

An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With Heart Failure

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 heart failure would be useful in quality assessment and improvement efforts.

Methods and Results—We developed a hierarchical regression model using Medicare claims data from 1998 that produces hospital risk-standardized 30-day mortality rates. We validated the model by comparing state-level standardized estimates with state-level standardized estimates calculated from a medical record model. To determine the stability of the model over time, we used annual Medicare cohorts discharged in 1999–2001. The final model included 24 variables and had an area under the receiver operating characteristic curve of 0.70. In the derivation set from 1998, the 25th and 75th percentiles of the risk-standardized mortality rates across hospitals were 11.6% and 12.8%, respectively. The 95th percentile was 14.2%, and the 5th percentile was 10.5%. In the validation samples, the 5th and 95th percentiles of risk-standardized mortality rates across states were 9.9% and 13.9%, respectively. Correlation between risk-standardized state mortality rates from claims data and rates derived from medical record data was 0.95 (SE=0.015). The slope of the weighted regression line from the 2 data sources was 0.76 (SE=0.04) with intercept of 0.03 (SE=0.004). The median difference between the claims-based state risk-standardized estimates and the chart-based rates was <0.001 (25th percentile=−0.003; 75th percentile=0.002). The performance of the model was stable over time.

Conclusions—This administrative claims-based model produces estimates of risk-standardized state mortality that are very good surrogates for estimates derived from a medical record model. (*Circulation*. 2006;113:1693-1701.)

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

Patients with heart failure, the most common cause of admission among Medicare beneficiaries,¹ have a high risk of mortality.^{2,3} Current publicly reported process measures for assessing heart failure care are quite limited⁴ and may fail to discriminate between healthcare providers on the basis of their overall quality of heart failure care. The direct measurement of healthcare outcomes may complement efforts to characterize performance by process measures.⁵ Although some variation in outcome is beyond the control of clinicians and hospitals, quality of care and safety would be expected to influence the risk of adverse events in these patients.

Clinical Perspective p 1701

To use outcomes as an indicator of healthcare performance in the care of patients with heart failure requires statistical

methods that can evaluate the comparative performance of regions, health systems, and hospitals, taking into account any differences in case mix. Such statistical models should have several key attributes.⁶ First, the model and its performance should be in the public domain so that it can be properly evaluated by the groups it is assessing. Second, because the spectrum of patients may vary among regions and institutions, the model must adjust for differences in demographic and clinical characteristics. Third, the model should use an approach that is appropriate for the hierarchical organization of the data (eg, patients nested within institutions). Finally, the model ideally should be properly validated in different populations of patients and across organizations and institutions to which it will be applied. Moreover, for efforts using administrative claims data, which have limitations but are the only source of information for national

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

© 2006 American Heart Association, Inc.

Circulation is available at <http://www.circulationaha.org>

DOI: 10.1161/CIRCULATIONAHA.105.611194

profiling, validation should ideally include a comparison with statistical models that use higher-quality clinical data. The focus of the comparison should be on the output of the models with respect to characterizing performance at the organizational level rather than patient-level discrimination.

Our objective was to develop a statistical model based on administrative claims data that would be appropriate to profile hospitals, regions, and states by their 30-day mortality rates for patients admitted with a diagnosis of heart failure. Because medical record data were only available in sufficient numbers to perform a state-level comparison of the output of the 2 models, we determined whether the state estimates of risk-standardized mortality rates from the claims model could be used as surrogates for the results of the medical record model. We also evaluated the stability of the model over time.

Methods

Derivation and Validation Cohorts

The Derivation Cohort

We randomly sampled half of the hospitalizations for heart failure (*International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM]* codes 402.01, 402.11, 402.91, 404.01, 404.11, 404.91, 428.0, 428.1, 428.9) in the 1998 Medicare Provider Analysis and Review (MEDPAR) files, clustered within hospitals. For risk adjustment we used information contained 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 ≥ 65 years because these patients are representative of the older heart failure population and had at least 1 year of Medicare utilization data before their hospitalization. For patients who were transferred, we linked the hospitalizations into an episode of care. The information about the patient was derived from the hospital to which the patient was initially admitted. The initial hospital was also designated as the responsible institution for the episode.

We excluded patients who were not in fee-for-service Medicare for 1 year before their admission. For patients with multiple admissions during the study period, we randomly selected a hospitalization. We also excluded patients who were discharged alive and not against medical advice with a total length of stay ≤ 1 day because it is unlikely that these patients were admitted with decompensated heart failure.

The Validation Cohorts

The primary validation was a comparison of risk-standardized mortality rates between the claims model and the medical record model. To conduct this comparison, we constructed a linked sample that contained both claims and medical chart abstracted data from the National Heart Care (NHC) Project, a national heart failure quality improvement project sponsored by the Centers for Medicare & Medicaid Services (CMS).⁷ The first NHC sample included hospitalizations with a principal discharge diagnosis of heart failure between April 1998 and March 1999, inclusive, and the second between July 2000 and June 2001, inclusive. In both time periods, all identified discharged patients in each of the 50 states, Washington, DC, and Puerto Rico were sorted by age, sex, race, and hospital. Within each state, up to 800 discharges in each of the 2 sampling frames were randomly selected; a census of records was obtained in states with < 800 eligible discharged patients. Records were reviewed in central data abstraction centers for clinical data. Data quality was ensured through the use of trained abstractors, electronic abstraction instruments, and record reabstraction. Patients without

valid Social Security numbers, receiving long-term hemodialysis, transferred to another hospital, or leaving against medical advice were excluded from the NHC cohorts, which consisted of 39 477 records in 1998–1999 and 39 405 records in 2000–2001.

To evaluate the stability of the claims model over time, we examined the performance of the Medicare claims model using the other half of the 1998 MEDPAR data and data for each of years 1999, 2000, and 2001. For each year we created the study sample using the same approach as that used for the derivation cohort.

Outcome

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

Model Derivation: Patient Predictors of Mortality

We developed candidate variables for the Medicare claims model from the claims codes. Because there are $> 15\ 000$ ICD-9-CM codes, we used the Hierarchical Condition Categories (HCC) to assemble clinically coherent codes into candidate variables.⁸ 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 and from the principal and secondary diagnosis codes from hospitalizations, institutional outpatient visits, and physician encounters in the 12 months before the index hospitalization.

We conducted a clinical review of the candidate variables to exclude secondary diagnoses from the index hospitalization that may have represented complications rather than conditions present on admission. For example, because shock as a secondary code for the index hospitalization may not have been present at the time of admission, we did not include that code. We combined categories of HCC variables on the basis of clinical judgment and bivariate associations and eliminated candidate variables with a $< 1\%$ frequency. Additional candidate variables included demographic (age, sex) and procedural factors (history of bypass surgery or percutaneous coronary intervention in the past year).

Model Development

Because of the natural clustering of the observations within hospitals, we estimated hierarchical generalized linear models (HGLM).^{9–11} We modeled the log-odds of mortality within 30 days of admission as a function of patient demographic and clinical characteristics and a random hospital-specific effect. This strategy accounts for within-hospital correlation of the observed outcomes and models the assumption that underlying differences in quality among the healthcare groups being evaluated lead to systematic differences in outcomes.

We first selected the covariates for the final claims model using a backward elimination procedure through the generalized linear model (GLM) with a logit link function approach. Because of the large number of patient observations, we chose an exit criterion of $P > 0.01$. For each model, we calculated several indices for assessing model performance¹² at the patient level: the area under the receiver operating characteristic (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. Large values for the ROC area, R^2 statistic, and a large difference in predicted probabilities between highest and lowest deciles provide evidence that the model has good discrimination. We further assessed model fit through examination of Pearson residuals. Finally, we reestimated the regression coefficients of the covariates identified from our backward elimination strategy using a HGLM.

Model Validation

Medical Record Model

We chose risk factors for the medical record model on the basis of the medical literature and clinical experience.^{2,13–16} Unlike the

TABLE 1. HF Initial Administrative Claims Sample

Year	Total	Exclusions, %					Final Sample	
		Age <65 y	Incomplete Information	Repeated HF Admission in Year	Transfer	LOS ≤1 d	n	% of Total
1998	785 493	9.6	13.0	25.0	1.4	5.3	444 581	56.6
1999	656 847	9.1	3.9	23.3	0.5	5.5	422 552	64.3
2000	675 814	9.5	4.8	23.3	0.5	5.8	426 576	63.1
2001	684 641	10.0	6.4	23.6	0.5	5.9	422 351	61.7

HF indicates heart failure; LOS, length of stay.

claims data, some covariates could be missing for patients in the sample. We categorized continuous variables into categories using the clinically meaningful cut points and added a category for missing values where applicable. For discrete-valued variables, we included an additional level that indicated the variable was 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. We computed measures of model fit and discrimination for the medical record model similar to those computed for the claims-based models.

Risk-Standardized Mortality Rates

We calculated risk-standardized mortality rates for each hospital using the estimated hospital-specific parameters from the respective hierarchical models. For this analysis we modeled the log-odds of mortality within 30 days of admission as a function of patient demographic and clinical characteristics and a random hospital-specific effect. This strategy accounts for within-hospital correlation of the observed outcomes and models the assumption that there are underlying differences in quality among hospitals. These rates are obtained as the ratio of predicted to expected mortality, multiplied by the national unadjusted rate.¹⁷ The ratio is predicted mortality in each hospital, given its patient mix and hospital-specific effect divided by the expected mortality in that hospital given the same patient mix and the average hospital-specific effect.¹⁷ Although other researchers have calculated the ratio of observed to expected outcomes, we use the predicted rates to avoid several analytical problems that have been cited.^{9,11,18} The expected outcome for each hospital is the number of 30-day deaths expected in 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 using all hospitals in our sample, applying the subsequent estimated regression coefficients to the patient characteristics observed in the hospital, and then summing. This is a form of indirect standardization. The predicted hospital outcome is the number of expected mortalities in 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 within the hospital, applying the hospital-specific regression coefficients to the patient characteristics in the hospital, and then summing.

To assess the validity of the administrative data model, we repeated the aforementioned process, but, rather than calculating hospital-specific risk-standardized mortality rates, we calculated state-specific risk-standardized mortality rates and compared these rates with risk-standardized rates obtained from a medical record model. We conducted this analysis because medical record data were only available in sufficient numbers to perform a state-level comparison of the output of the 2 models. We used 2 approaches to examine the relationship between the risk-standardized rates obtained from administrative data and chart data. First, we estimated a linear regression equation describing the association between the 2 rates, weighting each state 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 state 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 compared the performance of the claims model over time in 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 dataset, each time calculating a risk score using the regression estimates from our derivation model.

After computing the overfitting statistics, in each validation dataset we recalibrated the model so that we used the same variables but fit the model to the data for each specific cohort. For each model, we calculated the same indices for assessing model performance¹² as in the derivation model.

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 785 493 heart failure discharges from 5146 hospitals in the national fee-for-service administrative claims database, of which 9.6%, 13.0%, and 5.3% of discharges were excluded for age <65 years, incomplete information in the 12 months before admission, and length of stay of ≤1 day, respectively (Table 1). Another 1.4% of the hospitalizations represented transfer in admission and were combined with the admission at the initial hospital to create an episode of care. In addition, 25.0% of the hospitalizations were repeat admissions. We randomly selected a single admission for each patient.

The derivation sample consisted of 222 424 cases with an unadjusted 30-day mortality rate of 12.1%. The mean age of the cohort was 79.6±7.7 years. The cohort included 59.3% women and 14.8% nonwhite patients. There were 5087 hospitals in the derivation cohort, with a median

TABLE 2. Administrative Claims Model: Heart Failure 30-Day Mortality (Based on 1998 Derivation Sample; n=222 424)

Variable	Estimate	<i>t</i>	Odds Ratio	95% CI
Intercept	-3.42	-142.51		
Demographic				
Age, years over 65	0.05	49.03	1.05	1.04–1.05
Male	0.25	17.88	1.28	1.24–1.31
Cardiovascular				
History of PTCA	-0.57	-9.42	0.57	0.50–0.64
History of CABG	-0.42	-16.14	0.66	0.63–0.69
History of heart failure (HCC 80)	0.45	26.64	1.57	1.52–1.62
History of MI (HCC 81)	0.22	9.61	1.24	1.19–1.30
Unstable angina (HCC 82)	-0.10	-4.99	0.90	0.87–0.94
Chronic atherosclerosis (HCC 83 and 84)	0.02	1.25	1.02	0.99–1.05
Cardiopulmonary-respiratory failure and shock (HCC 79)	0.18	9.47	1.20	1.16–1.25
Valvular heart disease (HCC 86)	0.15	9.71	1.16	1.12–1.19
Comorbidity				
Hypertension (HCC 89 and 91)	-0.34	-24.44	0.71	0.69–0.73
Stroke (HCC 95 and 96)	0.13	5.73	1.14	1.09–1.19
Renal failure (HCC 131)	0.43	23.65	1.53	1.48–1.59
COPD (HCC 108)	0.14	9.79	1.15	1.12–1.18
Pneumonia (HCC 111, 112, 113)	0.14	8.94	1.16	1.12–1.19
Diabetes (HCC 15–20, 120)	0.11	7.36	1.11	1.08–1.14
Protein-calorie malnutrition (HCC 21)	0.74	30.64	2.09	1.99–2.19
Dementia (HCC 49–50)	0.38	21.18	1.47	1.42–1.52
Hemiplegia, paraplegia, paralysis, functional disability (HCC 100, 101, 102, 68, 69, 177, 178)	0.18	6.64	1.19	1.13–1.26
Peripheral vascular disease (HCC 104, 105)	0.12	7.37	1.12	1.09–1.16
Metastatic cancer (HCC 7,8)	0.80	28.66	2.22	2.11–2.35
Trauma in last year (HCC 154–156,158–162)	0.09	5.79	1.09	1.06–1.12
Major psychiatric disorders (HCC 54, 55, 56)	0.10	3.83	1.10	1.05–1.16
Chronic liver disease (HCC 25, 26, 27)	0.40	8.24	1.50	1.36–1.65

PTCA indicates percutaneous transluminal coronary angioplasty; CABG, coronary artery bypass graft surgery; MI, myocardial infarction; and COPD, chronic obstructive pulmonary disease. Between-hospital variance=0.05 (SE=0.005). All HCC variables represent “or” conditions from all sources. Complication codes are not included (index admission only ones excluded).

annual number of Medicare heart failure hospitalizations of 28 (25th and 75th percentiles, 11 and 62, respectively). The observed mortality rate ranged from 0.0% to 100.0% across these hospitals, and the 25th, 50th, and 75th percentiles were 6.9%, 11.5%, and 16.7%, respectively.

The claims model included 24 variables (2 demographic, 8 cardiovascular, and 14 comorbidity variables) (Table 2). The model had good discrimination, calibration, and fit (Table 3). The area under the ROC curve was 0.71. The observed mortality rate increased from 3.0% in the lowest predicted decile to 28.5% in the highest predicted decile, a range of 24.5%. The adjusted R^2 was 0.10. Figure 1A and 1B shows the distributions of the standardized 30-day mortality rates overall and stratified by hospital heart failure volume. The 25th and 75th percentiles were 11.6% and 12.8%, respectively. The 95th percentile was 14.2%, and the 5th percentile was 10.5%.

Medical Record Validation

The NHC validation sample contained 46 700 hospitalizations from 4285 hospitals in 50 states and a crude 30-day mortality rate of 11.9%. The medical record comparison model in this cohort included 28 variables (Table 4). The area under the ROC curve was 0.78. The observed mortality rate ranged from 1.8% in the lowest predicted decile to 42.4% in the highest. As expected, the explained variation was higher in the chart-based model (R^2 was 0.22) than in the claims-based model.

In this cohort the administrative model had an area under the ROC curve of 0.70, an observed mortality rate ranging from 2.9% in the lowest predicted decile to 28.4% in the highest predicted decile, and an adjusted R^2 of 0.09. The estimated state-specific standardized mortality rates derived from each model are displayed in Figure 2.

TABLE 3. Heart Failure Administrative Model and Medical Record Model Performance

Model	Overfitting Indices (Intercept, Slope)	Discrimination		
		Adjusted R^2 *	Predictive Ability† (Lowest Decile, Highest Decile)	ROC Curve Area
Administrative derivation sample				
1998 (n=222 424)	(0, 1)	0.10	3.0%–28.5%	0.71
Administrative validation samples				
1998 (n=222 157)	(−0.004, 0.99)	0.10	2.8%–29.0%	0.70
1999 (n=422 552)	(−0.0002, 0.99)	0.09	3.2%–28.2%	0.70
2000 (n=426 576)	(−0.06, 0.97)	0.10	3.1%–27.9%	0.70
2001 (n=422 351)	(−0.006, 0.99)	0.10	3.0%–28.1%	0.70
Medical record sample				
1998–2001 (n=46 700)	(0, 1)	0.22	1.8%–42.4%	0.78
Linked administrative sample				
1998–2001 (n=46 700)	(−0.005, 1.00)	0.09	2.9%–28.4%	0.70

*Max-rescaled R^2 .
†Observed rates.

The slope of the weighted regression line of the state-specific mortality rates is 0.76 (SE=0.04), and the intercept is 0.03 (SE=0.004). The correlation coefficient of the standardized mortality rates from the 2 models is 0.95 (SE=0.02). The

median difference between the models in the state-specific risk-standardized mortality rates was <0.001 (25th percentile, −0.003; 75th percentile, 0.02; 10th percentile, −0.006; 90th percentile, 0.004).

Model Performance in the Administrative Validation Set

In each validation cohort, the model fit was similar to that of the derivation cohort (Table 3). These comparisons spanned 3 years of Medicare admissions for heart failure. The unadjusted mortality ranged from 11.5% to 12.2% across years of data. The percent explained variation ranged from 0.09 to 0.10, and the area under the ROC curves was 0.70.

Discussion

We developed an administrative claims-based model for calculating 30-day case mix–adjusted heart failure mortality rates in Medicare fee-for-service patients. This model has recently been endorsed by the National Quality Forum. Although the deficiencies of administrative data are well known,²⁰ the risk-standardized estimates from this model at the state level are highly correlated with the estimates obtained from a medical record model. Thus, for the purposes of profiling, the claims model was a very good surrogate for the medical record model, and the model was very stable over time. The medical record model we used had good discrimination, rivaling a model that was recently published by a Canadian group.² We note that our comparison between data sources focused on risk-standardized estimates; investigators who wish to use different functions of the estimates, such as the percentage of hospitals falling into a particular quantile, will need to undertake an assessment of the comparability of the 2 data sources.

Importantly, the claims model only includes information about the patients that is known on admission. Secondary diagnoses in administrative claims may represent conditions present on admission or those that develop during the hospitalization. Thus, we omitted secondary diagnostic codes

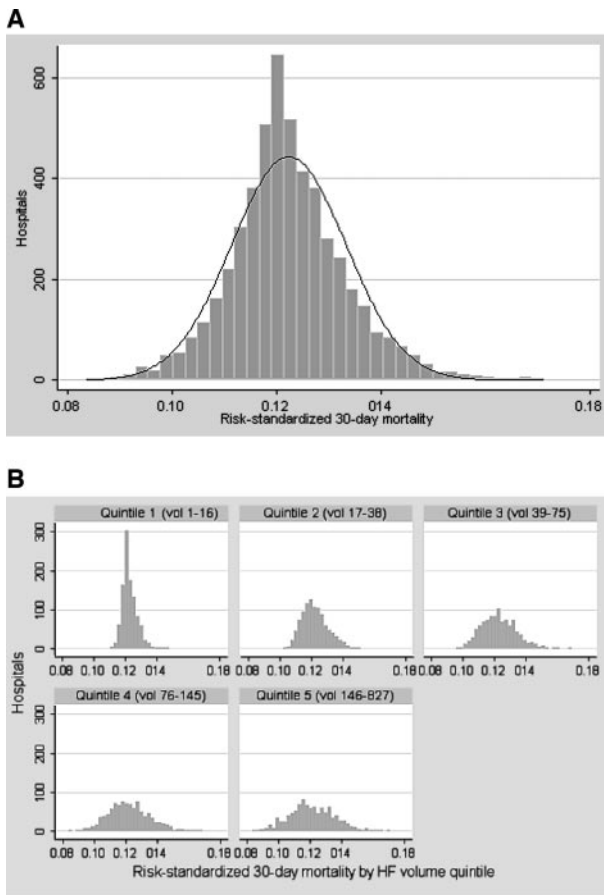


Figure 1. Distributions of hospital-level risk-standardized 30-day heart failure mortality rates using the administrative model on the basis of 1998 data (A, overall; B, stratified by volume). HF indicates heart failure.

TABLE 4. Chart-Based Model: Heart Failure 30-Day Mortality (Clinical-Based Model n=46 700)

Variable	Estimate	<i>t</i>	Odds Ratio	95% CI
Intercept	-3.93	-36.78		
Demographic				
Age, %				
75–84 y	0.38	8.81	1.47	1.35–1.59
≥85 y	0.76	16.46	2.13	1.95–2.33
Male	0.17	5.46	1.19	1.12–1.27
Noncardiac history				
Unable to walk independently	0.75	13.05	2.12	1.89–2.37
Dementia/Alzheimer's disease	0.63	14.87	1.88	1.73–2.04
Diabetes	-0.13	-4.00	0.87	0.82–0.93
Cardiac history				
Cerebrovascular accident	0.24	6.30	1.27	1.18–1.36
Hypertension	-0.24	-7.60	0.78	0.74–0.84
History of PTCA	-0.57	-4.37	0.56	0.43–0.73
History of CABG	-0.48	-8.02	0.62	0.55–0.70
Vital signs (first 48 h of admission)				
SBP <125 mm Hg	0.76	22.65	2.14	2.01–2.29
SBP unknown	0.81	1.36	2.26	0.70–7.27
DBP <90 mm Hg	0.39	8.68	1.48	1.35–1.61
DBP unknown	0.13	0.30	1.14	0.49–2.64
Heart rate, bpm				
<60	-0.23	-3.16	0.80	0.69–0.92
>100	0.33	9.43	1.39	1.30–1.49
Unknown	0.87	2.32	2.38	1.15–4.96
Cardiac symptoms (first 24 h of admission)				
Congestive heart failure	0.17	4.72	1.18	1.10–1.27
Cardiac arrest	3.06	34.73	21.29	17.92–25.31
Aortic stenosis	0.26	5.07	1.30	1.18–1.44
LVEF				
<40	0.38	2.79	1.46	1.12–1.90
Unknown	0.18	2.15	1.20	1.02–1.41
Initial laboratory results (first 24 h of admission)				
Sodium <145	0.30	4.41	1.35	1.18–1.55
Potassium >5	0.47	11.40	1.60	1.48–1.74
BUN/creatinine	0.99	28.60	2.70	2.52–2.89
BUN/creatinine unknown	0.84	7.76	2.31	1.87–2.85
White blood cell count, $\mu\text{L} \times 1000$ (first 24 h of admission)				
>20	0.90	11.01	2.45	2.09–2.87
Unknown	0.07	0.82	1.07	0.91–1.25

PTCA indicates percutaneous transluminal coronary angioplasty; CABG, coronary artery bypass graft surgery; SBP, systolic blood pressure; DBP, diastolic blood pressure; LVEF, left ventricular ejection fraction; and BUN, blood urea nitrogen. Between-state variance is 0.021 (SE=0.0068).

that may have represented complications, avoiding a scenario of making a hospital with a high rate of complications appear to be admitting patients with greater illness.

Another model using this methodology was developed for assessing the performance of the care of acute myocardial infarction.¹⁹ The findings in this study are similar, providing evidence that an administrative model can produce results that are comparable to those of a medical record model. In the

case of acute myocardial infarction, we were able to perform the comparison of the claims model with the medical record model at the hospital level, using data from the Cooperative Cardiovascular Project.²² For our study we were only able to compare performance at the state level because of the availability of data. At a hospital level there may be less agreement, but in the hospital-level analysis of acute myocardial infarction the findings were very similar.

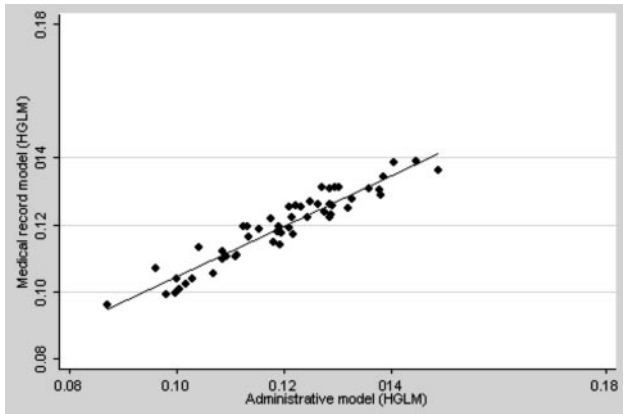


Figure 2. Comparison of the state-level risk-standardized mortality rates with the medical record model and the administrative model.

We counted only a single admission from each patient in each time period that was evaluated. Although this approach resulted in the loss of some information, it was necessary because mortality rates would be associated with readmission rates if all readmissions were included. Prior studies have indicated very high readmission rates for Medicare patients discharged after an episode of heart failure.²³

An important aspect of our model is that it is in the public domain. Other publicly reported models of heart failure outcomes are proprietary.²⁴ It is not possible to determine which variables are included or how well the model performs. Shielding this information from the public ensures that the validity of these models cannot be evaluated.

We employed hierarchical modeling in developing this model, which accounts for the clustering of the data (ie, patients within hospitals).^{10,25,26} Patients within hospitals have characteristics that are more highly correlated than patients in different hospitals. The relatedness of the observations can lead to underestimation of the SEs and cause the false appearance of statistically significant differences. In addition, hierarchical modeling can take into account differences in the amount of information provided by each hospital and allow small-volume hospitals to be retained in the analysis.

An important issue is whether 30-day mortality is a suitable metric for evaluating hospital performance for patients with heart failure. For some patients with end-stage heart failure, death is not an adverse event but rather the inevitable consequence of a long, chronic illness. For some patients who die, quality of care plays no role. However, our assumption is that most patients hospitalized with heart failure prefer to survive the hospitalization. Moreover, quality of care is associated with the risk of dying. Finally, the fairest way to assess hospital outcomes is to look at a standardized period of time that is fairly proximate to the time of the initial hospitalization. For these reasons, we chose to use 30-day mortality. We were not able to include resuscitation status in this model. However, even if it were available it is not clear that it should be used as a covariate. The designation of do-not-resuscitate (DNR) is distinct from a decision to provide comfort care only. Patients with DNR may want

high-quality care that will enhance the likelihood that they will survive. What DNR means for many individuals is that they do not want extraordinary means to keep them alive should their condition worsen. Thus, whether patients with a DNR status should not be considered in this assessment is not clear. We need a marker for patients who are admitted to the hospital for palliative care and for whom survival is not a goal of treatment. In the absence of this type of information, it will be difficult to incorporate the treatment goals into the outcomes assessment, which is a limitation of any such measure. For this to be a problem, there would have to be marked differences among the hospitals in the number of patients admitted who prefer comfort care only. We anticipate that this population is very small in comparison to the patients admitted who prefer high-quality care that will increase their likelihood of at least short-term survival.

We also note that in the linked sample, explained variation was low: 9% with administrative data and 22% with medical record data. Even the best models that predict outcomes in medicine have a substantial amount of unexplained variation. This unexplained variation is a result of unmeasured risk factors, quality of care, and random variation. In the medical record model, which is used for validation of the administrative claims 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 unlikely that they would have markedly increased the explained variation. We are left with the inference that much of the unexplained variation is the result of the care that was provided and random variation.

Administrative data from Medicare have limitations but are the only currently available national data that can assess hospital outcomes for heart failure. Approximately 70% of heart failure hospitalizations occur in patients aged ≥ 65 years. As a result, Medicare data are highly representative of this population. More timely chart information would be preferable, but the burden on institutions would be considerable with current technology.

Some hospitals accept in referral many patients with end-stage heart failure for transplantation or other high-technology interventions. Medicare patients, however, are not generally candidates for these approaches, and therefore the expectation is that heart failure centers will not suffer from adverse selection when only Medicare patients are considered. Improving routinely collected data holds great promise for enhancing our ability to track outcomes, to elevate risk adjustment approaches, and to avoid manipulation of coding. For now, we only have administrative data with which to perform this type of profiling.

This model was developed on the basis of the heart failure codes available during the time periods assessed. We could not validate the codes in the administrative dataset, but studies suggest that they have a very high specificity and positive predictive value.²⁷ The introduction of new codes is unlikely to affect these models because they would involve a redistribution of patients but would not be expected to move a patient out of a heart failure code. Nevertheless, these models are expected to undergo continual evaluation with

efforts to improve them over time as new codes and data become available.

The direction of the coefficients in the models deserves comment. Several variables, such as hypertension, history of procedures, and unstable angina, have negative coefficients. The direction of these coefficients is consistent with the chart review model except for unstable angina, which was not included. These variables may be related to cardiac function (independent of ejection fraction, which is included in the medical record model) or may be a marker for other patient characteristics that are associated with a favorable prognosis. More work is necessary to understand how these factors may mediate their association. In addition, one variable, diabetes, has a positive coefficient in the administrative model and a negative coefficient in the medical record model. It is important to note that the variables are very different. In the administrative claims we determined whether there had been a claim in the prior year for diabetes. This may identify a person with more severe diabetes and someone who is seeking care for the condition. In the medical record we sought any documentation of diabetes at any time and had no information on whether the patients were seeking care for it and did not require that they were being treated for it. The differences in the definitions and in the associated covariates likely led to the different directions of the β -coefficient of the covariate.

Another important consideration is that the claims models depend on data that are available only from CMS. The generalizability of the findings to other populations cannot be tested. Nevertheless, the vast majority of patients admitted to the hospital with heart failure are in the Medicare population, in which fee-for-service is the most common form of coverage. The ability to profile the performance of hospitals by their experience with this large patient group is useful. However, this approach may not represent performance with other groups of patients with heart failure.

In conclusion, we developed a model using administrative claims data that is suitable for profiling hospital heart failure outcomes. The model is in the public domain and demonstrates consistent performance over time. In addition, it produces results that can serve as a surrogate for those from a medical record model. Despite the limitations of currently available data, this model may be a valuable tool in assessing the outcomes achieved by states and hospitals in caring for patients with heart failure.

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 Jephtha 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

- Centers for Medicare and Medicaid Services. Available at: http://www.cms.hhs.gov/statistics/feeforservice/DRG_Rank_Discharge.pdf. Accessed December 27, 2005.
- Lee DS, Austin PC, Rouleau JL, Liu PP, Naimark D, Tu JV. Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model. *JAMA*. 2003;290:2581–2587.
- Rathore SS, Foody JM, Wang Y, Smith GL, Herrin J, Masoudi FA, Wolfe P, Havranek EP, Ordin DL, Krumholz HM. Race, quality of care, and outcomes of elderly patients hospitalized with heart failure. *JAMA*. 2003;289:2517–2524.
- Jencks SF, Huff ED, Cuedon T. Change in the quality of care delivered to Medicare beneficiaries, 1998–1999 to 2000–2001. *JAMA*. 2003;289:305–312.
- Donabedian A. The role of outcomes in quality assessment and assurance. *QRB Qual Rev Bull*. 1992;11:356–360.
- 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 Research Interdisciplinary Writing Group. *Circulation*. 2006;113:456–462.
- Masoudi FA, Ordin DL, Delaney RJ, Krumholz HM, Havranek EP. The National Heart Failure Project: a Health Care Financing Administration initiative to improve the care of Medicare beneficiaries with heart failure. *CHF*. 2000;6:337–339.
- 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.
- 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.
- 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.
- Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical aspects of institutional performance. *J Royal Stat Soc*. 1996;159:385–444.
- Harrell FE. *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis*. 3rd ed. New York, NY: Springer; 2001.
- Fonarow GC, Adams KF Jr, Abraham WT, Yancy CW, Boscardin WJ, for the ADHERE Scientific Advisory Committee, Study Group, and Investigators. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA*. 2005;293:572–580.
- Curtis JP, Sokol SI, Wang Y, Rathore SS, Ko DT, Jadbabaie F, Portnay EL, Marshalko SJ, Radford MJ, Krumholz HM. The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure. *J Am Coll Cardiol*. 2003;42:736–742.
- Wexler DJ, Chen J, Smith GL, Radford MJ, Yaari S, Bradford WD, Krumholz HM. Predictors of costs of caring for elderly patients discharged with heart failure. *Am Heart J*. 2001;142:350–357.
- Krumholz HM, Chen Y-T, Wang Y, Vaccarino V, Radford MJ, Horwitz RI. Predictors of readmission among elderly survivors of admission with heart failure. *Am Heart J*. 2000;139:72–77.
- 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.

18. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med.* 1997;127:764–768.
19. Deleted in proof.
20. 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.
21. Krumholz HM, Wang Y, Mattera JA, Wang Y, Han LF, Ingber MJ, Roman S, Normand S-LT. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation.* 2006;113:1683–1692.
22. 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.
23. Krumholz HM, Parent EM, Tu N, Vaccarino V, Wang Y, Radford MJ, Hennen J. Readmission after hospitalization for congestive heart failure among Medicare beneficiaries. *Arch Intern Med.* 1997;157:99–104.
24. Healthgrades. Available at: <http://www.healthgrades.com>. Accessed February 9, 2006.
25. 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.
26. DeLong E. Hierarchical modeling: its time has come. *Am Heart J.* 2003;145:16–18.
27. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. *Med Care.* 2005;43:480–485.

CLINICAL PERSPECTIVE

A model using administrative claims data that is suitable for profiling hospital performance for heart failure would be useful in quality assessment and improvement efforts. Administrative data from Medicare have limitations but are the only currently available national data that can assess hospital outcomes for heart failure. Only administrative claims data are widely available to perform these types of analyses. We developed a hierarchical regression model using Medicare claims data that produces hospital risk-standardized 30-day mortality rates and validated them at a state level against results 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 Heart Failure**

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;113:1693-1701; originally published online March 20, 2006;
doi: 10.1161/CIRCULATIONAHA.105.611194

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2006 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the
World Wide Web at:

<http://circ.ahajournals.org/content/113/13/1693>

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

Reprints: Information about reprints can be found online at:
<http://www.lww.com/reprints>

Subscriptions: Information about subscribing to *Circulation* is online at:
<http://circ.ahajournals.org/subscriptions/>