An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With an Acute Myocardial Infarction

Harlan M. Krumholz, MD, SM; Yun Wang, PhD; Jennifer A. Mattera, MPH; Yongfei Wang, MS; Lein Fang Han, PhD; Melvin J. Ingber, PhD; Sheila Roman, MD, MPH; Sharon-Lise T. Normand, PhD

Background—A model using administrative claims data that is suitable for profiling hospital performance for acute myocardial infarction would be useful in quality assessment and improvement efforts. We sought to develop a hierarchical regression model using Medicare claims data that produces hospital risk-standardized 30-day mortality rates and to validate the hospital estimates against those derived from a medical record model.

Methods and Results—For hospital estimates derived from claims data, we developed a derivation model using 140 120 cases discharged from 4664 hospitals in 1998. For the comparison of models from claims data and medical record data, we used the Cooperative Cardiovascular Project database. To determine the stability of the model over time, we used annual Medicare cohorts discharged in 1995, 1997, and 1999–2001. The final model included 27 variables and had an area under the receiver operating characteristic curve of 0.71. In a comparison of the risk-standardized hospital mortality rates from the claims model with those of the medical record model, the correlation coefficient was 0.90 (SE=0.003). The slope of the weighted regression line was 0.95 (SE=0.007), and the intercept was 0.008 (SE=0.001), both indicating strong agreement of the hospital estimates between the 2 data sources. The median difference between the claims-based hospital risk-standardized mortality rates and the chart-based rates was <0.001 (25th and 75th percentiles, −0.003 and 0.003). The performance of the model was stable over time.

Conclusions—This administrative claims-based model for profiling hospitals performs consistently over several years and produces estimates of risk-standardized mortality that are good surrogates for estimates from a medical record model. **(*Circulation*. 2006;113:1683-1692.)**

Key Words: health policy ■ quality of health care ■ myocardial infarction

Acute myocardial infarction (AMI), a common, high-risk event that requires timely intervention and extensive coordination among hospital clinicians and personnel, is the focus of several national efforts to improve quality of care.[1] The Centers for Medicare & Medicaid Services (CMS) and the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) publicly report 7 process measures for AMI, including the use of aspirin on admission and discharge, β-blockers on admission and discharge, angiotensin-converting enzyme inhibitors or angiotensin receptor blockers on discharge, time to reperfusion, and smoking cessation counseling.[2,3] These measures convey important information about healthcare quality but focus on a narrow spectrum of the overall care provided to patients and thus explain a relatively small portion of the variation across hospitals in risk-adjusted mortality rates.[4]

Clinical Perspective p 1692

Outcomes, in contrast to structure and process, provide a more global assessment of healthcare quality and represent what matters most to patients.[5] Process measurement is susceptible to the diversion of resources to what is being measured at the expense of what is not, potentially worsening overall quality of care and outcomes. Although outcomes are not entirely under the control of clinicians and hospitals, quality of care and safety can influence the risk of adverse events. Moreover, outcomes have the most relevance to patients.

Outcome measurement is challenging, however, because of variation among institutions in the risk profile of their patients.[6] Statistical methods can adjust for observed differ-

Received December 28, 2005; revision received February 13, 2006; accepted February 16, 2006.

From the Section of Cardiovascular Medicine, Department of Medicine (H.M.K., Yongfei Wang), Section of Health Policy and Administration, Department of Epidemiology and Public Health (H.M.K.), and Robert Wood Johnson Clinical Scholars Program (H.M.K.), Yale University School of Medicine, New Haven, Conn; Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, Conn (H.M.K., Yun Wang, J.A.M.); Centers for Medicare & Medicaid Services, Baltimore, Md (L.F.H., M.J.I., S.R.); Department of Health Care Policy, Harvard Medical School, Boston, Mass (S.T.N.); and Department of Biostatistics, Harvard School of Public Health, Boston, Mass (S.T.N.).

Guest Editor for this article was Donna K. Arnett, PhD.

Correspondence to Dr Harlan M. Krumholz, Yale University School of Medicine, Room I-456 SHM, 333 Cedar St, PO Box 208088, New Haven, CT 06520-8088. E-mail harlan.krumholz@yale.edu

Circulation **is available at http://www.circulationaha.org** **DOI: 10.1161/CIRCULATIONAHA.105.611186**

ences in patient risk, but the only nationally available data on hospitalizations in the United States are Medicare claims, which do not always accurately reflect the information in the medical record.[7] Thus, for a claims-based model to be suitable for public reporting, it should ideally be validated against a similar approach using medical record data.[8] The validation should not be assessed at the patient level but instead should assess how well the characterization of hospital performance by administrative data compares with that of the model based on medical record data.

We developed a hierarchical regression model using Medicare claims data that produces hospital risk-adjusted 30-day mortality rates. We aligned our approach with a recent American Heart Association Scientific Statement that defined standards for statistical models used for the public reporting of health outcomes.[8] This document recommends that such models be in the public domain, not misclassify complications as comorbidities, have a standardized period of outcome assessment, use statistical techniques that account for clustering of the data, and be validated with various cohorts and against the results of a model based on medical record data. We compared hospital risk-standardized mortality estimates derived from a claims-based model with rates determined from a model based on medical record data for 181 032 patients discharged from 4322 US hospitals from 1994 to 1996. To assess the stability of the model over time, we also assessed model performance in multiple years of Medicare claims data.

Methods

Derivation and Validation Cohorts

The Derivation Cohort

We randomly sampled half of the hospitalizations for AMI (*International Classification of Diseases, 9th Revision, Clinical Modification* [*ICD-9-CM*] codes 410.xx except for 410.x2) in the 1998 Medicare Provider Analysis and Review (MEDPAR) files, clustered within hospitals. For risk adjustment we used information in the MEDPAR files, physician files, and hospital outpatient files. The MEDPAR claims have data on each hospitalization for fee-for-service Medicare enrollees and include demographic information, principal and secondary diagnosis codes, and procedure codes. Diagnosis codes for comorbidities were also collected from physician and hospital outpatient files. These data were collected for the year before the index hospitalization.

We retained hospitalizations in which the patient was aged >64 years because these patients are representative of the older AMI population. We linked hospitalizations as an episode of care and attributed the outcome to the admitting hospital. To confirm the diagnosis, patients with AMI who were transferred from one facility to another were required to have a principal discharge diagnosis of AMI at both hospitals. For transferred patients, comorbid ("preexisting") conditions were identified from the initial (index) admission only so that these patients would not have the opportunity to have more conditions coded than the patients who were not transferred. We excluded patients with a total length of stay of ≤1 day and who were discharged alive and not against medical advice because it is unlikely that these patient suffered an AMI. We also excluded patients without 1 year of history in Medicare fee-for-service.

The Validation Cohorts

We constructed a linked sample that contains both claims and medical chart abstracted data for each patient. The medical record data were obtained from the Cooperative Cardiovascular Project (CCP), a national AMI quality improvement project in which >200 000 medical records were abstracted.[9] Hospitalizations for CCP were identified from hospital bills in the Medicare National Claims History File of claims submitted under fee-for-service. Hospitalizations for CCP occurred during an 8-month period between February 1994 and July 1995, except for the states in the pilot study (Alabama, Connecticut, Iowa, and Wisconsin), in which sampling took place during a 4-month period from August through November 1995. Predefined variables were abstracted from copies of hospital records. The reliability of the data was monitored by means of monthly reabstractions of randomly selected records; the accuracy of abstraction with respect to treatment variables was 95%.

To evaluate the stability of the claims model over time, we also evaluated the performance of the Medicare claims model using the other half of the 1998 MEDPAR data and data for each of years 1995, 1996, 1997, 1999, 2000, and 2001. In each case we constructed the sample in the same manner as for the derivation cohort.

Outcome

The primary outcome was hospital-specific risk-standardized all-cause 30-day mortality, defined as death from any cause 30 days after the index admission date. Mortality information was obtained from the Medicare enrollment files by linking unique patient identifiers.

Model Derivation: Patient Predictors of Mortality

Candidate variables for the Medicare claims model were developed primarily from the administrative diagnostic codes. Because there are >15 000 *ICD-9-CM* codes, we used the Hierarchical Condition Categories (HCC) to assemble clinically coherent codes into single variables.[10] This system, which includes 189 categories, was developed by physician and statistical consultants under a contract to CMS and is publicly available. The HCC candidate variables considered for this model were derived from the secondary diagnosis and procedure codes from the index hospitalization (all the principal diagnoses were 410.xx, except for 410.x2) and from the principal and secondary diagnosis codes from hospitalizations, hospital outpatient visits, and outpatient office encounters in the 12 months before the index hospitalization.

We conducted a clinical review of the candidate variables to eliminate the secondary diagnoses from the index hospitalization that could have represented complications rather than conditions present on admission. For example, we did not include hemorrhage as a secondary diagnosis on the index admission because it may have been present on admission or occurred during the hospitalization. We combined categories of HCC variables on the basis of clinical judgment and bivariate associations and eliminated candidate variables with <1% frequency. Additional candidate variables, based on clinical judgment and a review of the literature, included demographic factors (age, sex), location of the AMI, and procedural factors (history of bypass surgery or percutaneous coronary intervention in the past year).

Model

Because of the natural clustering of the observations within hospitals, we estimated hierarchical generalized linear models (HGLM).[11–13] We modeled the log-odds of mortality within 30 days of admission as a function of patient demographic and clinical variables and a random hospital-specific effect. This strategy accounts for within-hospital correlation of the observed outcomes, separates within-hospital variation from between-hospital variation, and models the assumption that underlying differences in quality among the institutions lead to systematic differences among hospital outcomes. The covariates for the model were first selected with the use of a backward elimination procedure through the generalized linear model (GLM) with a logit link function approach. Because of the large number of patient observations that heavily influences probability values, we chose an exit criterion for a variable of $P>0.01$. We also evaluated the full model compared with the model that contained only age and sex as covariates. We calculated the area under the receiver operating characteristic (ROC) curve and the percentage of explained variation (R^2). Finally, we reestimated the

regression coefficients of the covariates identified from our backward elimination strategy using a HGLM.

Model Validation

Medical Record Model

Candidate variables for the medical record model were selected on the basis of a literature review and clinical experience.[4,14] Unlike the claims data, some covariates could be missing for patients in the sample. When there were missing data for a continuous-valued variable, we created an additional variable that assumed a value of 0 if the variable was measured and a value of 1 if missing, and set the value of the original variable to its mean when missing. This method of modeling missing data assumes that data are missing at random and permits inclusion of all available cases, although it is not as efficient as multiple imputation procedures. For discrete-valued variables, we included an additional level that indicated the variable was missing. We computed measures of model fit and discrimination for the medical record model similar to those computed for the claims-based models.

Hospital Risk-Standardized Mortality Rates

We calculated risk-standardized mortality rates for each hospital using the estimated hospital-specific parameters from the respective hierarchical models. These rates are obtained as the ratio of "predicted" to expected mortality, multiplied by the national unadjusted rate.[15] Although other researchers have calculated the ratio of observed to expected outcomes, we use the predicted rates to avoid several analytic problems that have been cited.[11,13,16] The expected outcome for each hospital is the number of 30-day deaths expected at the hospital if the hospital's patients were treated at a "reference" hospital. Operationally this was accomplished by regressing the risk factors on the mortality with all hospitals in our sample, applying the subsequent estimated regression coefficients to the patient characteristics observed at the hospital, and then summing. This is a form of indirect standardization. The predicted hospital outcome is the number of expected mortalities at the "specific" hospital and not at a reference hospital. Operationally this was accomplished by estimating a hospital-specific random effect that represented baseline mortality risk for the hospital, applying the hospital-specific regression coefficients to the patient characteristics at the hospital, and then summing.

Using the 1994 to 1995 hospitalizations, we used 2 approaches to examine the relationship between the risk-standardized rates obtained from using administrative data and those using chart data. First, after creating a linked sample of admissions between the administrative claims data and the medical record data, we assessed the relationship between the risk-standardized mortality rates from the administrative claims model and from the chart model for each hospital through graphical and regression techniques. We estimated a linear regression equation describing the association between the 2 rates, weighting each hospital by the number of hospitalizations, and calculated the intercept and the slope of this equation. A slope close to 1 and an intercept close to 0 would provide evidence that the hospital rates from the 2 sources are very similar. Second, for each hospital we calculated the difference between the risk-standardized mortality rate based on the claims data and the medical record data and then summarized the distribution of these differences among the hospitals using the average, median, and maximum differences.

Stability of the Model Over Time

We validated the model over time by comparing its performance in the derivation set with various validation cohorts, as described above. To assess whether we included too many risk factors in our final model, we calculated indices that quantify overfitting. Specifically, we used the coefficients estimated from the derivation model to predict the log-odds of mortality in the validation cohorts. This was accomplished by multiplying the observed risk factors in each validation cohort and summing over the covariates for a subject to obtain a mortality score. Using these scores for each subject, we then estimated a logistic regression model in which the outcome was observed mortality and the single covariate was the risk score. The intercept and slope obtained from this model are referred to as overfitting indices. If there is overfitting, we would expect the slopes to be different from 1 and the intercepts to be different from 0. We repeated this process for each validation data set, each time calculating a risk score using the regression estimates from our derivation model.

After assessing overfitting, we recalibrated the models in each of the validation data sets so that we used the same variables but reestimated the regression coefficients to the data for each specific cohort. We then calculated several indices for assessing model performance[18]: the area under the ROC curve, explained variation as measured by the generalized R^2 statistic, and the observed outcomes in strata defined by the lowest and highest deciles based on predictive probabilities. Model fit was further assessed through examination of Pearson residuals.

All analyses were conducted with the use of SAS version 8.02 (SAS Institute Inc, Cary, NC). Models were fitted separately to each year of data. The hierarchical models were estimated with the use of the GLIMMIX macro in SAS.

The authors had full access to the data and take responsibility for its integrity. All authors have read and agree to the manuscript as written.

Results

Patient Characteristics and Administrative Model: Derivation Sample

The 1998 sample included 387 081 AMI discharges from 4828 hospitals that were retrieved from the national fee-for-service administrative claims database, of which 8.3%, 10.7%, and 2.0% of the discharges were excluded for age $<$65 years, incomplete information in the 12 months before admission, and length of stay of $\leq$1 day, respectively (Table 1). Another 10.7% of the hospitalizations represented transfer in admission and were combined with the admission at the initial hospital to create an episode of care.

The derivation sample consisted of 140 120 cases with an unadjusted 30-day mortality rate of 18.0%. The mean age of the cohort was 77.8$\pm$7.4 years. The cohort included 50.5% women and 9.7% nonwhite patients. There were 4664 hospitals in the derivation cohort, with a median annual number of Medicare AMI hospitalizations of 17 (25th and 75th percentiles, 6 and 40, respectively). The observed mortality rate ranged from 0.0% to 100.0% across these hospitals, and the 25th, 50th, and 75th percentiles were 13.1%, 16.9%, and 21.1%, respectively.

On the basis of a clinical review of the candidate variables, bivariate analysis, and stepwise GLM procedure, we identified 27 variables, including 2 demographic, 10 cardiovascular, and 15 comorbidity variables, for the final model. The model had good discrimination, calibration, and fit (Table 2). The area under the ROC curve was 0.71. Model discrimination was good, with the observed mortality rate ranging from 4.0% in the lowest predicted decile to 40.0% in the highest predicted decile, a range of 36.0%. The adjusted R^2 was 0.12. Figure 1A illustrates the overall distribution of risk-standardized 30-day mortality rates at the hospital level, and Figure 1B shows the distributions of risk-standardized 30-day mortality rates stratified by hospital volume. The 25th and 75th percentiles were 16.8% and 18.4%, respectively. The 95th percentile was 19.9%, and the 5th percentile was 15.5%.

The model that included only age and sex had worse model fit with an area under the ROC curve of 0.62. Model

TABLE 1. AMI Initial Administration Claims Sample

Data Source	Total	Exclusion, %				Final Sample	
		Age <65 y	Incomplete Information*	Transfer†	LOS ≤1 d‡	n	% of Total
1995	386 521	7.9	10.7	10.7	1.0	283 578	73.4
1996	393 845	8.1	10.9	10.9	1.3	285 953	72.6
1997	391 812	8.2	11.2	10.9	1.7	282 127	72.0
1998	387 081	8.3	10.7	10.7	2.0	280 098	72.4
1999	387 018	8.5	10.4	10.7	2.3	280 319	72.4
2000	346 595	8.3	5.2	11.6	2.6	263 124	75.9
2001	354 402	8.6	6.6	11.4	3.0	264 191	74.6

LOS indicates length of stay.

*Incomplete information in the 12-month, preindex admission period was excluded for the AMI sample.

†After linking the "transfer-in" hospital with the "transfer-out" (index admission) hospital, the records of the "transfer-in" hospital were deleted so that the case was assigned to the index admission hospital.

‡Discharged within first day of admission and alive, not against medical advice, not transferred.

discrimination was not as good: The observed mortality rate ranged from 10.2% in the lowest predicted decile to 31.0% in the highest predicted decile, a range of 20.8%. Explained variation, R^2, was only 0.04.

The model that did not include institutional outpatient and physician data as a source of data had a lower prevalence of comorbid conditions. For example, diabetes was present in 32.9% on the basis of inpatient, outpatient, and physician data but was only 28.9% for inpatient alone. Hypertension was present in 46.8% on the basis of inpatient, outpatient, and physician data but was only 31.7% for inpatient alone. Overall, among 27 variables, except age, male gender, history of percutaneous transluminal coronary angioplasty, history of percutaneous coronary intervention, and location of AMI in the index admission, many variables were affected by physician data; the absolute percent change in frequency ranged from 0.17% (chronic liver disease) to 15.9% (trauma in last year), with a mean of 5.8±4.6%. However, the model using only inpatient data had model fit that was close to the model based on inpatient, outpatient, and physician data, with an area under the ROC curve of 0.70. The observed mortality rate ranged from 3.9% in the lowest predicted decile to 39.8% in the highest predicted decile, a range of 35.9%. The adjusted R^2 was 0.11.

Medical Record Model

The final CCP validation sample contained 181 032 hospitalizations and a crude 30-day mortality rate of 18.8%. In this cohort, the administrative model had an area under the ROC curve of 0.69, an observed mortality rate ranging from 5.3% in the lowest predicted decile to 38.6% in the highest predicted decile, and an adjusted R^2 of 0.10. The medical record comparison model in this cohort included 31 variables (Table 3). The area under the ROC curve was 0.77. The observed mortality rate ranged from 2.9% in the lowest predicted decile to 59.0% in the highest. Explained variation, R^2, was 0.24. For all covariates that represented similar information in the 2 data sources, the respective estimated regression coefficients were in a similar direction. Bypass surgery was positive in the medical record model and negative in the claims model, but the definitions were different: In the claims model there was a requirement for a billing code in the year before the AMI, whereas in the medical record model there was written documentation of bypass surgery at any time.

Comparison of Hospital Mortality Rates: Claims and Medical Record Data

The estimated hospital-specific standardized 30-day mortality rates derived from each model are displayed in Figure 2A and stratified by volume in Figure 2B. The slope of the weighted regression line is 0.95 (SE=0.007) and the intercept is 0.008 (SE=0.001), both indicating strong agreement of the hospital risk-standardized mortality estimates between the 2 data sources. The correlation coefficient of the standardized mortality rates from the 2 models is 0.90 (SE=0.003). The median difference between the hospital-specific risk-standardized mortality rates estimated from the claims data and those estimated from the medical record data was <0.001 (25th and 75th percentiles, −0.003 and 0.003, respectively; 10th and 90th percentiles, −0.007 and 0.007, respectively).

Model Performance in Administrative Validation Set

In each validation cohort the model fit was similar to that of the derivation cohort (Table 4). These comparisons spanned 7 years of Medicare admissions for AMI. The unadjusted mortality ranged from 18.1% to 19.0%. The percent explained variation ranged from 0.11 to 0.12, and the area under the ROC curve ranged from 0.69 to 0.71. The overfitting statistics were all within an acceptable range, indicating that we had not overfitted the models.

Discussion

This study introduces an administrative claims-based model for reporting hospital-specific 30-day AMI mortality rates for Medicare beneficiaries with output that is an excellent surrogate for that produced by a model based on medical record

TABLE 2. AMI Administrative Model Based on 1998 Derivation Sample

Predictor	Estimate	*t*	Odds Ratio	95% CI
Intercept	−2.61	−99.78		
Demographic				
Age, years over 65	0.05	49.45	1.05	1.05–1.05
Male	0.06	4.34	1.07	1.04–1.10
Cardiovascular				
History of PTCA	−0.52	−9.83	0.60	0.54–0.66
History of CABG	−0.10	−3.23	0.91	0.85–0.96
History of heart failure (HCC 80)	0.42	23.03	1.52	1.47–1.58
History of AMI (HCC 81)	−0.44	−22.72	0.65	0.62–0.67
Anterior MI (*ICD-9* 410.00–410.19)	0.64	34.74	1.89	1.83–1.96
Inferior/lateral/posterior MI (*ICD-9* 410.20–410.69)	0.44	24.49	1.56	1.51–1.62
Unstable angina (HCC 82)	−0.13	−6.14	0.88	0.84–0.91
Chronic atherosclerosis (HCC 83 and 84)	−0.41	−25.91	0.67	0.65–0.69
Cardiopulmonary-respiratory failure and shock (HCC 79)	0.30	10.95	1.35	1.28–1.42
Valvular heart disease (HCC 86)	0.09	4.29	1.09	1.05–1.13
Comorbidity				
Hypertension (HCC 89 and 91)	−0.13	−8.43	0.88	0.85–0.91
Stroke (HCC 95 and 96)	0.23	8.49	1.26	1.20–1.33
Cerebrovascular disease (HCC 97, 98, 99, 103)	0.08	3.76	1.09	1.04–1.13
Renal failure (HCC 131)	0.36	13.25	1.43	1.36–1.51
COPD (HCC 108)	0.15	8.66	1.16	1.12–1.20
Pneumonia (HCC 111, 112, 113)	0.13	5.85	1.14	1.09–1.19
Diabetes (HCC 15–20, 120)	0.24	15.59	1.28	1.24–1.32
Protein-calorie malnutrition (HCC 21)	0.51	13.27	1.66	1.54–1.79
Dementia (HCC 49–50)	0.41	18.64	1.51	1.45–1.58
Hemiplegia, paraplegia, paralysis, functional disability (HCC 100, 101, 102, 68, 69, 177, 178)	0.31	9.60	1.36	1.28–1.45
Peripheral vascular disease (HCC 104, 105)	0.21	10.62	1.23	1.18–1.28
Metastatic cancer (HCC 7, 8)	0.58	14.89	1.78	1.65–1.92
Trauma in last year (HCC 154–156, 158–162)	0.10	5.73	1.11	1.07–1.15
Major psychiatric disorders (HCC 54, 55, 56)	0.23	7.01	1.25	1.18–1.34
Chronic liver disease (HCC 25, 26, 27)	0.61	8.10	1.84	1.59–2.13

PTCA indicates percutaneous transluminal coronary angioplasty; CABG, coronary artery bypass graft surgery; and COPD, chronic obstructive pulmonary disease. Estimate of between-hospital variance=0.041 (SE=0.0032).

data. From prior work we know that at the patient level, medical record data are better for discriminating patients who survive and those who do not. In profiling institutions, however, the emphasis is on a measure that averages information for all patients within the hospital rather than on the specific agreement of individual variables or patient-level discrimination. Because the claims-based model and a medical record model classified hospitals similarly with respect to their standardized mortality rates, this approach, which has been endorsed by the National Quality Forum, may be suitable for the public reporting of hospital outcomes for patients with AMI. We note that our comparison between data sources focused on risk-standardized estimates; investigators who wish to use the results in different ways will need to undertake an assessment of the comparability of the 2 data sources for that purpose.

The development of this model included several methodological improvements on currently utilized administrative data risk-adjustment models. The model was designed to include only diagnosis codes that indicate conditions present on admission, thus avoiding the problem of unwittingly crediting a hospital for more ill patients who just may have had more complications during the hospitalization, possibly as a result of worse care. We used clinical judgment to exclude secondary diagnosis codes in which complications could not be distinguished from preexisting conditions. In addition, we defined the outcome with a standardized period of follow-up rather than relying on the hospital stay that may vary by institution. In addition, we made use of healthcare utilization in the year before the index admission to improve the predictive ability of the model and minimize the potential for the gaming of codes at the admitting hospital.

A notable aspect of our approach is the validation achieved by comparing the output of the administrative claims model with that from a model based on medical record review data.

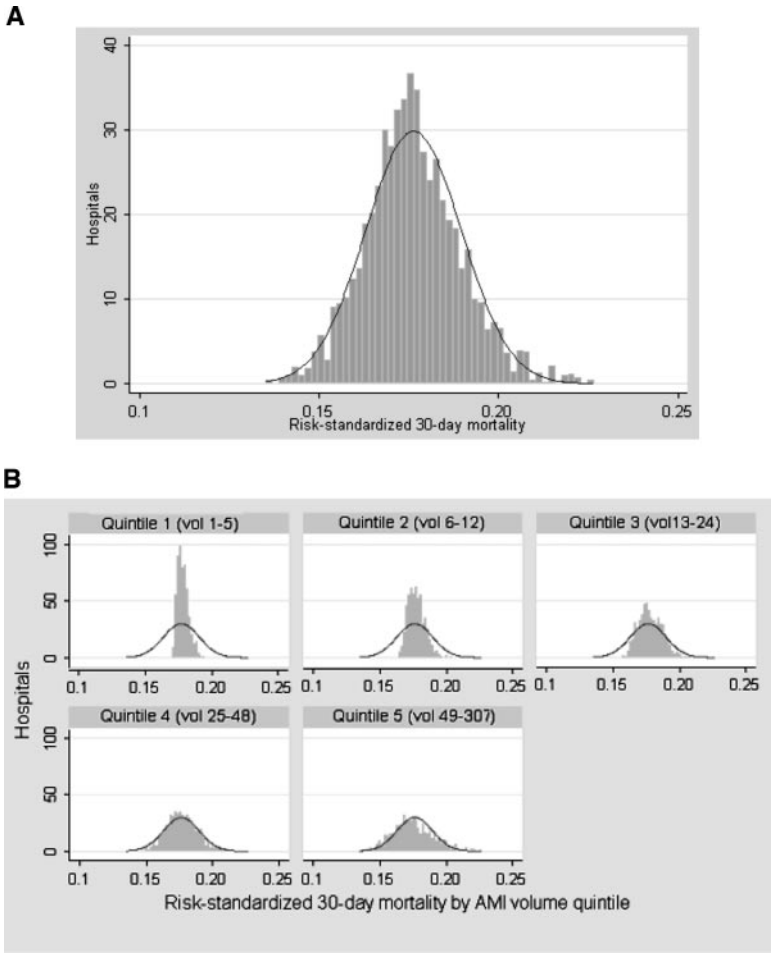

Figure 1. Distributions of risk-standardized 30-day AMI mortality rates overall (A) and stratified by volume (B).

Although even a model based on detailed information from medical records is not truly a gold standard, models based on this information currently provide the best opportunity to characterize the baseline risk of patients admitted to the hospital with an AMI. The Medicare claims data model did not perform as well as the medical record model at the patient level; however, the output of the models, ie, the profiling of the hospital performance, is the focus of this effort. We show that at the hospital level the administrative claims data, for all their limitations, can serve as a reasonable surrogate for the risk-standardized estimates from a model based on better data. The current cost of collecting medical record data precludes their collection as part of a national profiling effort. Therefore, the only current opportunity to develop a national hospital profiling effort can be based on administrative claims data. Our findings suggest that such an approach is possible.

We particularly sought an approach that could be released in the public domain. Many of the current publicly reported hospital profiling systems use proprietary approaches and do not provide information about model fit or validation against a gold standard.[19,20] This model, including its methodology, covariates, and performance, can be posted and disseminated.

Our analytic approach used hierarchical modeling, which takes into account nesting of the data (ie, patients within hospitals). Patients within hospitals would be expected to have a mortality risk that is more highly correlated than that of patients in different hospitals. This lack of independence of the observations can lead to underestimation of the SEs of risk factors and cause the appearance of statistically significant differences where none truly exist. In addition, sample sizes vary by hospital, and hierarchical modeling can take into account the differences in the amount of information provided by each hospital.[12,21,22]

An important question is whether 30-day mortality is a suitable metric for the comparison of hospital performance. Outcomes are 1 of the 3 domains for quality measurement of Donabedian.[23] Some organizations are currently using this measure, although their methods are often obscure.[20] From the hospital perspective, there are concerns about whether the true profile of their patients' risk can be taken into account. Although the percentage of explained variability was low, the probability of discriminating survivors and nonsurvivors, the most common metric for assessing binary-valued outcomes, was 70% with the use of administrative data and 77% with the use of medical record data. In addition, the unexplained variation is a result of unmeasured risk factors, quality of care, and random variation. In the medical record model, we have included the risk factors that are considered most important for early mortality. It is possible that novel risk factors will be identified or that some other unmeasured risk factors might have added incrementally to the model, but it is

TABLE 3. AMI Medical Record Model Based on CCP Data Set (1994 to 1995)

Variable Definition	Estimate	*t*	Odds Ratio	95% CI
Intercept	−3.45	−77.30		
Demographics				
Male	−0.18	−12.61	0.84	0.81–0.86
Age, years over 65	0.05	50.29	1.05	1.05–1.05
Noncardiac history				
Diabetes (any type)	0.10	6.87	1.10	1.07–1.14
Cardiac history				
CVA/stroke	0.28	16.06	1.32	1.28–1.37
Myocardial infarction	−0.06	−3.72	0.95	0.92–0.97
Hypertension	−0.21	−15.25	0.81	0.79–0.83
COPD	0.05	3.25	1.05	1.02–1.09
Bypass surgery	0.12	5.53	1.12	1.08–1.17
Angioplasty	−0.35	−11.43	0.70	0.66–0.75
Cardiac symptoms (first 48 h of admission)				
Systolic blood pressure, mm Hg				
≤100	0.00		1.00	
>100	−0.99	−51.30	0.37	0.36–0.39
Missing	0.97	13.22	2.64	2.29–3.05
Shock	1.49	42.76	4.42	4.12–4.73
Heart failure/PE/CHF on x-ray/rales/gallop rhythm or S3	0.39	25.00	1.48	1.44–1.53
Time since chest pain started (relative to hospital arrival)				
<6 h	0.00		1.00	
6–12 h	0.16	6.19	1.17	1.11–1.23
>12 h	0.18	8.57	1.20	1.15–1.24
No chest pain	0.33	19.29	1.39	1.34–1.44
Unable to determine chest pain time	0.32	14.56	1.38	1.32–1.44
Initial laboratory results (first 24 h of admission)				
BUN, mg/dL*	0.01	14.16	1.01	1.01–1.01
BUN missing	0.27	4.63	1.31	1.17–1.47
Creatinine, mg/dL*	0.58	31.88	1.78	1.72–1.85
Creatinine missing	0.87	14.30	2.38	2.11–2.68
White blood cell count, μL ×1000 (first 24 h of admission)				
<6			1.00	
6–12	0.29	9.23	1.34	1.26–1.43
>12	0.85	26.21	2.33	2.19–2.48
Missing	0.44	9.17	1.55	1.41–1.70
First ECG within 6 h before or after arrival				
ST-segment elevation	0.31	20.08	1.36	1.32–1.40
ECG unavailable	0.15	5.78	1.16	1.10–1.22
Left bundle-branch block	0.16	6.20	1.17	1.12–1.23
Right bundle-branch block	0.35	15.09	1.41	1.35–1.48
Second/third-degree heart block	0.28	5.55	1.32	1.20–1.46
Documented location(s) of AMI				
Anterior or lateral	0.56	37.64	1.75	1.70–1.81
No location determined	0.78	31.89	2.17	2.07–2.28

CVA indicates cerebrovascular accident; COPD, chronic obstructive pulmonary disease; PE, peripheral edema; CHF, congestive heart failure; and BUN, blood urea nitrogen. All $P<0.0001$.

Estimate of between-hospital variance=0.061 (SE=0.0052).

*Mean, if not measured.

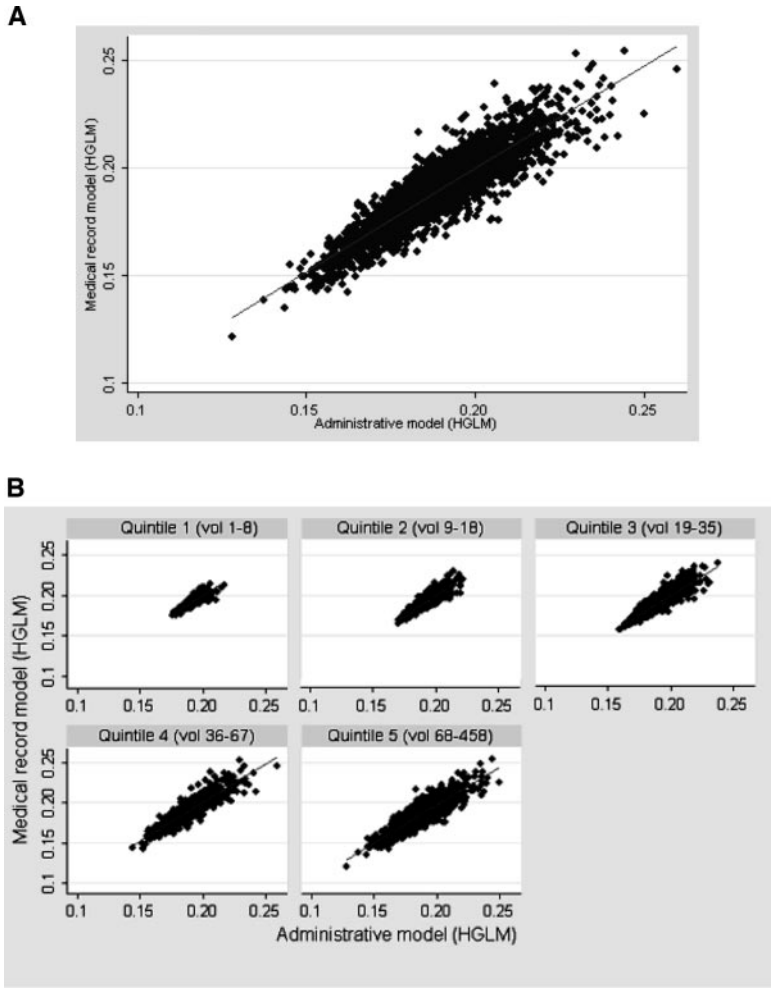

Figure 2. Comparison of the hospital-level risk-standardized mortality rates with the medical record model and the administrative model, overall (A) and stratified by volume (B).

unlikely that they would have markedly increased the explained variation. The value of outcome measurement over structure and process is that it represents what is most important to the patient. In addition, a focus on one process can lead to ignoring other important processes that are not measured in an attempt to be viewed positively. Assuming that mortality is appropriate for profiling, we selected 30 days because it is a standardized outcome within a time frame that

TABLE 4. AMI Administrative Model and Medical Record Model Performance

Model	Over-Fitting Indices (Intercept, Slope)	Discrimination		
		Adjusted R^{2*}	Predictive Ability† (Lowest Decile, Highest Decile)	ROC Curve Area
Administrative data derivation sample				
1998 (1st half) (n=140 120)	(0, 1)	0.12	4.0%–40.0%	0.71
Administrative data validation samples				
1998 (2nd half) (n=139 978)	(0.01, 0.99)	0.11	4.2%–40.1%	0.70
1995 (n=283 578)	(0.10, 0.99)	0.11	4.5%–39.3%	0.69
1996 (n=285 953)	(0.08, 1.00)	0.11	4.5%–39.1%	0.70
1997 (n=282 127)	(0.04, 1.001)	0.12	4.3%–39.0%	0.70
1999 (n=280 319)	(0.04, 1.01)	0.12	3.9%–40.6%	0.71
2000 (n=263 124)	(0.02, 1.00)	0.12	5.3%–40.6%	0.70
2001 (n=264 191)	(0.03, 1.00)	0.12	4.8%–41.1%	0.70
Linked administrative data model				
1994–1995 (n=181 032)	(0.02, 1.00)	0.10	5.3%–38.6%	0.69
Linked medical record data model				
1994–1995 (n=181 032)	(0, 1)	0.24	2.9%–59.0%	0.77

*Max-rescaled R^2.

†Observed rates in deciles determined by estimated model.

hospital quality would expect to influence and has been commonly used in clinical trials.

We did not exclude hospitals with small volumes because the hierarchical model takes the sample size of each unit into account. In general, hospitals that provide little information (have small volumes) will have predicted risk-standardized mortality rates that are near the national average because these institutions do not provide sufficient information for an informed estimate of their performance. Because these small-volume hospitals would cluster at the average, they would not have a strong influence on the correlation between the 2 methods of calculating hospital performance. Moreover, no matter what analytic approach is adopted, it is difficult to calculate precise estimates of outcome rates for hospitals or other healthcare units with small volumes.

Our analysis has some issues to consider. This study is the first to validate a claims model in US hospitals using this approach. We required a large national sample of claims and medical record data. The data were only available in the CCP, a quality improvement initiative that abstracted records from 1994 to 1995. Nevertheless, even a chart-based model has limitations, and there is no true gold standard against which to compare the claims model. However, the chart-based model currently provides the best characterization of the demographic and clinical characteristics of the patients. Another issue to consider is that coding practices have changed since that time, as has the definition of AMI, although our evaluation of the patient-level performance of the model over time revealed that it did not change. Finally, although the output from the administrative data model was highly correlated with that from the medical record data, our findings should not deflect attention from the need to improve the quality of data available for profiling. We focused on the relationship between risk-standardized estimates derived from 2 data sources; if the primary goal is to identify high-quality hospitals, ie, those in the upper 10th percentile, then it will also be important to assess the sensitivity and specificity of the claims-based estimates for this activity.[24] Improving *ICD* codes holds great promise for enhancing our ability to track outcomes such as complications, to elevate risk-adjustment approaches, and to avoid manipulation of coding. Advances in electronic health records that would provide medical record data in digital format for use in profiling hospital performance will also advance our ability to adjust for hospital differences in case mix. The use of administrative data should only be a temporary solution as higher quality data become available through the adoption of electronic health records.

This model is specific to Medicare fee-for-service patients and may not be generalizable to other data sources and patient populations. Nevertheless, Medicare patients represent a majority of the patients hospitalized with an AMI. Within Medicare, the fee-for-service patients are the vast majority. This effort was limited by the availability of national claims and medical record data. However, the ability to profile institutions by the experience of the Medicare fee-for-service patients represents an advance. As better, more comprehensive databases become available for broader populations, it will be natural to extend this approach.

Growing interest in the public reporting of outcomes has focused attention on the need for models that can adjust for differences in case mix among hospitals. Only administrative claims data are widely available to perform these types of analyses. In this study we derived an administrative claims model that characterizes the performance of hospitals in a manner similar to that produced by the medical record model. Thus, this claims model can produce information about hospital performance that is comparable to that obtained from higher-quality data that are much more expensive to obtain.

Acknowledgments

The analyses on which this publication is based were performed under contract 500-05-CO01, entitled "Utilization and Quality Control Quality Improvement Organization for the State of Colorado," sponsored by CMS (formerly Health Care Financing Administration), Department of Health and Human Services. The content of the publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. The authors assume full responsibility for the accuracy and completeness of the ideas presented. This article is a direct result of the Health Care Quality Improvement Program initiated by CMS, which has encouraged identification of quality improvement projects derived from analysis of patterns of care, and therefore required no special funding on the part of this contractor. Ideas and contributions to the authors concerning experience in engaging with issues presented are welcomed. The authors thank Dr Jeptha Curtis, Dr JoAnne Foody, Dr Robert McNamara, Deron Galusha, and Amy Rich from the Yale University School of Medicine; Neil Gittings from CMS; and Debra Chromik from the Colorado Foundation for Medical Care for their contributions to this work. CMS reviewed and approved the use of its data for this work and approved submission of the manuscript.

Disclosures

Dr Krumholz is a consultant to United Healthcare. The other authors report no conflicts.

References

1. Spertus JA, Radford MJ, Every NR, Ellerbeck EF, Peterson ED, Krumholz HM. Challenges and opportunities in quantifying the quality of care for acute myocardial infarction: summary from the Acute Myocardial Infarction Working Group of the American Heart Association/American College of Cardiology First Scientific Forum on Quality of Care and Outcomes Research in Cardiovascular Disease and Stroke. *Circulation*. 2003;107:1681–1691.
2. Centers for Medicare and Medicaid Services. Available at: http://www.cms.hhs.gov. Accessed February 8, 2006.
3. Joint Commission on Accreditation of Healthcare Organizations. Available at: http://www.jcaho.org. Accessed February 8, 2006.
4. Normand ST, Glickman ME, Sharma RG, McNeil BJ. Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients: results from the Cooperative Cardiovascular Project. *JAMA*. 1996;275:1322–1328.
5. Donabedian A. The role of outcomes in quality assessment and assurance. *QRB Qual Rev Bull*. 1992;11:356–360.
6. Brook RH, Iezzoni LI, Jencks SF, Knaus WA, Krakauer H, Lohr KN, Moskowitz MA. Symposium: case-mix measurement and assessing quality of hospital care. *Health Care Financ Rev*. 1987;Spec No:39–48.
7. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. *Ann Intern Med*. 1993;119:844–850.
8. Krumholz HM, Brindis RG, Brush JE, Cohen DJ, Epstein AJ, Furie K, Howard G, Peterson ED, Rathore SS, Smith SC, Spertus JA, Wang Y, Normand S-LT. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement

from the Quality of Care and Outcomes Research Interdisciplinary Writing Group. *Circulation*. 2006;113:456–462.
9. Marciniak TA, Ellerbeck EF, Radford MJ, Kresowik TF, Gold JA, Krumholz HM, Kiefe CI, Allman RM, Vogel RA, Jencks SF. Improving the quality of care for Medicare patients with acute myocardial infarction: results from the Cooperative Cardiovascular Project. *JAMA*. 1998;279:1351–1357.
10. Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Iezzoni LI, Ingber MJ, Levy JM, Robst J. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev*. 2004;25:119–141.
11. Normand SLT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. *J Am Stat Assoc*. 1997;92:803–814.
12. Shahian DM, Normand SL, Torchiana DF, Lewis SM, Pastore JO, Kuntz RE, Dreyer PI. Cardiac surgery report cards: comprehensive review and statistical critique. *Ann Thorac Surg*. 2001;72:2155–2168.
13. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical aspects of institutional performance. *J Royal Stat Soc*. 1996;159:385–444.
14. Krumholz HM, Chen J, Wang Y, Radford MJ, Chen YT, Marciniak TA. Comparing AMI mortality among hospitals in patients 65 years of age and older: evaluating methods of risk adjustment. *Circulation*. 1999;99:2986–2992.
15. Shahian DM, Torchiana DF, Shemin RJ, Rawn JD, Normand S-LT. The Massachusetts cardiac surgery report card: implications of statistical methodology. *Ann Thorac Surg*. 2005;80:2106–2113.
16. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med*. 1997;127:764–768.
17. Deleted in proof.
18. Harrell FE. *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis*. 3rd ed. New York, NY: Springer; 2001.
19. US News and World Report. Best hospitals 2005. Available at: http://www.usnews.com/usnews/nycu/health/hosptl/tophosp.htm. Accessed February 8, 2006.
20. Healthgrades. Available at: www.healthgrades.com. Accessed February 8, 2006.
21. Austin PC, Tu JV, Alter DA. Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: should we be analyzing cardiovascular outcomes differently? *Am Heart J*. 2003;145:27–35.
22. DeLong E. Hierarchical modeling: its time has come. *Am Heart J*. 2003;145:16–18.
23. Donabedian A. The quality of care: how can it be assessed? *JAMA*. 1988;260:1743–1748.
24. Normand S-LT, Wolf RE, Ayanian JZ, McNeil BJ. Assessing the accuracy of hospital clinical performance measures. *Med Decision Making*. In press.

CLINICAL PERSPECTIVE

Growing interest in the public reporting of outcomes has focused attention on the need for models that can adjust for differences in case mix among hospitals. Only administrative claims data are widely available to perform these types of analyses. A model using administrative claims data that is suitable for profiling hospital performance for acute myocardial infarction would be useful in quality assessment and improvement efforts. We developed a hierarchical regression model using Medicare claims data that produces hospital risk-standardized 30-day mortality rates that are similar to what can be derived from a medical record model. Thus, the results of this administrative model can be considered a surrogate for the results from the medical record model. This model has been endorsed by the National Quality Forum as a measure of hospital performance.

An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With an Acute Myocardial Infarction

Harlan M. Krumholz, Yun Wang, Jennifer A. Mattera, Yongfei Wang, Lein Fang Han, Melvin J. Ingber, Sheila Roman and Sharon-Lise T. Normand

Circulation. 2006;113:1683-1692; originally published online March 20, 2006;
doi: 10.1161/CIRCULATIONAHA.105.611186
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0009-7322. Online ISSN: 1524-4539