "worst" ($P<.01$ of so many deaths) hospitals shows higher quality in targeted hospitals as often as not. Three-year targeting identifies significantly lower explicit quality in targeted hospitals treating patients with CHF and a nearly significant trend in the same direction for patients with AMI. Targeting by HCFA yields similar results, but the significantly low quality estimate for patients with CHF in targeted hospitals is suspect because it is based on only 26 patients.

COMMENT

We stated our main study objectives as follows: (1) to determine if hospitals with high death rates provide lower-quality care or have more severely ill patients than do hospitals with lower death rates and (2) to determine at the patient level how the probability of death is related to severity of illness and quality of care.

With respect to the first objective, we determined that hospitals targeted with unexpectedly high age-sex-race-disease–specific death rates do not provide lower quality of care than do untargeted hospitals, and that any differences in quality of care that lie within estimated confidence bounds have minimal effects on death rates. We found that higher average severity for patients with AMI in targeted hospitals accounts for about 25% of the difference in 30-day death rates, but differences in severity of illness do not explain higher death rates for patients with CHF.

With respect to the second objective, we determined that, at an individual patient level, higher severity of illness markedly increases the probability of death, and, to a lesser extent, better quality of care reduces the probability of death.

Finally, we are left with an unexplained excess of 2.9 or 0.6 deaths per 100 patients admitted with CHF for hospitals targeted on inpatients or 30-day death rates. For patients with AMI, the figures are 2.7 and 2.0 excess deaths per 100 admitted patients. The excess for 30-day targeting could possibly result from misestimated severity and quality differences between targeted and untargeted hospitals, but the uncertainty in these estimates is not large

enough to explain the excess for inpatient targeting. The excess for inpatient targeting could be the effect of unmeasured severity differences, unmeasured quality differences, or longer average lengths of stay in targeted hospitals.

Unmeasured severity of illness could be responsible for some of this excess, but another study has demonstrated that it will be difficult to improve the measurement of severity using just data in a medical record for patients with these two conditions.[24] If the hypothesis that hospital death rate differences are due to unmeasured severity is to be tested, then prospectively collected data from the patient or physician must be obtained.

Even the sophisticated explicit measures of quality used herein certainly cannot capture all of the potentially important differences in hospital care, if for no other reason than that they measure the process of care predominantly during the first 3 days following admission. A previous study demonstrated a relationship between quality of hospital care and the hospital death rates, but only when using implicit physician judgment to measure quality (not preset criteria).[14] That study found no significant relationship between death rates and a quality score explicitly calculated from medical record data, based on preset criteria. Any defects in quality in targeted hospitals appeared to be in areas not easily assessed by explicit measurement.

Another possibility is that the excess death rate in inpatient targeted hospitals results in part from longer average patient stays in those hospitals. Patients with CHF stay on average more than 4 days longer in inpatient targeted hospitals (14.1 days vs 9.7), and patients with AMI stay more than 1 day longer (13.1 vs 11.7). If these differences were entirely due to customary practice variations and exogenous constraints on patient discharges such as lack of nursing home beds, then differences in length of stay could easily explain the remaining gaps between death rates in targeted and untargeted hospitals, as we demonstrate in reference 19. But, of course, differences in length of stay can also arise from differences in severity of illness or in quality of care.

If one believes that quality differs among hospitals and that it is important to detect the differences, the more important question is whether a targeting mechanism can be devised that better identifies hospitals providing lower-quality care. For that reason, we retargeted the hospitals in our data set to take account of the following possible limitations in our targeting method.

Table 5.—Differences in Death Rates Between Targeted and Untargeted Hospitals That Correspond to Estimated Differences in Severity, DNR, and Quality*

Explanatory Variable	Death Rate Differences Between Targeted and Untargeted Hospitals (Deaths per 100 Admissions)	
	CHF Patients	AMI Patients
Inpatient Targeting		
Due to severity difference	−0.25†(−0.72 to 0.24)‡	1.20 (−0.21 to 2.69)
Due to DNR difference	−0.19†(−0.33 to −0.05)	−0.02†(−0.12 to 0.09)
Due to process difference	−0.11†(0.01 to −0.23)	0.02 (0.28 to −0.25)
30-d Targeting		
Due to severity difference	0.16 (−0.79 to 1.18)	2.81 (0.89 to 4.85)
Due to DNR difference	0.16 (−0.16 to 0.49)	0.08 (−0.12 to 0.28)
Due to process difference	−0.04†(0.19 to −0.26)	−0.12†(0.12 to −0.36)

*CHF indicates congestive heart failure; AMI, acute myocardial infarction; and DNR, "do not resuscitate."
†Minus sign indicates that lower severity, lower DNR, or higher quality in targeted hospitals contributes to the difference to be explained, rather than helping to explain the difference.
‡Numbers in parentheses are 95% confidence intervals.

Table 6.—Explaining Excess Death Rates in Targeted vs Untargeted Hospitals (Deaths per 100 Admissions)*

	CHF Patients		AMI Patients	
	Inpatient Targeting	30-d Targeting	Inpatient Targeting	30-d Targeting
Observed difference	7.4	5.0	10.2	10.9
Less expected due to binomial variation or selection effect	5.1	4.1	6.3	6.1
Unexplained excess after binomial variation	2.3	0.9	3.9	4.8
Less				
Expected due to differences in severity of illness	−0.3 (NS)	0.2 (NS)	1.2 (NS)	2.8
Expected due to differences in DNR at admission	−0.2 (NS)	0.2 (NS)	−0.0 (NS)	0.1 (NS)
Expected due to differences in quality of care	−0.1 (NS)	−0.0 (NS)	0.0 (NS)	−0.1 (NS)
Unexplained after binomial variation and systematic differences	2.9	0.6	2.7	2.0

*CHF indicates congestive heart failure; AMI, acute myocardial infarction; DNR, "do not resuscitate"; NS, no significant difference between targeted and untargeted hospitals ($P>0.05$); and minus sign, that lower severity, lower DNR, or higher quality in targeted hospitals contributes to the difference to be explained, rather than helping to explain the difference.

Table 7.—Differences in Quality of Care Between Targeted and Untargeted Hospitals Using Alternative Targeting Methods*

Retargeting Method	CHF Patients		AMI Patients	
	Untargeted Hospitals	Targeted Hospitals†	Untargeted Hospitals	Targeted Hospitals
Best (lower than expected deaths) vs worst ($P<.01$) inpatient targeting				
Patients in sample	265	370	328	226
Mean quality score†	0.17 (0.05)	0.21 (0.05)	0.25 (0.05)	0.36 (0.05)
Best (lower than expected deaths) vs worst ($P<.01$) 30-d targeting				
Patients in sample	399	96	402	95
Mean quality score†	0.12 (0.04)	−0.05 (0.10)	0.25 (0.04)	0.15 (0.10)
Hospitals targeted by the HCFA for 1986 ($P<.05$)				
Patients in sample	1080	26	984	141
Mean quality score†	0.11 (0.03)	−0.42‡(0.16)	0.28 (0.03)	0.18 (0.07)
3-y Targeting				
Patients in sample	950	156	1006	117
Mean quality score†	0.13 (0.03)	−0.20‡(0.08)	0.28 (0.03)	0.13 (0.08)

*CHF indicates congestive heart failure; AMI, acute myocardial infarction; and HCFA, Health Care Financing Administration.
†Higher score is better care; see text and reference 24. Values in parentheses are SEs.
‡Significant difference between untargeted and targeted hospitals, $P<.01$, expected direction.

One possibility is that a $P<.05$ cutoff for targeting (where P is the probability that the hospital would have as many deaths as it did) is not strict enough, or that differences are obscured by contrasting hospitals below the probability cutoff with *all* other hospitals (which may include small hospitals with high death rates and low-quality care, but too few patients to reach the statistical significance necessary for targeting). Another possibility is that our method does not adequately control for severity of illness. The method used by HCFA is similar to ours but adjusts for severity as thoroughly as possible, using administrative data; we adjusted only for age, sex, race, and disease. A third possibility is that random variation obscures any real differences in a single-year analysis. When we retargeted to minimize or avoid these possible problems, the results were mixed and, except possibly for 3-year targeting, did not add much to our original analysis.

In summary, our analyses of a representative sample of patients with CHF and AMI in four populous states have not produced much evidence that hospitals with higher than expected death rates based only on administrative data actually, on review of their medical records, provide lower-quality care. What can be done to improve our ability to identify such hospitals? There is some evidence that targeting hospitals with consistently high death rates over periods longer than 1 year may identify potential quality problems. Further research needs to be performed to identify the optimal targeting interval (2, 3, or 4 years), but the time interval cannot be so long as to make the results irrelevant to current patient care. The current HCFA mortality release reports results for 3 years of data, but they are analyzed 1 year at a time.[3]

Our targeting method did not use all of the information available in the administrative data to control for severity of illness, only age, sex, race, and a clinical meaningful grouping of principal diagnoses from the *International Classification of Diseases, Ninth Revision*. Thus, it is closer to the method used by HCFA in its initial 1986 data release (*New York Times*, March 12, 1986:1) (when this study was designed) than it is to HCFA's current method.[3] Would targeting that included other severity adjustment measures that are available from administrative data have substantially affected the results of this study? We think not. Green et al[17] found that the severity adjustment used in HCFA's 1988 release[2] explained only 2.5% of the variance in outcome on average for the five broadly defined diseases investigated, including 3.9% for severe acute heart disease. The 1989 analysis of HCFA's, which controls for principal diagnosis grouped into homogeneous death rate categories,[3] might explain on the order of 8% of the variance in our already fairly homogeneously defined AMI, and probably only 2% for our CHF. Thus, even if we had used the most recent HCFA method to adjust for severity, we would have reduced the width of the targeting confidence intervals by no more than 4% or so, and the targeting probabilities calculated with and without the additional adjustment would have been highly correlated.

Over the years HCFA has substantially improved its targeting method, and it continues to improve. But, given our results, we believe that the improved methods should be tested to see if they are indeed targeting lower-quality hospitals. First, such methods should be tested against simulation models to confirm that they are not just picking hospitals whose high death rates result solely from random variation. Second, the quality of care in targeted vs untargeted hospitals should be compared using clinical data from medical records. Third, that comparison should include both implicit and explicit assessment of quality. Fourth, sufficient public discussion about both the targeting methods and results should occur so that perhaps their acceptability within the medical profession and the hospital community will increase.[26] And finally, if targeting based on administrative data cannot be improved, then serious attention needs to be given to whether detailed data on severity of illness at time of admission should be collected routinely and nationally. If hospitals were targeted based on detailed severity data, would the targeting be more accurate? Would the additional accuracy be worth the cost?

This work was supported by the Health Care Financing Administration. We appreciate the good work of Caren Kamberg and Barbara Genovese, who were responsible for most peer review organization contact and hard-copy data management for this study.

References

1. Bowen OR, Roper WL. *Medicare Hospital Mortality Information, 1986*. Health Care Financing Administration, US Dept of Health and Human Services; 1987. HCFA publication 01-002.
2. Bowen OR, Roper WL. *Medicare Hospital Mortality Information, 1987*. Health Care Financing Administration, US Dept of Health and Human Services; 1988. HCFA publication 00651.
3. Sullivan LW, Hays LB. *Medicare Hospital Mortality Information, 1986, 1987, 1988*. Health Care Financing Administration, US Dept of Health and Human Services; 1989. HCFA publication 00701.
4. Greenfield S, Aronow HU, Elashoff RM, Watanabe D. Flaws in mortality data: the hazards of ignoring comorbid disease. *JAMA*. 1988;260:2253-2255.
5. Blumberg MS. Comments on HCFA hospital death rate statistical outliers. *Health Serv Res*. 1987;21:715-740.
6. Wagner DP, Knaus WA, Draper EA. The case for adjusting hospital death rates for severity of illness. *Health Aff*. 1986;5:148-153.
7. Fink A, Yano EM, Brook RH. The condition of the literature on differences in hospital mortality. *Med Care*. 1989;27:315-336.
8. 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.
9. Flood AB, Scott WR, Ewy W. Does practice make perfect? Part II: the relation between volume and outcomes and other hospital characteristics. *Med Care*. 1984;22:115-125.
10. Flood AB, Scott WR. *Hospital Structure and Performance*. Baltimore, Md: The Johns Hopkins Press; 1987.
11. Hannan EL, O'Donnell JF, Kilburn H, Bernard HR, Yazici A. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. *JAMA*. 1989;262:503-510.
12. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. *N Engl J Med*. 1979;301:1364-1369.
13. Riley G, Lubitz J. Outcomes of surgery among the Medicare aged: surgical volume and mortality. *Health Care Financ Rev*. 1985;7:37-47.
14. Dubois RW, Rogers WH, Moxley JH, et al. Hospital inpatient mortality: is it a predictor of quality? *N Engl J Med*. 1987;317:1674-1680.
15. Pollack MM, Ruttmann UE, Getson PR, and Members of the Multi-Institutional Study Group. Accurate prediction of the outcome of pediatric intensive care: a new quantitative method. *N Engl J Med*. 1987;316:134-139.
16. Jencks SF, Daley J, Draper D, Thomas N, Lenhart G, Walker J. Interpreting hospital mortality data: the role of clinical risk adjustment. *JAMA*. 1988;260:3611-3624.
17. Green J, Wintfeld N, Sharkey P, Passman LJ. The importance of severity of illness in assessing hospital mortality. *JAMA*. 1990;263:241-246.
18. Chassin MR, Park RE, Lohr KN, et al. Differences among hospitals in Medicare patient mortality. *Health Serv Res*. 1989;24:1-31.
19. Park RE, Brook RH, Kosecoff J, et al. *Explaining Variations in Hospital Death Rates: Randomness, Severity of Illness, Quality of Care*. Santa Monica, Calif: The RAND Corp. In press.
20. Kahn KL, Chassin MR, Rubenstein LV, et al. *Medical Record Abstraction Form and Guidelines for Assessing Quality of Care for Hospitalized Patients With Congestive Heart Failure*. Santa Monica, Calif: The RAND Corp; 1988. N-2798-HCFA.
21. Kosecoff J, Rubenstein LV, Kahn KL, et al. *Medical Record Abstraction Form and Guidelines for Assessing Quality of Care for Hospitalized Patients With Acute Myocardial Infarction*. Santa Monica, Calif: The RAND Corp; 1988. N-2799-HCFA.
22. Kahn KL, Rubenstein LV, Draper D, et al. The effect of the DRG-based prospective payment system on quality of care for hospitalized Medicare patients: introduction to the series. *JAMA*. In press.
23. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med*. 1986;104:410-418.
24. Keeler EB, Kahn KL, Draper D, et al. Changes in sickness at admission following the introduction of prospective payment. *JAMA*. In press.
25. Kahn KL, Rogers WH, Rubenstein LV, et al. Measuring quality of care with explicit process criteria pre- and post-implementation of the DRG-based prospective payment system. *JAMA*. In press.
26. Berwick DM, Wald DL. Hospital leaders' opinions of the HCFA mortality data. *JAMA*. 1990; 263:247-249.