Results of Report Cards for Patients with Congestive Heart Failure Depend on the Method Used To Adjust for Severity

Roy M. Poses, MD; Donna K. McClish, PhD; Wally R. Smith, MD; Elizabeth C. Huber, MD; F. Lynne W. Clemo, MD; Brian P. Schmitt, MD; Donna Alexander, PhD; Edward M. Racht, MD; and Christopher C. Colenda III, MD

Background: **The validity of outcome report cards may depend on the ways in which they are adjusted for risk.**

Objectives: **To compare the predictive ability of generic and disease-specific survival prediction models appropriate for use in patients with heart failure, to simulate outcome report cards by comparing survival across hospitals and adjusting for severity of illness using these models, and to assess the ways in which the results of these comparisons depend on the adjustment method.**

Design: **Analysis of data from a prospective cohort study.**

Setting: **A university hospital, a Veterans Affairs (VA) medical center, and a community hospital.**

Patients: **Sequential patients presenting in the emergency department with acute congestive heart failure.**

Measurements: **Unadjusted 30-day and 1-year mortality across hospitals and 30-day and 1-year mortality adjusted by using disease-specific survival prediction models (two sickness-at-admission models, the Cleveland Health Quality Choice model, the Congestive Heart Failure Mortality Time-Independent Predictive Instrument) and generic models (Acute Physiology and Chronic Health Evaluation [APACHE] II, APACHE III, the mortality prediction model, and the Charlson comorbidity index).**

Results: **The community hospital's unadjusted 30-day survival rate (85.0%) and the VA medical center's unadjusted 1-year survival rate (60.9%) were significantly lower than corresponding rates at the university hospital (92.7% and 67.5%, respectively). No severity model had excellent ability to discriminate patients by survival rates (all areas under the receiver-operating characteristic curve < 0.73). Whether the VA medical center, the community hospital, both, or neither had worse survival rates on simulated report cards than the university hospital depended on the prediction model used for adjustment.**

Conclusions: **Results of simulated outcome report cards for survival in patients with congestive heart failure depend on the method used to adjust for severity.**

Ann Intern Med. 2000;133:10-20.
For author affiliations, current addresses, and contributions, see end of text.

Report cards may be used to compare the quality of health care according to use of services; performance of health care processes; or outcomes across providers, hospitals, or other organizations. Recently, outcome report cards in particular have received much attention. Managed care organizations and commercial health care consultants are using them more frequently to measure quality (1), and the former may use them to make decisions about financial incentives and provider participation (2). The Joint Commission on Accreditation of Healthcare Organizations now requires hospitals to measure their performance by using commercial systems, including some that incorporate outcome report cards (3). The American Board of Medical Specialties is considering requiring outcome report card evaluations for certification (4). Pennsylvania (5), California (6), Florida (7), New Jersey (8), Connecticut (9), Ontario, Canada (10), and the United Kingdom's National Health Service (11) have disseminated comparisons of hospitals' mortality outcomes to consumers. Hospitals with extreme mortality rates have received considerable publicity (6, 9, 12–14), which may affect their fortunes in an increasingly competitive health care market.

However, the validity of outcome report cards may depend on the methods of risk adjustment used to construct them (15). Fair report cards should distinguish the effects of providers or organizations on outcomes from the effects of chance and patients' underlying prognoses (16). Report cards that do not adequately adjust for patients' prognostic characteristics could adversely affect quality by penalizing facilities that care for the sickest patients (17).

Risk adjustment methods based only on administrative data (18) or generic methods (that is, methods applicable to diagnostically heterogeneous groups of patients) based on clinical data may not fully adjust for all prognostic differences among patients of different providers (19). Results of outcome report cards that use generic risk adjustment methods may vary depending on the choice of method (20).

There is hope that newer, disease-specific risk adjustment methods based on clinical data may perform better

Table 1. **Unadjusted Survival Rates across Hospitals**

Hospital	30-Day Survival		1-Year Survival	
	Patients Surviving	Relative Risk (95% CI)	Patients Surviving	Relative Risk (95% CI)
	n/n (%)		*n/n (%)*	
University hospital	433/467 (92.7)	1.0 (referent)	315/467 (67.5)	1.0 (referent)
Veterans Affairs medical center	580/649 (89.4)	0.96 (0.93–1.001)	395/649 (60.9)	0.90 (0.83–0.99)
Community hospital	187/220 (85.0)	0.92 (0.86–0.97)	132/220 (60.0)	0.89 (0.78–1002)

(21–23). For example, some models predict survival specifically for patients with congestive heart failure (24–26), an increasingly prevalent disease (27) and a leading cause of hospitalization for elderly persons (28). Studies have compared survival across hospitals for specific diseases, including heart failure, by using disease-specific risk adjustment methods (29, 30). However, whether outcome report cards that use these methods will provide consistent and valid results is unknown.

In this study, we compared the predictive ability of several prediction models specific to heart failure and several generic models appropriate for patients with heart failure; compared survival across hospitals, adjusting for severity of illness by using these models (thus simulating survival outcome report cards); and assessed the extent to which the results of these comparisons depended on the adjustment method.

Methods

Study Design, Sites, and Sample

We performed a retrospective cohort study of patients from the Predictions and Outcomes of Congestive Heart Failure (POCHF) study (31). From 5 November 1990 to 31 December 1992, we prospectively enrolled consecutive patients older than 16 years of age who had received a clinical diagnosis of acute heart failure from their evaluating physicians in the emergency departments of a university hospital, a Veterans Affairs (VA) medical center, or a community hospital in one small metropolitan area. We excluded patients with possible or definite acute myocardial infarction. Institutional review boards at all three hospitals approved the study. Patients were enrolled regardless of their disposition after their initial emergency department visit and could be enrolled each time they visited the emergency department with acute heart failure.

Housestaff from a single internal medicine program provided care at the university hospital and the VA medical center. Many of the attending physicians at one hospital had some clinical or teaching duties at the other hospital. Approximately 47% of housestaff but only 7% of attending physicians cared for study patients at both the university hospital and the VA medical center. The community hospital had a separate, private attending staff; its only housestaff were on a single team that was present during part of the study.

Survival Assessment

The main study outcome variable was survival, measured 30 days and 1 year after the emergency department visit. For each patient, we first attempted to assess survival using one of the following primary data sources: a data form completed by the emergency department physician, chart review, interview of the patient or surrogate within approximately 4 days of the visit, follow-up telephone interviews 90 days and 1 year after the visit, or clinical databases of patient encounters and laboratory results at the university hospital and the VA medical center. For patients whose vital status could not be unambiguously ascertained from these primary sources (7% of patients at 30 days and 8% at 1 year), we searched state and federal vital status databases. We recorded death dates from these secondary sources only for patients for whom we had an exact match on two and at least an approximate match on one of three characteristics: name, birth date, and Social Security number.

Choice of Mortality Prediction Models

The risk adjustment methods we chose were survival prediction models that were designed to assess quality of care or institutional performance for acutely ill patients, were designed exclusively for or at least to include patients with heart failure, and had required data collected for them by the POCHF study. We chose four models that were specific to heart failure: two sickness-at-admission (SA) models designed to predict 30- and 180-day survival (25), a revised version (32) of the original Cleveland Health Quality Choice (CHQC) model (29), and the Congestive Heart Failure Mortality Time-Insensitive Predictive Instrument (CHFM-TIPI) (26). We chose four generic models that were designed to predict mortality for hospitalized patients with various admitting diagnoses, including heart failure: the Acute Physiology and Chronic Health Evaluation (APACHE) II (33); APACHE III (34); the admission mortality prediction model (MPM_0) (35); and the Charlson comorbidity index (CCI) (36).

The SA models (37), the CHQC model (29), CHFM-TIPI (26), APACHE II (38), and APACHE III (39, 40) have been used in published comparisons of hospital mortality. The predictive performance of all models except CHFM-TIPI has been validated on a sample of patients distinct from the sample that participated in model development (29, 33, 35, 40–45). The models were designed to predict survival at hospital discharge (CHFM-TIPI, APACHE II, APACHE III, MPM_0) or at 30 days (SA-30, CHQC), 180 days (SA-180), or 1 year after admission (CCI).

All models provide a score indicating a patient's probability of survival, which is calculated by using data from multiple variables. No model uses the same set of variables as another, and each has a unique scheme to weight and combine variables. A list of variables is given in the **Appendix Table.**

The CHFM-TIPI, MPM_0, and CCI used only data that were available before admission. The others used some data collected within 1 to 2 days after hospitalization. The two SA models, the CHQC model, and APACHE III were designed to impute normal values when data items about physiologic status or laboratory test results were missing for a particular patient. The CHFM-TIPI, APACHE II, and MPM_0 required that such patients be excluded. The CCI used only data about preexisting comorbid conditions.

Data Collection for Prediction Models and Calculation of Model Scores

We used retrospective chart review to collect most of the data required by the models. Trained physicians and nurses who were unaware of the study questions completed a standardized data form, and a senior cardiology fellow reviewed emergency department electrocardiograms. When information about the presence of preexisting conditions was unavailable for a particular visit, we used information obtained from the patient's previous study enrollment when it was available.

For each patient, we computed a score for each of the eight models, following as closely as possible the procedures laid out in the original publications describing the models. Because all of the generic models used admitting diagnosis as a variable but our sample was restricted to patients with heart failure, we omitted the admitting diagnosis when computing scores.

Because four of the eight models were designed to impute normal values for missing data about physiology and diagnostic tests, we imputed as normal the values of all such missing variables for all models to make fair comparisons. We found that 3.8% to 58.8% of all patients and 1.2% to 44.5% of all hospitalized patients had at least one

Table 2. **Clinical Characteristics of Patients across Hospitals**

Characteristic	Patients at University Hospital (*n* = 467)	Patients at Veterans Affairs Medical Center (*n* = 649)	Patients at Community Hospital (*n* = 220)	*P* Value*
Female, %	51.0	1.0	56.4	0.001
White, %	17.6	55.8	85.4	0.001
Dyspnea on mild exertion or rest, %	61.5	75.5	51.8	0.001
Acute orthopnea, %	78.4	78.0	55.0	0.001
Paroxysmal nocturnal dyspnea, %	64.2	66.0	43.6	0.001
Receiving a diuretic, %	67.9	70.7	59.1	0.006
Receiving angiotensin-converting enzyme inhibitors, %†	47.3	50.0	30.9	0.001
Hypertension, %	78.4	78.1	55.5	0.001
Intensive care unit admission in past year, %	42.2	43.3	46.4	>0.2
Possible valvular heart disease, %	28.9	31.1	19.6	<0.001
Coronary artery disease, %†	40.5	56.6	40.5	0.001
Mean age, *y*	63.0	68.2	74.2	<0.001

* Comparing differences in proportions across hospitals.
† Past angina, coronary artery bypass graft surgery, percutaneous transluminal coronary angioplasty, or myocardial infarction.

Table 3. **Proportions of Patients Whose Scores Were Ordered Differently by at Least One Quintile in Each Possible Pair of Survival Models***

Model	Model							
	SA-30	SA-180	CHQC	CHFM-TIPI	APACHE II	APACHE III	MPM_0	CCI
SA-30		0.28 0.28	0.33 0.28	0.32 0.36	0.34 0.34	0.30 0.30	0.33 0.35	0.40 0.36
SA-180			0.34 0.33	0.34 0.39	0.33 0.32	0.33 0.35	0.34 0.37	0.39 0.35
CHQC				0.34 0.34	0.37 0.38	0.37 0.40	0.36 0.35	0.41 0.37
CHFM-TIPI					0.39 0.35	0.36 0.35	0.35 0.28	0.42 0.37
APACHE II						0.31 0.32	0.35 0.35	0.40 0.37
APACHE III							0.35 0.35	0.40 0.37
MPM_0								0.39 0.35

* The top figure in each cell is the proportion of patients who were put in a more severe quintile of mortality risk by the model to the left of the row compared with the model at the top of the column. The bottom figure in each cell is the proportion of patients who were put in a more severe quintile of mortality risk by the model at the top of the column compared with the model to the left of the row. APACHE = Acute Physiology and Chronic Health Evaluation; CCI = Charlson comorbidity index; CHFM-TIPI = Congestive Heart Failure Mortality Time-Insensitive Predictive Instrument; CHQC = Cleveland Health Quality Choice; MPM_0 = mortality prediction model; SA-30 = sickness at admission (30-day survival); SA-180 = sickness at admission (180-day survival).

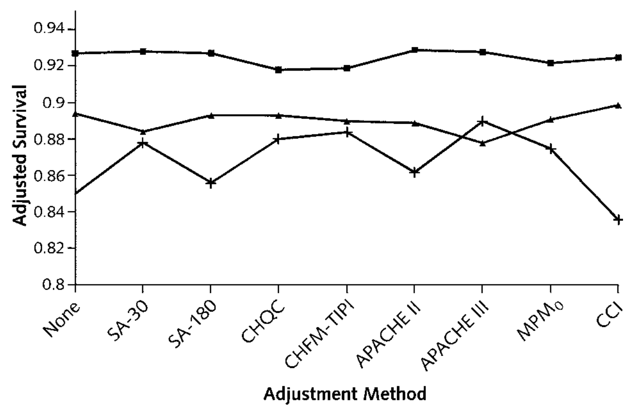
missing model variable. Three of the models used only data available in the emergency department before possible hospital admission. Therefore, to avoid bias caused by changes that occurred in the value of variables as a result of in-hospital management decisions, we used only data on physiologic and diagnostic test variables obtained before the patient left the emergency department.

Statistical Analysis

We excluded patients with missing charts or survival data. Our main analyses included all patients regardless of whether they were hospitalized, but additional analyses excluded patients who were not hospitalized.

To assess predictive performance, we constructed logistic regression models in which the prediction model score was the independent variable and survival was the dependent variable. These analyses yielded a C-statistic for each model, which is equivalent to the area under the receiver-operating characteristic (ROC) curve, a measure of the prediction model's ability to discriminate among patients by ordering them according to rates of the outcome event (46). These logistic regression analyses recalibrated the models to reflect patients' overall survival rate and yielded Hosmer–Lemeshow statistics as indices of the accuracy of the recalibrated models (47). The researchers who developed some of the prediction models provided schemes to map model scores onto predicted rates of survival. We did not evaluate the accuracy of these schemes, however, because the mappings reflected the survival rates of the samples used to develop them rather than rates of our sample (48).

Figure 1. **Results of simulated report cards for 30-day survival for the entire cohort.**

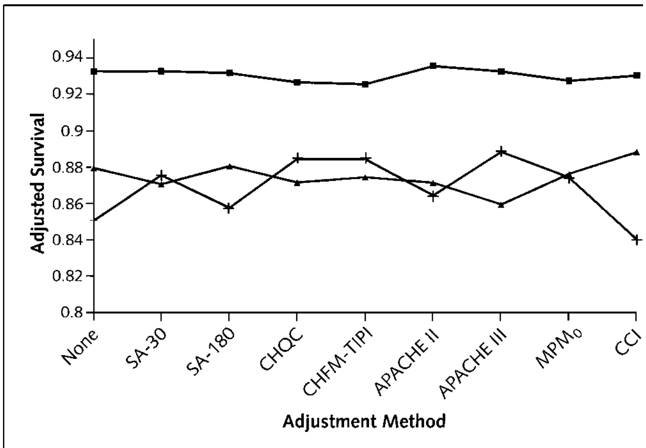


Lines with squares represent the university hospital, lines with triangles represent the Veterans Affairs (VA) medical center, and lines with crosses represent the community hospital. The university hospital and the VA medical center differed significantly for survival adjusted by the sickness-at-admission (30-day survival) (*SA-30*) model, Acute Physiology and Chronic Health Evaluation (*APACHE*) II, and APACHE III. The VA medical center and the community hospital differed significantly for survival adjusted by Charlson comorbidity index (*CCI*). The university hospital and the community hospital differed significantly for unadjusted survival and for survival adjusted by the SA-30 model, the sickness-at-admission (180-day survival) (*SA-180*) model, APACHE II, admission mortality prediction model (*MPM_0*), and CCI. For all comparisons, $P < 0.05$. CHFM-TIPI = Congestive Heart Failure Mortality Time-Insensitive Predictive Instrument; CHQC = Cleveland Health Quality Choice.

To assess disagreement among the models' survival predictions for individual patients, we ordered patients by quintiles according to their scores from each model and then tabulated the number of patients for which each possible pair of models disagreed by at least one quintile.

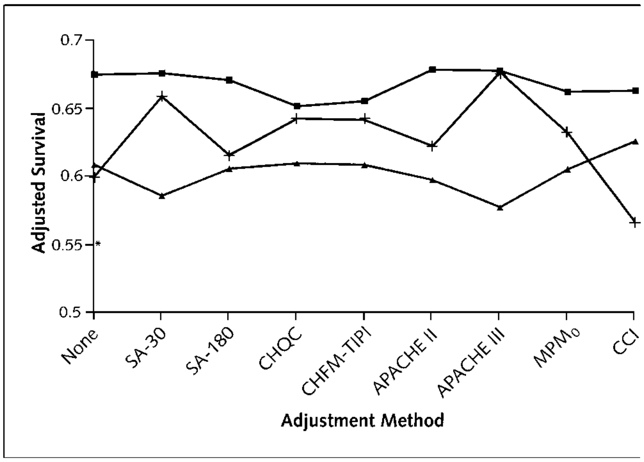
To assess the effect of adjustment method on simulated outcome report cards, we again used logistic regression analyses. We built models of the form logit(survival) = $b_0 + b_1 UH + b_2 CH + b_3 VA + b_4$ score, where *UH*, *CH*, and *VA* were indicator variables for the university hospital, the community hospital, and the VA medical center, respectively, and *score* was the score of the particular survival prediction model calculated for each patient. We built sets of logistic regression models for 30-day and 1-year survival, adjusting for each of the eight models. To determine the statistical significance of differences between survival rates of pairs of hospitals after adjustment, we tested whether their regression coefficients were equal. We then calculated each hospital's survival rate after adjustment according to each prediction model by using the marginal prediction method: We calculated the adjusted survival rate for the hospital as if its patient population had the same distribution of predicted survival probabilities as the entire study sample (49).

Figure 2. **Results of simulated report cards for 30-day survival after exclusion of patients who were not hospitalized.**



Lines with squares represent the university hospital, lines with triangles represent the Veterans Affairs (VA) medical center, and lines with crosses represent the community hospital. The university hospital and the VA medical center differed significantly for survival adjusted by the sickness-at-admission (30-day survival) (*SA-30*) model; the sickness-at-admission (180-day survival) (*SA-180*) model; the Cleveland Health Quality Choice (*CHQC*) model; the Congestive Heart Failure Mortality Time-Insensitive Predictive Instrument (*CHFM-TIPI*); Acute Physiology and Chronic Health Evaluation (*APACHE*) II; APACHE III; admission mortality prediction model (MPM_0); and Charlson comorbidity index (*CCI*). The VA medical center and the community hospital did not differ significantly. The university hospital and the community hospital differed significantly for unadjusted survival and for survival adjusted by the SA-30 model, the SA-180 model, APACHE II, MPM, and CCI. For all comparisons, $P < 0.05$.

Figure 3. **Results of simulated report cards for 1-year survival for the entire cohort.**



Lines with squares represent the university hospital, lines with triangles represent the Veterans Affairs (VA) medical center, and lines with crosses represent the community hospital. The university hospital and the VA medical center differed significantly for unadjusted survival and for survival adjusted by the sickness-at-admission (30-day survival) (*SA-30*) model; the sickness-at-admission (180-day survival) (*SA-180*) model; the Cleveland Health Quality Choice (*CHQC*) model; the Congestive Heart Failure Mortality Time-Insensitive Predictive Instrument (*CHFM-TIPI*); Acute Physiology and Chronic Health Evaluation (*APACHE*) II; APACHE III; and admission mortality prediction model (MPM_0). The VA medical center and the community hospital differed significantly for survival adjusted by APACHE III. The university hospital and the community hospital differed significantly for survival adjusted by the SA-180 model, APACHE II, and CCI. For all comparisons, $P < 0.05$.

Results

We examined 1603 emergency department visits of 1173 patients. We excluded 94 visits of patients who also presented with myocardial infarction, 20 visits without vital status data, and 153 visits for which no charts were available, leaving 1336 visits eligible for our main analyses.

The overall survival rate was 1200 of 1336 (89.8%) at 30 days and 842 of 1336 (63.0%) at 1 year. The 30-day survival rate at the community hospital and the 1-year survival rate at the VA medical center were significantly lower than the corresponding rates at the university hospital (**Table 1**). The patients' clinical characteristics were typical of patients with heart failure and varied across the hospitals (**Table 2**).

None of the prediction models had very good discriminating ability, as measured by the areas under the ROC curves. For two random patients, one who survived and one who did not, the area under the ROC curve is equivalent to the probability that the patient who survived had a better model score (that is, a greater predicted probability of survival). The CHQC model had the best discriminating ability for 30-day survival (ROC curve area, 0.73), followed by the SA-30 model and MPM_0 (ROC curve area, 0.67 for both). The models' performance at predicting 1-year survival was similar, as was their performance at predicting survival only for hospitalized patients. However, after recalibration by logistic regression, we found that no model was badly calibrated ($P > 0.05$, data not shown).

The models often disagreed in how they ordered individual patients by probability of survival. **Table 3** shows the proportions of patients that particular pairs of models placed in different quintiles of survival probability. For each possible pairing of models, more than half of the patients were ranked differently.

We determined the effects of adjusting simulated report cards by using different survival prediction models for 30-day survival (**Figure 1**), 30-day survival after exclusion of patients who were not hospitalized (**Figure 2**), and 1-year survival (**Figure 3**). Each of the figures displays the unadjusted survival rate for the three hospitals and their survival rates adjusted by each of the eight prediction models. These figures show that whether any pair of hospitals had a significantly different survival rate depended on the methods used for the comparison: the choice of the model used for adjustment, the time frame for survival, and whether we excluded patients who were not hospitalized.

Some simulated report cards, including those for 30-day survival adjusted by the CHQC and CHFM-TIPI models, showed no significant differences in survival among the three hospitals.

Simulated report cards for 30-day survival adjusted by APACHE III; those for 30-day survival of only hospitalized patients adjusted by the CHQC model, CHFM-TIPI, and APACHE III; and those for 1-year survival adjusted by the SA-30 model, the CHQC model, CHFM-TIPI, and MPM_0 showed that only the VA medical center had significantly worse survival than the university hospital. Simulated report cards for 30-day survival, unadjusted and adjusted by the SA-180 model and MPM_0; those for unadjusted 30-day survival of only hospitalized patients; and those for 1-year survival adjusted by CCI showed that only the community hospital had significantly worse survival than the university hospital. Simulated report cards for 30-day survival adjusted by the SA-30 model and APACHE II; those for 30-day survival of only hospitalized patients adjusted by the SA-30 model, the SA-180 model, APACHE II, and MPM_0; and those for 1-year survival adjusted by the SA-180 model, APACHE II, and APACHE III showed that the VA medical center and the community hospital had significantly worse survival than the university hospital. One simulated report card, for 30-day survival adjusted by CCI, showed that the community hospital had significantly worse survival than both the university hospital and the VA medical center.

Discussion

Using eight survival prediction models to adjust for initial severity of illness, we evaluated mortality for patients with heart failure who were seen at three hospitals. No analysis showed that patients at the VA medical center or the community hospital had better survival rates than those at the university hospital. Other studies of outcome report cards have suggested that patients with congestive heart failure at teaching hospitals have survival rates better than (29) or at least the same as (30) survival rates among patients at other hospitals.

However, whether patients at the VA medical center or community hospital seemed to have survival rates different from or worse than survival rates at the university hospital depended on the survival prediction model used to adjust for severity. Our findings corroborate reports by Iezzoni and colleagues, which demonstrated that results of mortality outcome report cards for patients with several diseases may vary depending on the generic survival prediction model used to adjust for underlying disease severity (47, 50–53). They also support a study by Fine and colleagues (54), which showed that results of mortality outcome report cards for patients with pneumonia varied depending on whether a generic or disease-specific model was used, and a study by Orr and colleagues (55), which showed that results of mortality report cards for patients undergoing coronary artery bypass grafting varied depending on which of four disease- and procedure-specific models was used. Our study, however, assessed outcomes of patients with heart failure by adjusting for severity with both disease-specific and generic prediction models.

Our results raise several questions. Why did different prediction models produce dissimilar results? The models ranked many individual patients differently, perhaps because of differences in the variables included in the models. There was some overlap, but the models also differed sub-

Appendix Table. **Variables Used in Models***

Variable Type	Model			
	SA-30	**SA-180**	**CHQC**	**CHFM-TIPI**
Demographic	Age	Male Nursing home resident	Age Nursing home resident	Age
History and comorbid conditions		Previous hospitalization Difficulty with limbs Noncompliance Ventricular depressants Antiarrhythmic medications	Cancer COPD CVA or TIA Cirrhosis Steroid use	
Physical examination	Blood pressure Coma score	Mean blood pressure†	Cardiac arrest Mechanical ventilation Neurologic assessment Heart rate Respiratory rate Systolic blood pressure†	Systolic blood pressure†
Diagnostic test	CHF on chest radiography BUN level† Sodium level† Hematocrit† Digitalis level‡ AST level†	BUN level†	Infiltrate on chest radiograph BUN level† Creatinine concentration† BUN/creatinine† Sodium level† LDH level† AST level† Leukocyte count† Albumin level† Bicarbonate level‡ pH‡ P_{CO_2}‡ P_{O_2}	Flat T waves on ECG Normal sinus rhythm on ECG
Other predictive models	APACHE II APS	SA-30 APACHE II APS		

* APACHE = Acute Physiology and Chronic Health Evaluation; APS = acute physiology score; AST = aspartate aminotransferase; BUN = blood urea nitrogen; CCI = Charlson comorbidity index; CHF = congestive heart failure; CHFM-TIPI = Congestive Heart Failure Mortality Time-Insensitive Predictive Instrument; CHQC = Cleveland Health Quality Choice; COPD = chronic obstructive pulmonary disease; CVA = cerebrovascular accident; ECG = electrocardiogram; LDH = lactate dehydrogenase; MPM_0 = mortality prediction model; NYHA = New York Heart Association; TIA = transient ischemic attack; SA-30 = sickness at admission (30-day survival); SA-180 = sickness at admission (180-day survival).
† Treated as a categorical variable.
‡ Treated as a dichotomous variable.

stantially in the specific variables that were used and in the weighting of common data items.

We found that no model was clearly superior to the others in the ability to discriminate among patients with different actual rates of survival. In fact, none of the models predicted mortality in our patient population as well as they did in their previously published applications.

Unavoidably, several features of our study design that were meant to reduce bias may have also reduced the ability of the models to predict survival. We required the models to predict 30-day and 1-year survival; however, most were designed to predict in-hospital survival, which may be biased by discharge decision making (56, 57). The longer a patient stays in the hospital, the more likely it is that he or she will die there (19). Nonetheless, the SA-30 and CHQC models were specifically designed to predict 30-day mortality.

We computed model scores by using only data available in the emergency department, even though most models were designed to also use data from the first days after hospitalization. Data available later in the hospitalization may be more predictive of survival because they may reflect management during hospitalization and may be more thorough. In addition, the assumption that missing data are

Appendix Table—Continued

Model				
APACHE II	**APACHE III**	**MPM_0**		**CCI**
Age†	Age†	Age†		
Liver insufficiency	Cirrhosis	Cirrhosis	Myocardial infarction	Hemiplegia
NYHA class IV cardiac disease	Hepatic failure	Chronic renal failure	Peripheral vascular disease	Moderate to severe renal disease
Severe chronic respiratory disease	Lymphoma	Metastatic cancer	Cerebrovascular disease	Diabetes with end-organ damage
Chronic renal failure on dialysis	Leukemia or multiple myeloma	Acute renal failure	Dementia	Tumor
Immunocompromise	Acute renal failure		COPD	Leukemia
	Immunosuppression		Connective tissue disease	Lymphoma
	AIDS		Ulcer disease	Moderate to severe liver disease
			Mild liver disease	Metastatic tumor
			Diabetes	
Mean arterial pressure†	Mean blood pressure†	Cardiopulmonary arrest		
Heart rate†	Heart rate†	Mechanical ventilation		
Respiratory rate†	Respiratory rate†	Heart rate‡		
Temperature†	Temperature†	Systolic blood pressure‡		
Glasgow coma score†	Neurologic abnormality†			
	Urine output			
Creatinine concentration†	BUN level†	Dysrhythmia		
Sodium level†	Creatinine concentration†			
Potassium level‡	Sodium level†			
Hematocrit†	Hematocrit†			
Leukocyte count†	Leukocyte count†			
Po_2†	pH†			
pH†	Pco_2†			
	Po_2†			
	Arterial–alveolar oxygen gradient†			
	Albumin level†			
	Bilirubin level†			
	Glucose level†			

likely to be normal may be more accurate for data obtained later. Data obtained after hospitalization, however, may be biased by the way the patient is managed in the hospital (58, 59). Even interventions made within the first hours of hospitalization may affect outcomes (60). Nonetheless, the CHFM-TIPI, MPM_0, and CCI were designed to use only data available on presentation to the emergency department.

Our main analyses included all patients regardless of whether they were admitted. All of the models were developed with data from patients who were hospitalized or in intensive care. Excluding patients who were not admitted, however, risks bias caused by the admitting decision. For example, failing to admit the sickest patients could improve a hospital's apparent survival rate. Covinsky and colleagues (61) also found that the admission acute physiology score from APACHE II was a mediocre predictor of in-hospital and 1-year survival (ROC areas, 0.67 and 0.58) for elderly hospitalized patients. Furthermore, excluding patients who were not hospitalized did not improve the performance of any of the models.

Our study had some specific limitations. Our assessments of generic models for patients with heart failure may not apply to the ways in which these generic models perform for patients with other clinical problems. Our study was done in three hospitals in one small metropolitan area. There is no good reason to suspect, however, that our results would not apply to other hospitals and in other geographic areas. Our study data were not new. They were newer than data used in outcome studies that adjusted for severity of illness by using some of the eight models (26, 37, 38) and not much older than the data used in other such studies (29, 32, 39). Furthermore, we know of no new effective interventions for the management of acute congestive heart failure introduced since our data collection that might reduce its generalizability to today's patients.

In conclusion, the results of survival report cards for patients with acute congestive heart failure may depend on the choice of methods used to adjust for risk. Relying on survival outcome report cards that use one of the currently available risk adjustment methods may lead to erroneous conclusions about the quality of care at particular hospitals and may unfairly penalize individual hospitals in today's competitive market. Further research to develop improved risk prediction models for the purpose of adjusting outcome report cards for patients' initial severity of illness is clearly warranted. Models incorporating more data about the disease of interest (for example, in patients with heart failure, information about cause, physiologic characteristics, functional status [62], and biopsychosocial issues) may make more discriminating predictions and give more consistent results when used to adjust outcome report cards. We should also continue to develop other methods of quality assessment, such as process report cards. Process report cards may be valid in clinical situations in which strong evidence favors certain processes over others. Such evidence-based quality assessment (63–66) is beginning to receive more support (67).

From Brown University and Memorial Hospital of Rhode Island, Pawtucket, Rhode Island; Virginia Commonwealth University, Richmond, Virginia; Northwestern University Medical School, Chicago, Illinois; Northern Virginia Community College, Manassas, Virginia; Emergency Medical Services, Austin, Texas; and Michigan State University, East Lansing, Michigan.

Grant Support: By a grant from the Agency for Healthcare Research and Quality (HS-06274) and a Robert Wood Johnson Foundation Generalist Physician Faculty Scholarship (Dr. Smith).

Requests for Single Reprints: Roy M. Poses, MD, Brown University Center for Primary Care and Prevention, Memorial Hospital of Rhode Island, 111 Brewster Street, Pawtucket, RI 02860; e-mail, royposes@brownvm.brown.edu.

Requests To Purchase Bulk Reprints (minimum, 100 copies): Barbara Hudson, Reprints Coordinator; phone, 215-351-2657; e-mail, bhudson@mail.acponline.org.

Current Author Addresses: Dr. Poses: Brown University Center for Primary Care and Prevention, Memorial Hospital of Rhode Island, 111 Brewster Street, Pawtucket, RI 02860.
Dr. McClish: Department of Biostatistics, Virginia Commonwealth University, MCV Campus, Box 980032, Richmond, VA 23298-0032.
Dr. Smith: Division of Quality Health Care, Department of Internal Medicine, Virginia Commonwealth University, MCV Campus, Box 980306, Richmond, VA 23298-0306.
Drs. Huber and Clemo: Division of General Internal Medicine, Department of Internal Medicine, Virginia Commonwealth University, MCV Campus, Box 980102, Richmond, VA 23298-0102.
Dr. Schmitt: Chicago Veterans Affairs Medical Center, 333 East Huron Street, Chicago, IL 60611.
Dr. Alexander: Department of Psychology, John Tyler Community College, 13101 Jefferson Davis Highway, Chester, VA 23831-5316.
Dr. Racht: City of Austin/Travis County Emergency Medical Services, 517 South Pleasant Valley, Austin, TX 78741-1902.
Dr. Colenda: Department of Psychiatry, College of Human Medicine, A-222 East Fee Hall, Michigan State University, East Lansing, MI 48824-1316.

Author Contributions: Conception and design: R.M. Poses, D.K. McClish, W.R. Smith, E.C. Huber, F.L.W. Clemo, B.P. Schmitt, E.M. Racht, C.C. Colenda.
Analysis and interpretation of the data: R.M. Poses, D.K. McClish, W.R. Smith, E.C. Huber, F.L.W. Clemo, B.P. Schmitt.
Drafting of the article: R.M. Poses, D.K. McClish, W.R. Smith, E.C. Huber, F.L.W. Clemo.
Critical revision of the article for important intellectual content: R.M. Poses, D.K. McClish, W.R. Smith, E.C. Huber, F.L.W. Clemo, B.P. Schmitt, D. Alexander, C.C. Colenda.
Final approval of the article: R.M. Poses, D.K. McClish, W.R. Smith, E.C. Huber, F.L.W. Clemo, B.P. Schmitt.
Provision of study materials or patients: R.M. Poses, E.C. Huber, F.L.W. Clemo, E.M. Racht.
Statistical expertise: R.M. Poses, D.K. McClish.
Obtaining of funding: R.M. Poses.
Administrative, technical, or logistic support: R.M. Poses, F.L.W. Clemo, B.P. Schmitt, A. Alexander.
Collection and assembly of data: R.M. Poses, D.K. McClish, W.R. Smith, E.C. Huber, F.L.W. Clemo, B.P. Schmitt, D. Alexander, E.M. Racht.

References

1. **Trillo RA.** Clinical outcomes: the new driving force in health care. American Medical News. 1997;40:17.
2. **Kassirer JP.** The use and abuse of practice profiles [Editorial]. N Engl J Med. 1994;330:634-5.

3. Joint Commission on Accreditation of Health Care Organizations. Oryx Outcomes: The Next Evolution in Accreditation. Oakbrook, IL: Department of Publication, Joint Commission on Accreditation of Health Care Organizations; 1997.
4. **Prager LO.** Upping the certification ante. American Medical News. 1999; 42:1, 31, 32.
5. **Jollis JG, Romano PS.** Pennsylvania's Focus on Heart Attack—grading the scorecard [Editorial]. N Engl J Med. 1998;338:983-7.
6. **Marquis J.** Heart attack report dings two Orange County centers. Los Angeles Times. 15 December 1997:B1.
7. **Landry S.** Hospital report cards get mixed results. St. Petersburg Times. 26 February 1996:1B.
8. **Fields J.** Surviving bypass surgery: doctors' and hospitals' death rates compared. The Record (Bergen County, NJ). 6 March 1999:A01.
9. Connecticut: hospital report card causes controversy. American Health Line. 9 August 1996.
10. **Papp L, Boyle T.** Heart attack hospitals rated. Toronto Star. 24 February 1999:A8.
11. Public Health Development Unit, NHS. Quality and Performance in the NHS: Clinical Indicators. London: National Health Service; 1999.
12. **Pallarito K.** N.Y. hospital's specialty lies at heart of facility's fiscal, operational health. Mod Healthc. 1995;25:35-6.
13. **Mazzolini J.** Hospital report finds fewer care differences. Cleveland Plain Dealer. 3 December 1998:4B.
14. Lancaster County hospitals earn strong report cards. Intelligencer Journal (Lancaster, PA). 29 June 1999:B1.
15. **Selker HP.** Systems for comparing actual and predicted mortality rates: characteristics to promote cooperation in improving health care [Editorial]. Ann Intern Med. 1993;118:820-2.
16. **Dubois RW, Rogers WH, Moxley JH 3d, Draper D, Brook RH.** Hospital inpatient mortality. Is it a predictor of quality? N Engl J Med. 1987;317:1674-80.
17. **Epstein A.** Performance reports on quality—prototypes, problems, and prospects [Editorial]. N Engl J Med. 1995;333:57-61.
18. **Iezzoni LI.** Assessing quality using administrative data. Ann Intern Med. 1997;127(8 Pt 2):666-74.
19. **Kahn KL, Brook RH, Draper D, Keeler EB, Rubenstein LV, Rogers WH, et al.** Interpreting hospital mortality data. How can we proceed? [Editorial] JAMA. 1988;260:3625-8.
20. **Iezzoni LI.** The risks of risk adjustment. JAMA. 1997;278:1600-7.
21. **McNeil BJ, Pedersen SH, Gatsonis C.** Current issues in profiling quality of care. Inquiry. 1992;29:298-307.
22. **Orchard C.** Comparing healthcare outcomes. BMJ. 1994;308:1493-6.
23. **Iezzoni LI.** Risk adjustment for medical effectiveness research: an overview of conceptual and methodological considerations. J Investig Med. 1995;43:136-50.
24. **Daley J, Jencks S, Draper D, Lenhart G, Thomas N, Walker J.** Predicting hospital-associated mortality for Medicare patients. A method for patients with stroke, pneumonia, acute myocardial infarction, and congestive heart failure. JAMA. 1988;260:3617-24.
25. **Keeler EB, Kahn KL, Draper D, Sherwood MJ, Rubenstein LV, Reinisch EJ, et al.** Changes in sickness at admission following the introduction of the prospective payment system. JAMA. 1990;264:1962-8.
26. **Selker HP, Griffith JL, D'Agostino RB.** A time-insensitive predictive instrument for acute hospital mortality due to congestive heart failure: development, testing, and use for comparing hospitals: a multicenter study. Med Care. 1994;32:1040-52.
27. **Schocken DD, Arrieta MI, Leaverton PE, Ross EA.** Prevalence and mortality rate of congestive heart failure in the United States. J Am Coll Cardiol. 1992;20:301-6.
28. **Ho KK, Pinsky JL, Kannel WB, Levy D.** The epidemiology of heart failure: the Framingham Study. J Am Coll Cardiol. 1993;22(4 Suppl A):6A-13A.
29. **Rosenthal GE, Harper DL, Quinn LM, Cooper GS.** Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. Results of a regional study. JAMA. 1997;278:485-90.
30. **Taylor DH Jr, Whellan DJ, Sloan FA.** Effects of admission to a teaching hospital on the cost and quality of care for Medicare beneficiaries. N Engl J Med. 1999;340:293-9.
31. **Poses RM, Smith WR, McClish DK, Huber EC, Clemo FL, Schmitt BP, et al.** Physicians' survival predictions for patients with acute congestive heart failure. Arch Intern Med. 1997;157:1001-7.
32. **Rosenthal GE, Baker DW, Norris DG, Way LE, Harper DL, Snow RJ.** Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. Health Serv Res. 2000;34:1449-68.
33. **Knaus W, Wagner D, Draper E.** Chapter 2. Development of APACHE. Crit Care Med. 1989;17:S181-5.
34. **Knaus WA, Wagner D, Draper E, Zimmerman JE, Bergner M, Bastos PG, et al.** The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest. 1991;100:1619-36.
35. **Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J.** Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. JAMA. 1993;270:2478-86.
36. **Charlson ME, Pompei P, Ales KL, MacKenzie CR.** A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40:373-83.
37. **Keeler EB, Rubenstein LV, Kahn KL, Draper D, Harrison ER, McGinty MJ, et al.** Hospital characteristics and quality of care. JAMA. 1992;268:1709-14.
38. **Knaus WA, Draper EA, Wagner DP, Zimmerman JE.** An evaluation of outcome from intensive care in major medical centers. Ann Intern Med. 1986;104:410-8.
39. **Sirio CA, Shepardson LB, Rotondi AJ, Cooper GS, Angus DC, Harper DL, et al.** Community-wide assessment of intensive care outcomes using a physiologically based prognostic measure: implications for critical care delivery from Cleveland Health Quality Choice. Chest. 1999;115:793-801.
40. **Pappachan JV, Millar B, Bennett ED, Smith GB.** Comparison of outcome from intensive care admission after adjustment for case mix by the APACHE III prognostic system. Chest. 1999;115:802-10.
41. **Park RE, Brook RH, Kosecoff J, Keesey J, Rubenstein L, Keeler E, et al.** Explaining variations in hospital death rates. Randomness, severity of illness, quality of care. JAMA. 1990;264:484-90.
42. **Wong DT, Crofts SL, Gomez M, McGuire GP, Byrick RJ.** Evaluation of the predictive ability of the APACHE II system and hospital outcome in Canadian intensive care unit patients. Crit Care Med. 1995;23:1177-83.
43. **Knaus WA, Draper EA, Wagner DP, Zimmerman JE.** APACHE II: a severity of disease classification system. Crit Care Med. 1985;13:818-29.
44. **Krousel-Wood MA, Abdoh A, Re R.** Comparing comorbid-illness

indices assessing outcome variation: the case of prostatectomy. J Gen Intern Med. 1996;11:32-8.
45. **Poses RM, McClish DK, Smith WR, Bekes C, Scott WE.** Prediction of survival of critically ill patients by admission comorbidity. J Clin Epidemiol. 1996;49:743-7.
46. **Braitman LE, Davidoff F.** Predicting clinical states in individual patients. Ann Intern Med. 1996;125:406-12.
47. **Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD.** Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. Ann Intern Med. 1995;123:763-70.
48. **Poses RM, Cebul RD, Collins M, Fager SS.** The importance of disease prevalence in transporting clinical prediction rules. The case of streptococcal pharyngitis. Ann Intern Med. 1986;105:586-91.
49. **Wilcosky TC, Chambless LE.** A comparison of direct adjustment and regression adjustment of epidemiologic measures. J Chronic Dis. 1985;38:849-56.
50. **Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD.** Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. Am J Public Health. 1996;86:1379-87.
51. **Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD.** Predicting in-hospital mortality for stroke patients: results differ across severity-measurement methods. Med Decis Making. 1996;16:348-56.
52. **Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD.** Severity measurement methods and judging hospital death rates for pneumonia. Med Care. 1996;34:11-28.
53. **Iezzoni LI, Ash AS, Shwartz M, Landon BE, Mackiernan YD.** Predicting in-hospital deaths from coronary artery bypass graft surgery. Do different severity measures give different predictions? Med Care. 1998;36:28-39.
54. **Fine MJ, Hanusa BH, Lave JR, Singer DE, Stone RA, Weissfeld LA, et al.** Comparison of a disease-specific and generic severity of illness measure for patients with community-acquired pneumonia. J Gen Intern Med. 1995;10:359-68.
55. **Orr RK, Maini BS, Sottile FD, Dumas EM, O'Mara P.** A comparison of four severity-adjusted models to predict mortality after coronary artery bypass graft surgery. Arch Surg. 1995;130:301-6.
56. **Green J, Wintfeld N.** Report cards on cardiac surgeons. Assessing New York State's approach. N Engl J Med. 1995;332:1229-32.
57. Dicing with death rates [Editorial]. Lancet. 1993;341:1183-4.
58. **Hughes JS, Iezzoni LI, Daley J, Greenberg L.** How severity measures rate hospitalized patients. J Gen Intern Med. 1996;11:303-11.
59. **Hannan EL, Racz MJ, Jollis JG, Peterson ED.** Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? Health Serv Res. 1997;31:659-78.
60. **Boyd O, Grounds M.** Can standardized mortality ratio be used to compare quality of intensive care unit performance? Crit Care Med. 1994;22:1706-9.
61. **Covinsky KE, Justice AC, Rosenthal GE, Palmer RM, Landefeld CS.** Measuring prognosis and case mix in hospitalized elders. The importance of functional status. J Gen Intern Med. 1997;12:203-8.
62. **Davis RB, Iezzoni LI, Phillips RS, Reiley P, Coffman GA, Safran C.** Predicting in-hospital mortality. The importance of functional status information. Med Care. 1995;33:906-21.
63. Evidence-based care: 1. Setting priorities: how important is the problem? Evidence-Based Care Resource Group. CMAJ. 1994;150:1249-54.
64. Evidence-based care: 2. Setting guidelines: how should we manage this problem? CMAJ. 1994;150:1417-23.
65. Evidence-based care: 3. Measuring performance: how are we managing this problem? Evidence-Based Care Resource Group. CMAJ. 1994;150:1575-9.
66. Evidence-based care: 4. Improving performance: how can we improve the way we manage the problem? Evidence-Based Care Resource Group. CMAJ. 1994;150:1793-6.
67. **McColl A, Roderick P, Gabbay J, Smith H, Moore M.** Performance indicators for primary care groups: an evidence based approach. BMJ. 1998;317:1354-60.