
Advancing Methods

Development and Validation of a Disease-Specific Risk Adjustment System Using Automated Clinical Data

Ying P. Tabak, Xiaowu Sun, Karen G. Derby, Stephen G. Kurtz, and Richard S. Johannes

Objective. To develop and validate a disease-specific automated inpatient mortality risk adjustment system primarily using computerized numerical laboratory data and supplementing them with administrative data. To assess the values of additional manually abstracted data.

Methods. Using 1,271,663 discharges in 2000–2001, we derived 39 disease-specific automated clinical models with demographics, laboratory findings on admission, ICD-9 principal diagnosis subgroups, and secondary diagnosis-based chronic conditions. We then added manually abstracted clinical data to the automated clinical models (manual clinical models). We compared model discrimination, calibration, and relative contribution of each group of variables. We validated these 39 models using 1,178,561 discharges in 2004–2005.

Results. The overall mortality was 4.6 percent ($n = 58,300$) and 4.0 percent ($n = 47,279$) for derivation and validation cohorts, respectively. Common mortality predictors included age, albumin, blood urea nitrogen or creatinine, arterial pH, white blood counts, glucose, sodium, hemoglobin, and metastatic cancer. The average c -statistic for the automated clinical models was 0.83. Adding manually abstracted variables increased the average c -statistic to 0.85 with better calibration. Laboratory results displayed the highest relative contribution in predicting mortality.

Conclusions. A small number of numerical laboratory results and administrative data provided excellent risk adjustment for inpatient mortality for a wide range of clinical conditions.

Key Words. Automated clinical data, laboratory data, predicting mortality, risk adjustment, performance reporting, comparative effectiveness research

Comparison of health care outcomes is of interest to both the clinical community and public (Halm and Chassin 2001; Fonarow and Peterson 2009; VanLare, Conway, and Sox 2010). New funding for comparative research

from the American Recovery and Reinvestment Act of 2009 (U.S. Congress), coupled with the health care reform, has generated renewed interest as well as concern about methods of comparative effectiveness research and performance reporting (Fonarow and Peterson 2009; Gibbons et al. 2009). When comparing health care outcomes in large populations, clinically credible risk adjustment methodology that can be implemented on a large scale at low cost is important. Although clinical trials are the standard method of assessing health care effectiveness, they have high data collection costs, tend to be conducted on relatively small and homogeneous patient populations, and are not practical for all types of research. As a complement, observational studies enable large-scale investigations of outcomes, which may be more applicable to real world settings (VanLare, Conway, and Sox 2010). The observational studies have been further advanced by the development and proliferating use of technology that enables electronic capture of clinical data. A 2008 survey on representative U.S. hospitals found that 77 percent had fully implemented and an additional 14 percent had been partially or were in the process of implementing electronic laboratory reports (Jha et al. 2009).

Recent publications demonstrated that automated laboratory data offer clinical credibility, objectivity, parsimony, and cost-effectiveness for risk adjustment (Jordan et al. 2007; Tabak, Johannes, and Silber 2007; Escobar et al. 2008; Render et al. 2008). Laboratory data were found to contribute most in predicting mortality among demographics, comorbidities, and other groups of variables (Tabak, Johannes, and Silber 2007; Escobar et al. 2008; Render et al. 2008). However, existing studies either did not assess contribution of additional clinical data, such as vital signs, in predicting mortality (Escobar et al. 2008; Render et al. 2008) or limited patient population to primarily male and ICU patients (Render et al. 2008). Tabak, Johannes, and Silber (2007) developed and validated disease-specific mortality predictive models, evaluating both cumulative and relative contributions of laboratory data in relation to demographics, administrative, and other manually collected clinical data. Their analysis, nevertheless, was limited to only six common clinical conditions. In a large patient population using disease-specific modeling, we sought

Address correspondence to Ying P. Tabak, Ph.D., Director, Biostatistics, Clinical Research, MedMined Services, CareFusion, 400 Nickerson Road, Marlborough, MA 01752; e-mail: ying.tabak@carefusion.com. Xiaowu Sun, Ph.D., Karen G. Derby, B.A., and Richard S. Johannes, M.D., M.S., are with Clinical Research, MedMined Services, CareFusion, Marlborough, MA. Richard S. Johannes, M.D., M.S., is also with the Division of Gastroenterology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA. Stephen G. Kurtz, M.S., is with MASSPRO, Waltham, MA.

to extend the previous work to a broad array of clinical conditions by addressing whether promising laboratory results observed in a few common clinical conditions are reproducible for other less frequently studied conditions. We further evaluated the value of manually extracted vital signs and mental status data in model predictive ability in relation to electronically captured laboratory results, demographics, and diagnosis-based administrative data. Because health care data are complex, prioritizing the electronic capture and utilization of the most standardized data elements for population-based research seems prudent. In addition to numerical laboratory results, vital signs are also objective and quantitative. Hence, determining the value of vital signs in risk adjustment may inform policy makers regarding the relative importance and priority of electronic data capture, storage, and transmission, given the federal government's commitment to invest billions of dollars in the coming years to encourage the widespread adoption of health information technology in the United States (Blumenthal 2010).

METHODS

Data

We used one of the Clinical Research Databases from CareFusion (Formerly Cardinal Health Clinical Research Database [Clinical Research Services, Marlborough, MA]). This database has been used for research since the late 1980s and the data collection system has been fully described elsewhere (Iezzoni and Moskowitz 1988; Silber et al. 1995; Fine et al. 1997; Kollef et al. 2005; Aujesky et al. 2006; Shorr et al. 2006, 2009; Pine et al. 2007; Tabak, Johannes, and Silber 2007; Hollenbeak et al. 2008; Tabak et al. 2009; Weigelt, Lipsky, and Tabak 2010). The current study population consisted of 1,271,663 discharges in 2000–2001 from 217 hospitals for the derivation cohort and 1,178,561 discharges in 2004–2005 from 191 hospitals for the validation cohort.

The study database included imported hospital administrative data that was comprised of demographics, principal diagnosis, and up to 25 secondary diagnosis codes. The database also contained electronically imported or manually abstracted laboratory data, vital signs, and other clinical findings. The derivation and validation cohorts had similar laboratory data completion rates. A total of 96 percent of patients had laboratory data on the day of admission. For the 2 percent of patients who did not have laboratory data recorded on admission day, data collection extended to 30 hours after admission. For surgical patients, laboratory data were eligible before surgery

starting time if surgery was within the admission window. If surgery was later than the admission data collection window, data collected within the admission window was used. About 2 percent of cases were recorded as missing laboratory data for the specified data collection window. For patients with multiple laboratory assessments on admission day, the worst value was collected.

For this study, we selected 39 major disease groups based on volume of admissions and associated inpatient mortality rate. These disease groups covered clinical conditions of all major organ systems, including the nervous, circulatory, digestive, hepatobiliary/pancreatic, musculoskeletal, metabolic, and kidney/urinary systems, as well as infectious diseases. Patients were classified into one of these mutually exclusive disease groups based on their principal diagnosis. Each patient had only one principal diagnosis for a given admission.

Model Development and Validation

Model Development. We first developed 39 automated clinical models—one for each disease group—using demographics, numerical laboratory findings on admission, principal diagnosis subgroups based on ICD-9 codes, and chronic conditions based on secondary diagnoses. For each disease group, we examined the distribution of each continuous variable in relation to in-hospital death. We partitioned each continuous variable into multiple discrete levels. A category for patients with missing laboratory data was created and the mortality of this group was compared and pooled into a reference group (Pine et al. 2007; Tabak, Johannes, and Silber 2007). This approach allowed us to use data on all the patients and is more practical for large-scale implementation than imputation or dropping patients with missing data. All candidate variables that were statistically associated with mortality ($p < .05$) were included as potential covariates. Variable selection in multivariable regression models was based on clinical plausibility and statistical significance.

We added manually abstracted clinical data to the automated clinical models for the 39 manual clinical models. We only considered vital signs (systolic blood pressure, diastolic blood pressure, respiration, heart beat, and temperature) and altered mental status, which was assessed by the Glasgow Coma Scale or a designation of disoriented, stupor, or coma as charted by the attending physicians. We did not include other manually collected clinical variables beyond vital signs and mental status, because previous studies have

found that the contribution of these variables to model discrimination is negligible (Tabak, Johannes, and Silber 2007; Hollenbeak et al. 2008).

We compared changes in c -statistics when vital sign and mental status variables were added to the models. Because the c -statistic may be insensitive in distinguishing between models on calibration and the traditional Hosmer–Lemeshow χ^2 test is not suitable when the sample size is very large, we evaluated the change of model calibration using joint distributions of predicted mortality risk by the two sets of models (Cook 2007). This method allowed us to evaluate whether models with manually extracted data would more accurately stratify individuals into higher or lower mortality risk strata compared with models without these data.

Model Validation. We validated each model internally using bootstrapping in the derivation cohort by sampling with replacement for 200 iterations (Efron and Tibshirani 1993). Variables that never changed coefficient signs and were significant in more than 70 percent of reiterations were retained in the model. For external validation, we recalibrated all models using 1,178,561 cases discharged in 2004–2005 because of the significant decrease in in-hospital mortality observed across years.

Relative Contributions of Variables. We examined changes in the model-fit log-likelihood value when each group of variables was retained and removed from the full model (Escobar et al. 2008; Render et al. 2008). We calculated the relative contributions of age, laboratory results, ICD-9 code-based variables, and additional manually abstracted variables for each model.

Comparison of Hospital Performance Using Automated versus Manual Clinical Models

Large-scale implementation of a clinical risk adjustment system requires cost efficiency. Because electronic capture of vital signs and mental status may require more comprehensive implementation of electronic medical records, which is currently less available compared with electronically captured numerical laboratory data (Jha et al. 2009), we evaluated whether models without vital signs and mental status (automated clinical models) can serve as surrogates for models with these additional data that currently requires manual extraction for the majority of hospitals. Specifically, we fit two sets of hierarchical models to compare hospital performance (Normand et al. 1997;

Tabak, Johannes, and Silber 2007). First, we obtained hospital ranking for each disease group using risk-standardized mortality rates generated from automated clinical models. Second, we obtained another set of ranking using manual clinical models. We used the Spearman rank correlation coefficient to assess the agreement. A high-level agreement between the two sets of results would suggest that the automated clinical models can be used as surrogates for the manual clinical models.

RESULTS

Patient Baseline Characteristics by Derivation versus Validation Cohorts

Overall, the median (interquartile range) age was 72 (57, 81) versus 71 (56, 81) years for the derivation versus validation cohorts, respectively. Approximately 45.4 percent of both cohorts were men. A total of 50.3 versus 58.3 percent of cases were from teaching hospitals and 15.6 versus 14.7 percent were from rural hospitals, respectively, for the two cohorts. The overall mortality was 4.6 percent ($n = 58,300$) for the derivation cohort and 4.0 percent ($n = 47,279$) for the validation cohort. Table 1 displays the distribution of patients and mortality by disease group for the derivation versus validation cohorts.

Mortality Predictors

The most common mortality predictors across disease groups included age, albumin, BUN or creatinine, arterial pH, white blood cell (WBC) counts, blood glucose, sodium, hemoglobin, and other abnormal metabolic, or hematologic parameters (Table 2). The most common chronic conditions predicting mortality included metastatic cancer or cancer of major organ systems. The overall results were similar in the recalibrated validation cohort.

Model Discrimination, Calibration, and Relative Contribution of Predictors

The average c -statistic for the automated models was 0.83 for the derivation cohort (Table 3). The addition of vital signs and mental status increased the average c -statistic to 0.85. It also improved model calibration when predicted mortality risk strata were evaluated in the joint distributions (Table 4). Models with vital signs and mental status reclassified 17.3 percent of cases into risk strata that were more accurate representations of observed mortality risks. For example, 57,483 cases in the 1–5 percent mortality risk stratum were reclassified into <1 percent mortality risk stratum, which was a more accurate

Table 1: Study Population by Disease Groups and Derivation and Validation Cohorts

Major Disease Category	Derivation Cohort		Validation Cohort	
	Discharges, n (%) [*]	Mortality, n (%) [†]	Discharges, n (%)	Mortality, n (%)
Disease Group				
Total	1,271,663 (100.0)	58,300 (4.6)	1,178,561 (100.0)	47,279 (4.0)
Nervous				
Ischemic stroke	44,102 (3.5)	2,929 (6.6)	36,870 (3.1)	2,093 (5.7)
Hemorrhagic stroke	13,860 (1.1)	3,128 (22.6)	12,694 (1.1)	2,909 (22.9)
Respiratory				
Asthma	23,011 (1.8)	102 (0.4)	28,066 (2.4)	125 (0.4)
Chronic obstructive pulmonary disease	64,438 (5.1)	1,970 (3.1)	57,791 (4.9)	1,369 (2.4)
Tuberculosis	495 (0.0)	24 (4.8)	470 (0.0)	36 (7.7)
Pneumonia	94,411 (7.4)	6,114 (6.5)	89,959 (7.6)	4,413 (4.9)
Aspiration pneumonia	18,289 (1.4)	3,354 (18.3)	17,805 (1.5)	2,503 (14.1)
Respiratory failure	23,087 (1.8)	4,520 (19.6)	24,883 (2.1)	4,702 (18.9)
Pulmonary embolism	9,680 (0.8)	534 (5.5)	11,613 (1.0)	452 (3.9)
Circulatory				
Arrhythmia	83,445 (6.6)	1,126 (1.3)	78,464 (6.7)	870 (1.1)
Arterial aneurysm	12,408 (1.0)	817 (6.6)	9,960 (0.8)	603 (6.1)
Venous disorder	20,162 (1.6)	226 (1.1)	17,851 (1.5)	174 (1.0)
Valvular heart disease	5,115 (0.4)	240 (4.7)	3,796 (0.3)	177 (4.7)
Acute myocardial infarction	84,581 (6.7)	7,034 (8.3)	64,363 (5.5)	4,552 (7.1)
Chest pain	171,526 (13.5)	754 (0.4)	136,826 (11.6)	410 (0.3)
Heart failure	131,748 (10.4)	6,008 (4.6)	120,745 (10.2)	4,377 (3.6)
Catastrophic vascular disease	3,727 (0.3)	1,146 (30.7)	3,056 (0.3)	761 (24.9)
Catastrophic cardiovascular disease	2,059 (0.2)	1,098 (53.3)	1,216 (0.1)	492 (40.5)
Other circulatory disease	53,933 (4.2)	649 (1.2)	49,738 (4.2)	366 (0.7)
Digestive				
Esophageal disease	13,959 (1.1)	78 (0.6)	12,610 (1.1)	39 (0.3)

continued

Table 1. Continued

Major Disease Category	Derivation Cohort		Validation Cohort	
	Discharges, n (%) [*]	Mortality, n (%) [†]	Discharges, n (%)	Mortality, n (%)
Gastrointestinal disorder	11,722 (0.9)	90 (0.8)	10,285 (0.9)	52 (0.5)
Intestinal disease	52,212 (4.1)	674 (1.3)	50,849 (4.3)	558 (1.1)
Gastroenterologic bleeding	47,372 (3.7)	1,561 (3.3)	44,431 (3.8)	1,199 (2.7)
Gastroenterologic obstruction	23,649 (1.9)	926 (3.9)	21,439 (1.8)	745 (3.5)
Gastroenterologic perforation	2,321 (0.2)	461 (19.9)	1,971 (0.2)	322 (16.3)
Infectious diarrhea	10,686 (0.8)	224 (2.1)	16,473 (1.4)	369 (2.2)
Colorectal cancer	6,374 (0.5)	498 (7.8)	4,376 (0.4)	284 (6.5)
Other gastroenterologic disease	24,162 (1.9)	231 (1.0)	24,901 (2.1)	169 (0.7)
Hepatobiliary/pancreatic				
Gallbladder disease	17,288 (1.4)	225 (1.3)	14,196 (1.2)	145 (1.0)
Acute liver disease	3,739 (0.3)	339 (9.1)	4,365 (0.4)	380 (8.7)
Liver/bacillary disease	10,946 (0.9)	917 (8.4)	10,901 (0.9)	730 (6.7)
Pancreas disease	20,379 (1.6)	364 (1.8)	21,096 (1.8)	251 (1.2)
Musculoskeletal				
Hip/upper femur fracture	35,534 (2.8)	955 (2.7)	27,498 (2.3)	670 (2.4)
Metabolic				
Diabetes	34,459 (2.7)	535 (1.6)	34,837 (3.0)	342 (1.0)
Kidney/urinary				
Chronic renal disease	11,150 (0.9)	653 (5.9)	9,033 (0.8)	468 (5.2)
Acute renal failure	13,940 (1.1)	1,407 (10.1)	24,218 (2.1)	1,682 (6.9)
Genital/urinary infection	35,881 (2.8)	666 (1.9)	38,790 (3.3)	507 (1.3)
Infection				
Sepsis	30,950 (2.4)	5,279 (17.1)	36,389 (3.1)	6,704 (18.4)
High-risk infection	4,663 (0.4)	444 (9.5)	3,737 (0.3)	279 (7.5)

*Discharges presented as n (column %).

†Mortality presented as n (row %). The overall standardized mortality rate for the validation cohort was 3.9%.

Table 2: Summary of Odds Ratios for Common Mortality Predictors

Variables	No. of Models Present	Derivation Cohort			Validation Cohort		
		Median	1st Quartile	3rd Quartile	Median	1st Quartile	3rd Quartile
Demographic							
Age	38	1.044	1.036	1.053	1.039	1.031	1.049
Male	7	1.28	1.10	1.48	1.15	1.09	1.47
Laboratory results							
Blood urea nitrogen	34	1.91	1.58	2.52	1.77	1.55	2.42
White blood cell count	33	1.56	1.41	1.76	1.47	1.31	1.87
Bands	31	1.50	1.28	1.77	1.59	1.45	1.98
Albumin	30	1.81	1.59	2.14	1.84	1.60	2.20
pH arterial	30	2.29	1.90	2.79	2.22	1.81	3.08
Sodium	29	1.41	1.15	1.67	1.47	1.23	1.57
PT INR/PT second	28	1.34	1.28	1.54	1.40	1.27	1.75
Glucose	27	1.40	1.21	1.72	1.31	1.19	1.52
Platelets	27	1.50	1.32	1.92	1.39	1.32	1.62
Troponin I/CK MB	27	1.69	1.43	1.93	1.73	1.52	1.97
Aspartate aminotransferase	26	1.53	1.36	1.90	1.49	1.34	2.05
Potassium	25	1.31	1.23	1.45	1.25	1.20	1.68
Base units	24	1.89	1.23	2.63	1.85	1.41	2.18
pCO ₂ arterial	23	1.65	1.43	2.29	1.60	1.38	2.22
PO ₂ /O ₂ saturation	23	1.57	1.35	1.91	1.49	1.25	2.03
Total bilirubin	21	1.60	1.47	2.11	1.85	1.39	2.48
Calcium	18	1.25	1.13	1.48	1.26	1.19	1.59
Creatinine	17	1.58	1.39	1.93	1.65	1.28	1.79
Hemoglobin	15	1.30	1.26	1.88	1.55	1.30	2.03
Creatine phosphokinase	13	1.37	1.21	1.60	1.42	1.23	1.73

continued

Table 2. Continued

Variables	No. of Models Present	Derivation Cohort			Validation Cohort		
		Median	1st Quartile	3rd Quartile	Median	1st Quartile	3rd Quartile
ICD-9 principal diagnosis subgroup	27	3.56	2.01	4.54	3.30	1.81	5.10
ICD-9 principal DX subgroups							
ICD-9 secondary diagnosis code-based comorbidity	29	2.64	2.32	3.47	2.73	2.14	3.73
Metastatic cancer	20	2.32	1.96	2.67	2.25	1.84	2.83
Solid organ cancer	14	2.15	1.86	2.39	2.03	1.64	2.65
Hematological cancer	16	2.20	1.87	3.12	2.16	1.70	2.89
Hypertensive renal disease with renal failure	16	1.33	1.27	1.49	1.36	1.23	1.64
Chronic obstructive pulmonary disease	15	1.77	1.36	1.95	1.64	1.54	1.91
Cardiomyopathy	13	1.86	1.54	2.58	2.68	2.12	3.11
Chronic renal failure	11	1.60	1.42	2.77	2.14	1.45	3.09
Chronic liver disease	10	2.08	1.64	2.47	1.90	1.51	2.48
Dialysis status	10	2.31	2.02	2.50	1.67	1.60	1.90
Cachexia	7	1.96	1.31	2.13	1.72	1.57	2.45
Chronic pulmonary heart disease	6	1.66	1.46	2.55	1.56	1.24	1.88
Pulmonary fibrosis/chronic pulmonary disease	6	1.56	1.43	2.21	1.62	1.52	1.75
Chronic ischemic heart disease	6	0.76	0.74	0.83	0.72	0.68	0.75
Hypertension not malignant							
Vital signs and mental status							
Systolic/diastolic blood pressure	34	1.86	1.59	2.05	1.65	1.50	2.05
Pulse	30	1.52	1.38	1.82	1.55	1.34	1.75
Respirations	30	1.92	1.57	2.16	1.46	1.28	1.91
Oral temperature	27	1.53	1.31	1.82	1.47	1.35	1.60
Altered mental status	36	3.17	2.40	4.54	2.95	2.24	5.27

CK MB, MB isoenzymes of creatine kinase; PT, prothrombin time; PTT, partial thromboplastin time; PT INR, prothrombin time/international normalized ratio.

Table 3: Average C-statistics for Derivation and Validation Cohorts by Major Disease Category

Major Disease Category	No. of Models	Derivation Cohort			Validation Cohort		
		Automated Model (Age+LAB+ICD)	Manual Model (Age+LAB+ICD+VS+AMS)	Automated Model (Age+LAB+ICD)	Manual Model (Age+LAB+ICD+VS+AMS)	Automated Model (Age+LAB+ICD)	Manual Model (Age+LAB+ICD+VS+AMS)
Nervous	2	0.80	0.87	0.80	0.86	0.80	0.86
Respiratory	7	0.79	0.82	0.78	0.80	0.78	0.80
Circulatory	10	0.84	0.86	0.81	0.83	0.81	0.83
Digestive	9	0.85	0.86	0.82	0.84	0.82	0.84
Hepatobiliary/pancreatic	4	0.85	0.87	0.84	0.86	0.84	0.86
Musculoskeletal	1	0.79	0.80	0.76	0.78	0.76	0.78
Metabolic	1	0.87	0.89	0.87	0.88	0.87	0.88
Kidney/urinary	3	0.80	0.82	0.79	0.82	0.79	0.82
Infection	2	0.80	0.83	0.79	0.81	0.79	0.81
All	39	0.83	0.85	0.81	0.83	0.81	0.83

AMS, altered mental status; ICD, ICD-9 principal diagnosis subgroup and ICD-9 secondary diagnosis-based comorbidity; LAB, laboratory results; VS, vital signs.

Table 4: Comparison of Observed and Predicted Mortality Risk*

	Predicted Risk Stratum by Models without VS and AMS					Predicted Risk Stratum by Models with VS and AMS				
	< 1%	1 to < 5%	5 to < 10%	10 to < 20%	20%+					
<i>n</i>	531,085	21,997	147	18	1					
Row% [†]	96.0	4.0	0.0	0.0	0.0					
Observed mortality (%) [‡]	0.2%	1.7%	11.6%	16.7%	100.0%					
1 to < 5%										
<i>n</i>	57,483	317,170	21,750	3,554	510					
Row%	14.4	79.2	5.4	0.9	0.1					
Observed mortality (%)	0.7%	2.1%	7.3%	15.8%	29.6%					
5 to < 10%										
<i>n</i>		38,815	53,013	13,717	2,676					
Row%		35.9	49.0	12.7	2.5					
Observed mortality (%)		4.1%	7.1%	14.0%	29.7%					
10 to < 20%										
<i>n</i>		1,860	22,015	32,329	11,052					
Row%		2.8	32.7	48.1	16.4					
Observed mortality (%)		5.2%	8.2%	14.5%	29.7%					
20%+										
<i>n</i>		31	979	10,263	38,096					
Row%		0.1	2.0	20.8	77.2					
Observed mortality (%)		12.9%	11.5%	18.1%	43.2%					
Total	588,568	379,873	97,904	59,881	52,335					

*This comparison uses aggregated predicted inpatient mortality generated from models with and without vital signs and altered mental status. All predicted and observed risks represent risk of in-hospital mortality. This table presents aggregated validation cohort ($n = 1,178,561$).

[†]Percent classified in each risk stratum by the models with vital signs and altered mental status.

[‡]Observed proportion of patients deceased in the hospital in each risk stratum.

AMS, altered mental status; VS, vital signs.

representation of observed mortality of 0.7 percent for these patients. It should be noticed that 96.7 percent of reclassified cases shifted only to the immediately adjacent stratum.

Overall, the laboratory variables contributed most in predicting mortality with an average relative contribution of 43.2 percent across all 39 models in the derivation cohort (Table 5). The next highest contributor in predicting mortality was age (17.4 percent). The ICD-9 code base comorbidities, vital signs, and altered mental status each contributed about 10 percent and ICD-9 principal diagnosis group contributed about 7 percent. The results for the validation cohort were very similar.

Hospital Performance Ranking

The hospital performance ranks generated by automated clinical models were highly correlated with those generated by manual clinical models. The average Spearman rank correlation coefficients on hospital performance ranking was 0.97 for both derivation and validation cohorts.

DISCUSSION

A clinically credible and low-cost risk adjustment system is important for comparative outcome studies and performance reporting. Numerical laboratory data are objective, precise, and parsimonious when used for risk adjustment. The finding that the same small set of numerical laboratory results can serve as the basis for excellent predictions of inpatient mortality for a large, diverse set of clinical conditions further opens the way to collect these data on all hospitalized patients for whom the tests are clinically indicated.

Why are the numerical laboratory results important in predicting mortality? Biomarkers such as serum chemistry, blood cell counts, blood gas, and other metabolic and hematologic parameters provide objective assessments of organ system function. They minimize variations in assessments of clinical conditions of patients and eliminate over- or under-coding issues that are of concern for variables based on diagnosis codes. Secondly, the numerical laboratory data have a “dose–response” relationship with mortality outcome; the farther the laboratory result deviates from the reference level, the higher the risk of death. This graded quantitative effect leads to more accurate differentiation of organ system dysfunction than dichotomous variables captured in diagnosis codes. Third, a set of two dozen numerical laboratory tests encompasses assessment of major organ system functions that are needed to keep

Table 5: Average Relative Contribution of Variable Group by Major Disease Category

Major Disease Category	Derivation Cohort						Validation Cohort					
	Age	Lab	ICD-PDX	ICD-COM	VS	AMS	Age	Lab	ICD-PDX	ICD-COM	VS	AMS
Nervous	8.5	18.3	5.9	3.7	7.5	56.1	8.4	25.0	1.9	2.2	6.2	56.3
Respiratory	18.0	48.8	4.8	8.0	13.4	7.0	24.2	45.6	3.4	8.5	10.5	7.8
Circulatory	14.4	41.4	10.0	13.5	10.9	9.9	15.7	34.0	14.3	13.2	11.8	11.0
Digestive	20.7	44.3	7.1	13.7	8.1	6.1	20.8	40.1	7.8	17.8	7.8	5.8
Hepatobiliary/pancreatic	15.2	52.2	10.0	6.3	8.2	8.1	17.3	60.0	6.0	6.1	3.3	7.3
Musculoskeletal	21.5	37.3	0.0	25.4	8.3	7.5	27.1	24.9	0.0	31.8	12.1	4.2
Metabolic	19.1	35.9	8.4	16.1	6.9	13.7	23.3	35.2	17.7	8.8	8.8	6.2
Kidney/urinary	23.0	41.9	0.9	9.1	14.5	10.6	23.6	39.8	1.9	8.3	12.7	13.8
Infection	18.8	42.4	5.4	9.5	13.8	10.2	19.4	46.5	3.2	9.9	12.2	8.9
Average of all 39 models	17.4	43.2	6.9	11.1	10.5	10.8	19.5	40.6	7.5	11.9	9.5	11.0

AMS, altered mental status; ICD-COM, ICD-9 secondary diagnosis-based comorbidity; ICD-PDX, ICD-9 principal diagnosis subgroup; LAB, laboratory results; VS, vital signs.

patients alive, making the use of laboratory data for risk adjustment parsimonious and efficient from both scientific and economic perspectives. From a clinical perspective, some deranged laboratory findings might not have one-on-one corresponding code to capture the complete spectrum of clinical complexity seen in laboratory results. For example, an abnormally low albumin could indicate chronic malnutrition, liver failure, renal dysfunction, secondary manifestation of cardiac dysfunction, or even acute severe sepsis possibly due to capillary leakage of albumin. Identification and classification of diagnosis codes to cover the broad spectrum of clinical conditions might be more arduous than directly using the laboratory test results themselves.

Our study built on previous studies on automated laboratory data (Tabak, Johannes, and Silber 2007; Escobar et al. 2008; Render et al. 2008) by extending previous research to both male and female patients admitted for a broad range of diseases in a diverse group of acute care hospitals in terms of teaching status, bed size, and rural location. We used a large database consisting of administrative, numerical laboratory, and manually collected clinical data. We found that laboratory data contributed most in predicting mortality even when we included manually collected key clinical findings of vital signs and mental status. Although the absolute value of the relative contribution of the laboratory data was slightly smaller in our study compared with previous studies that did not include additional manually collected clinical data as covariates (Escobar et al. 2008; Render et al. 2008), our finding is consistent with previous publication that included additional variables in the models and used a different statistical method to calculate the relative contribution of laboratory data (Tabak, Johannes, and Silber 2007). These findings further validated the stability and consistency of objective and numerical laboratory data when applied to different patient populations across a wide array of disease groups and over time.

We found that vital signs and altered mental status added an average 0.02 in c -statistic above and beyond automated models. The small cumulative c -statistic increase was in line with a previous study on eight clinical conditions (Pine et al. 2009). However, adding vital signs and mental status improved model calibration in joint distribution analysis, which was not investigated previously. The incremental improvement in calibration might be particularly meaningful if risk stratification is of the interest (Cook 2007). Furthermore, these physiologic variables have clinical face validity as mortality risk factors. With the increasing use of full electronic medical records, automated collection, storage, and transmission of voluminous vital signs will likely become more practical. Hence, our findings may have policy implications for setting

the next priority of inclusion of electronic clinical data for health services research.

The finding that altered mental status contributed more than laboratory data in predicting mortality among patients suffering from neurologic disorders such as ischemic and hemorrhagic stroke patients is clinically plausible. From a clinical perspective, laboratory results do not necessarily capture neurologic function. Further study on using “present on admission” (POA) for ICD-9 diagnosis codes indicating “coma” may shed light on a practical way to electronically capture and utilize the information of altered mental status on admission for risk adjustment, especially for diseases of the neurological system. It should be noted that current coding conventions may preclude coding of signs and symptoms that are “integral part of” or “associated routinely with a disease process” (CMS 2009). Our finding on the importance of altered mental status in risk adjustment may aid responsible parties to discuss and consider clarifying and modifying rules so that clinically important signs and symptoms, such as “coma,” can be consistently coded across hospitals.

Adoption of a full electronic medical record system that enables inter-hospital collection, storage, and transmission of vital signs and mental status throughout the United States will likely take time. Hence, a system using data that is already captured electronically across a vast majority of hospitals would have practical value. Our analysis showed that hospital performance ranks generated by automated clinical models (numerical laboratory data and administrative data) were highly correlated with those generated by models with additional vital signs and mental status. As a bridge, hybrid models incorporating the most widely automated numerical laboratory results and information from administrative data may serve as a reasonable intermediate step for aggregated performance reporting.

Our study has limitations. It may be debatable on how to best group a heterogeneous patient population into clinically homogeneous subgroups. Currently, there were multiple clinical grouping systems (Pine et al. 2007; Tabak, Johannes, and Silber 2007; Escobar et al. 2008; Elixhauser, Steiner, and Palmer 2010). The Clinical Classifications Software (*CCS*) recently updated by the Agency for Healthcare Research and Quality consists of 285 diagnosis groups (Elixhauser et al. 2010). Although the *CCS* system offers granularity of grouping patients into homogeneous diagnosis-related groups, it would require even larger database than we currently have to insure adequate number of cases and outcome events for model development and validation, especially for those low-volume disease groups. Implementing a more granulated disease grouping system, such as the *CCS*, for clinical risk adjustment

modeling may be achievable in the future if a nation-wide automated clinical database is established for health services research. Our study provided further evidence in support of establishing such a national database to advance health services research.

The methodology surrounding the use of numerical laboratory data in risk adjustment modeling also varies. Our disease-specific modeling approach encompassed three phases. (1) It was based on review of disease-specific risk adjustment tools for a general inpatient population published by the clinical community, which showed differences in variable selection and weight of the same variable in different risk adjustment models for patients hospitalized for different clinical conditions (Goldman et al. 1996; Fine et al. 1997; Fonarow et al. 2005; Aujesky et al. 2006; Tabak, Johannes, and Silber 2007; Tabak et al. 2009). (2) Our empirical review of the distribution of each variable in relation to the outcome by disease group showed significant differences across disease groups. For example, for patients with WBC counts ($10^9/L$) of ≤ 4.3 , 4.4–10.9, 11.0–14.1, 14.2–19.8, or ≥ 19.9 , the corresponding observed inpatient mortality was 7.8, 3.6, 4.5, 5.8, or 7.6 percent if pneumonia was the principal diagnosis whereas, for the same laboratory findings, the corresponding mortality for patients of chronic obstructive pulmonary disease (COPD) was 1.7, 1.7, 2.7, 4.0, and 5.5 percent. These data revealed that neutropenia (WBC $\leq 4.3 [10^9/L]$) was associated with the highest mortality risk for pneumonia patients, but not for COPD patients, for whom, the mortality was about the same, the lowest (1.7 percent), whether their WBC was below or in the normal range. (3) The feedback from our clinical advisory panels preferred disease-specific models for easy understanding of risk factors and their relative weights pertinent to caregivers' specialties. Our finding that risk factors and their relative weights vary depending upon the clinical conditions being considered supports this viewpoint.

The disease-specific modeling approach differs from generic modeling approaches used by other researchers. These approaches include APACHE IV (Zimmerman et al. 2006) and the Kaiser Permanent risk adjustment systems (Escobar et al. 2008), for which a generic physiological score using numerical variables was devised first and then an aggregated physiology score was reentered into the multivariable model with other variables, including disease groups. The generic method has merits. It requires only a reasonably sized database for model development and validation and it might be easier to dissimilate and implement. In contrast, development and validation of a disease-specific risk adjustment system requires a very large database and the application of such a system may necessitate the incorporation of more

complex electronic systems. Although a direct comparison of these two modeling approaches from statistical perspectives might be interesting, it is beyond the scope of the current study. Perhaps more pertinent to health services research is the fact that both modeling approaches yielded convergent results on the importance of numerical laboratory and vital sign data in risk adjustment, which provide compelling evidence for policy makers in setting priority of health care information technology in capturing and utilizing these numerical data.

Our comorbidity variables using secondary diagnoses did not reflect the recent coding change of identification of acute clinical conditions POA. Future studies may further examine directly consistency, reliability, and validity of POA coding in the administrative data as well as the relative contribution of these new data in relation to electronically captured numerical laboratory and vital sign data when they all become widely available. When evaluating the value of these data, it is important to balance objectivity, parsimony, and cost, in addition to statistical performance.

CONCLUSIONS

A small number of laboratory findings provide objective, quantitative, and parsimonious measures of the risk of inpatient mortality in a large array of clinical conditions. Clinical models using electronic numerical laboratory and administrative data can be used for population-based comparative outcome studies and hospital performance reporting. Vital signs and mental status should be included in the automated risk adjustment systems when the electronic collection, storage, and transmission of these data become widely available. Based on automated data, these models are cost-efficient to implement as a risk adjustment system.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: An abstract based on preliminary results of this manuscript was selected as one of the “most outstanding” abstracts and was presented at the Academy Health 25th Annual Research Meeting on June 9, 2008, Washington, DC. The slides presentation was posted at the Academy Health website.

All authors were previous employees at Cardinal Health. Y. P. T., X. S., K. G. D., and R. S. J. reported current employment at CareFusion. S. G. K. reported current employment at Massachusetts Peer Review Organization. R. S. J. also reported employment at the Division of Gastroenterology, Brigham and Women's Hospital and Harvard Medical School.

We would like to thank Linda Hyde at CareFusion for her technical support. We acknowledge many helpful and constructive comments from the two anonymous reviewers.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Aujesky, D., D. S. Obrosky, R. A. Stone, T. E. Auble, A. Perrier, J. Cornuz, P. M. Roy, and M. J. Fine. 2006. "A Prediction Rule to Identify Low-Risk Patients with Pulmonary Embolism." *Archives of Internal Medicine* 166 (2): 169–75.
- Blumenthal, D. 2010. "Launching HITECH." *New England Journal of Medicine* 362 (5): 382–5.
- CMS. 2009. "ICD-9-CM Official Guidelines for Coding and Reporting" [accessed on April 9, 2010]. Available at <http://www.cdc.gov/nchs/data/icd9/icdguide09.pdf>
- Cook, N. R. 2007. "Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction." *Circulation* 115 (7): 928–35.
- Efron, B., and R. Tibshirani. 1993. *An Introduction to the Bootstrap*. London: Chapman & Hall.
- Elixhauser, A., C. Steiner, and L. Palmer. 2010. "Clinical Classifications Software (CCS), 2010. U.S. Agency for Healthcare Research and Quality" [accessed on January 19, 2010]. Available at <http://www.hcup-us.ahrq.gov/toolsoftware/ccs/CCSUsersGuide.pdf>
- Escobar, G. J., J. D. Greene, P. Scheirer, M. N. Gardner, D. Draper, and P. Kipnis. 2008. "Risk-Adjusting Hospital Inpatient Mortality Using Automated Inpatient, Outpatient, and Laboratory Databases." *Medical Care* 46 (3): 232–9.
- Fine, M. J., T. E. Auble, D. M. Yealy, B. H. Hanusa, L. A. Weissfeld, D. E. Singer, C. M. Coley, T. J. Marrie, and W. N. Kapoor. 1997. "A Prediction Rule to Identify Low-Risk Patients with Community-Acquired Pneumonia." *New England Journal of Medicine* 336 (4): 243–50.
- Fonarow, G. C., K. F. Adams Jr., W. T. Abraham, C. W. Yancy, and W. J. Boscardin. 2005. "Risk Stratification for In-Hospital Mortality in Acutely Decompensated Heart Failure: Classification and Regression Tree Analysis." *Journal of American Medical Association* 293 (5): 572–80.
- Fonarow, G. C., and E. D. Peterson. 2009. "Heart Failure Performance Measures and Outcomes: Real or Illusory Gains." *Journal of American Medical Association* 302 (7): 792–4.

- Gibbons, R. J., T. J. Gardner, J. L. Anderson, L. B. Goldstein, N. Meltzer, W. S. Weintraub, and C. W. Yancy. 2009. "The American Heart Association's Principles for Comparative Effectiveness Research: A Policy Statement from the American Heart Association." *Circulation* 119 (22): 2955–62.
- Goldman, L., E. F. Cook, P. A. Johnson, D. A. Brand, G. W. Rouan, and T. H. Lee. 1996. "Prediction of the Need for Intensive Care in Patients Who Come to the Emergency Departments with Acute Chest Pain." *New England Journal of Medicine* 334 (23): 1498–504.
- Halm, E. A., and M. R. Chassin. 2001. "Why Do Hospital Death Rates Vary?" *New England Journal of Medicine* 345 (9): 692–4.
- Hollenbeak, C. S., C. P. Gorton, Y. P. Tabak, J. L. Jones, A. Milstein, and R. S. Johannes. 2008. "Reductions in Mortality Associated with Intensive Public Reporting of Hospital Outcomes." *American Journal of Medical Quality* 23 (4): 279–86.
- Iezzoni, L. I., and M. A. Moskowitz. 1988. "A Clinical Assessment of MedisGroups." *Journal of American Medical Association* 260 (21): 3159–63.
- Jha, A. K., C. M. DesRoches, E. G. Campbell, K. Donelan, S. R. Rao, T. G. Ferris, A. Shields, S. Rosenbaum, and D. Blumenthal. 2009. "Use of Electronic Health Records in U.S. Hospitals." *New England Journal of Medicine* 360 (16): 1628–38.
- Jordan, H. S., M. Pine, A. Elixhauser, D. C. Hoaglin, D. Fry, K. Coleman, D. Deitz, D. Warner, J. Gonzales, and Z. Friedman. 2007. "Cost Effective Enhancement of Claim Data to Improve Comparisons of Patient Safety." *Journal of Patient Safety* 3: 82–90.
- Kollef, M. H., A. Shorr, Y. P. Tabak, V. Gupta, L. Z. Liu, and R. S. Johannes. 2005. "Epidemiology and Outcomes of Health-Care-Associated Pneumonia: Results from a Large US Database of Culture-Positive Pneumonia." *Chest* 128 (6): 3854–62.
- Normand, S.-L. T., M. E. Glickman, and C. A. Gatsonis. 1997. "Statistical Methods for Profiling Providers of Medical Care: Issues and Applications." *Journal of the American Statistical Association* 92: 803–14.
- Pine, M., H. S. Jordan, A. Elixhauser, D. E. Fry, D. C. Hoaglin, B. Jones, R. Meimban, D. Warner, and J. Gonzales. 2007. "Enhancement of Claims Data to Improve Risk Adjustment of Hospital Mortality." *Journal of American Medical Association* 297 (1): 71–6.
- . 2009. "Modifying ICD-9-CM Coding of Secondary Diagnoses to Improve Risk-Adjustment of Inpatient Mortality Rates." *Medical Decision Making* 29 (1): 69–81.
- Render, M. L., J. Deddens, R. Freyberg, P. Almenoff, A. F. Connors Jr., D. Wagner, and T. P. Hofer. 2008. "Veterans Affairs Intensive Care Unit Risk Adjustment Model: Validation, Updating, Recalibration." *Critical Care Medicine* 36 (4): 1031–42.
- Shorr, A. F., V. Gupta, X. Sun, R. S. Johannes, J. Spalding, and Y. P. Tabak. 2009. "Burden of Early-Onset Candidemia: Analysis of Culture-Positive Bloodstream Infections from a Large US Database." *Critical Care Medicine* 37 (9): 2519–26.

- Shorr, A. F., Y. P. Tabak, A. D. Killian, V. Gupta, L. Z. Liu, and M. H. Kollef. 2006. "Healthcare-Associated Bloodstream Infection: A Distinct Entity? Insights from a Large U.S. Database." *Critical Care Medicine* 34 (10): 2588–95.
- Silber, J. H., P. R. Rosenbaum, J. S. Schwartz, R. N. Ross, and S. V. Williams. 1995. "Evaluation of the Complication Rate as a Measure of Quality of Care in Coronary Artery Bypass Graft Surgery." *Journal of American Medical Association* 274 (4): 317–23.
- Tabak, Y. P., R. S. Johannes, and J. H. Silber. 2007. "Using Automated Clinical Data for Risk Adjustment: Development and Validation of Six Disease-Specific Mortality Predictive Models for Pay-for-Performance." *Medical Care* 45 (8): 789–805.
- Tabak, Y. P., X. Sun, R. S. Johannes, V. Gupta, and A. F. Shorr. 2009. "Mortality and Need for Mechanical Ventilation in Acute Exacerbations of Chronic Obstructive Pulmonary Disease: Development and Validation of a Simple Risk Score." *Archives of Internal Medicine* 169 (17): 1595–902.
- U.S. Congress. "Public Law 111–5. American Recovery and Reinvestment Act of 2009" [accessed on January 20, 2010]. Available at <http://www.gpo.gov/fdsys/pkg/PLAW-111publ5/content-detail.html>
- VanLare, J. M., P. H. Conway, and H. C. Sox. 2010. "Five Next Steps for a New National Program for Comparative-Effectiveness Research." *New England Journal of Medicine* 362 (11): 970–3.
- Weigelt, J. A., B. A. Lipsky, and Y. P. Tabak. 2010. "Surgical Site Infections: Causative Pathogens and Associated Outcomes among Hospitalized Patients, 2003 to 2007." *American Journal of Infection Control* 38 (2): 112–20.
- Zimmerman, J. E., A. A. Kramer, D. S. McNair, and F. M. Malila. 2006. "Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital Mortality Assessment for Today's Critically Ill Patients." *Critical Care Medicine* 34 (5): 1297–310.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.