Mortality Probability Model III and Simplified Acute Physiology Score II

Assessing Their Value in Predicting Length of Stay and Comparison to APACHE IV

Eduard E. Vasilevskis, MD; Michael W. Kuzniewicz, MD, MPH; Brian A. Cason, MD; Rondall K. Lane, MD, MPH; Mitzi L. Dean, MS, MHA; Ted Clay, MS; Deborah J. Rennie, BA; Eric Vittinghoff, PhD; and R. Adams Dudley, MD, MBA

Background: **To develop and compare ICU length-of-stay (LOS) risk-adjustment models using three commonly used mortality or LOS prediction models.**

Methods: **Between 2001 and 2004, we performed a retrospective, observational study of 11,295 ICU patients from 35 hospitals in the California Intensive Care Outcomes Project. We compared the accuracy of the following three LOS models: a recalibrated acute physiology and chronic health evaluation (APACHE) IV-LOS model; and models developed using risk factors in the mortality probability model III at zero hours (MPM_0) and the simplified acute physiology score (SAPS) II mortality prediction model. We evaluated models by calculating the following: (1) grouped coefficients of determination; (2) differences between observed and predicted LOS across subgroups; and (3) intraclass correlations of observed/expected LOS ratios between models.**

Results: **The grouped coefficients of determination were APACHE IV with coefficients recalibrated to the LOS values of the study cohort (APACHE IVrecal) [$R^2 = 0.422$], mortality probability model III at zero hours (MPM_0 III) [$R^2 = 0.279$], and simplified acute physiology score (SAPS II) [$R^2 = 0.008$]. For each decile of predicted ICU LOS, the mean predicted LOS vs the observed LOS was significantly different ($p \leq 0.05$) for three, two, and six deciles using APACHE IVrecal, MPM_0 III, and SAPS II, respectively. Plots of the predicted vs the observed LOS ratios of the hospitals revealed a threefold variation in LOS among hospitals with high model correlations.**

Conclusions: **APACHE IV and MPM_0 III were more accurate than SAPS II for the prediction of ICU LOS. APACHE IV is the most accurate and best calibrated model. Although it is less accurate, MPM_0 III may be a reasonable option if the data collection burden or the treatment effect bias is a consideration.** ***(CHEST 2009; 136:89–101)***

Abbreviations: APACHE = acute physiology and chronic health evaluation; APACHE IVorig = acute physiology and chronic health evaluation using coefficients described by the original publication of the acute physiology and chronic health evaluation IV length-of-stay model; APACHE IVrecal = acute physiology and chronic health evaluation IV with coefficients recalibrated to the length-of-stay values of the study cohort; CABG = coronary artery bypass graft; CALICO = California Intensive Care Outcomes; CI = confidence interval; DNR = do not resuscitate; LOS = length of stay; MPM_0 III = mortality probability model III at zero hours; SAPS = simplified acute physiology score; SLOSR = standardized length of stay ratio

The ICU provides advanced and resource-intensive treatment for the sickest hospitalized patients. Care in the ICU accounts for approximately 13% of hospital costs and 4.2% of national health expenditures.[1] These costs are largely explained by the length of stay (LOS) in the ICU.[2,3] There is significant variation in ICU LOS among hospitals that persists even after adjusting for patient risk factors.[4–6] This possibly reflects variations in ICU organization, safety, quality, or other hospital or community factors such as the availability of non-ICU beds.[7–10]

An important objective is to identify ICUs requiring longer or shorter LOSs after accounting for differences in patient characteristics. Comparing risk-adjusted ICU LOSs among ICUs may prove complementary to risk-adjusted mortality and process measures in assessing ICU performance.[11] The Joint Commission[12] and others[13] have expressed interest in public reporting of risk-adjusted ICU LOS.

The acute physiology and chronic health evaluation (APACHE [a registered trademark of Cerner Corporation; Kansas City, MO])[14,15] system is the only validated ICU risk-adjustment model that provides performance information about two separate outcomes of care (mortality and ICU LOS). The APACHE IV model is the most recent version. Two other validated ICU mortality prediction models, the mortality probability model III at zero hours (MPM_0 III) and the simplified acute physiology score (SAPS) II, use alternative risk-adjustment methods to assess mortality, although they have not been used for LOS prediction.[16,17] MPM_0 III and SAPS are important to consider for LOS risk adjustment because, as with APACHE, using the data collected for mortality prediction may provide an efficient means of assessing LOS. In addition, both models are used for the purposes of risk adjustment.[18,19] In contrast to APACHE, they have fewer risk factors and impose less of a data collection burden.[20]

We used data from $>11{,}000$ patients in the California Intensive Care Outcomes (CALICO) project to develop and compare the performance of APACHE IV, MPM_0 III, and SAPS II models in LOS prediction. In addition, we explored additional patient and hospital factors that may influence ICU LOS or hospital rankings.

Materials and Methods

Hospital Selection

All California hospitals were sent a recruiting packet. A network of volunteer hospitals was established through mailings and regional presentations.

Patient Selection

Data were collected between 2001 and 2004. Inclusion criteria were age ≥ 18 years and ICU stay ≥ 4 h. We excluded patients with conditions that were not examined across each risk-adjustment model, including burns, trauma, and coronary artery bypass graft (CABG) patients. In addition, we excluded patients who had been readmitted to the ICU, consistent with prior studies, and only abstracted data from the index ICU admission. We utilized a proportional sampling method where the goal sample size depended on the hospitals' annual number of ICU admissions.[20]

Risk Models and Variables

We used the MPM_0 III and SAPS II variables specified in their mortality model publications to create a LOS predictive model.[16,17] For the APACHE IV model, we used predictor variables detailed in the ICU LOS model publication.[15] Trained nurses from participating hospitals abstracted data for all models. ICU LOS, defined in hours and minutes, was the time at discharge from the ICU (either death or physical departure from the unit) minus the time of admission (first recorded vital sign on the ICU flow sheet). The LOS was calculated in days to the second significant digit and truncated at 30 days to minimize the impact of outliers, as previous investigators have done.[14,15] MPM_0 III

From the Philip R. Lee Institute for Health Policy Studies (Drs. Vasilevskis, Kuzniewicz, Lane, and Dudley, Ms. Dean, Mr. Clay, and Ms. Rennie), the Divisions of General Internal Medicine (Dr. Vasilevskis), Hospital Medicine (Dr. Vasilevskis), Neonatology (Dr. Kuzniewicz), and Pulmonary and Critical Care Medicine (Dr. Dudley), and the Departments of Anesthesiology and Perioperative Medicine (Drs. Cason and Lane), and Epidemiology and Biostatistics (Dr. Vittinghoff), University of California at San Francisco, San Francisco, CA; Veterans Affairs Medical Center (Dr. Cason), San Francisco, CA; the Department of Medicine (General Internal Medicine and Public Health) [Dr. Vasilevskis], Vanderbilt University, Nashville, TN; and Geriatric Research Education and Clinical Care (Dr. Vasilevskis), and the Clinical Research Training Center of Excellence (Dr. Vasilevskis), Department of Veterans Affairs, Tennessee Valley Healthcare System, Nashville, TN.

Dr. Vasilevskis had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Responsibility for areas of the study were as follows: study concept and design: Drs. Vasilevskis, Kuzniewicz, and Dudley; acquisition of data: Drs. Kuzniewicz, Cason, Lane, and Dudley, and Ms. Dean; analysis and interpretation of data: Drs. Vasilevskis, Kuzniewicz, Cason, Lane, Vittinghoff, and Dudley, Ms. Dean, Mr. Clay, and Ms. Rennie; drafting of the manuscript: Drs. Vasilevskis and Dudley; critical revision of the manuscript for important intellectual content: Drs. Vasilevskis, Kuzniewicz, Cason, Lane, Vittinghoff, and Dudley, Ms. Dean, Mr. Clay, and Ms. Rennie; statistical analysis: Drs. Vasilevskis and Vittinghoff, and Mr. Clay; obtained funding: Dr. Dudley; administrative, technical, or material support: Drs. Cason and Lane, Ms. Dean, and Ms. Rennie; and study supervision: Ms. Dean and Dr. Dudley.

The views expressed in this article are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs.

This work was supported by the California Office of Statewide Health Planning and Development and the Agency for Healthcare Research and Quality (R01 HS13919-01). Dr. Dudley's work was also supported by an Investigator Award in Health Policy from the Robert Wood Johnson Foundation. Dr. Vasilevskis was supported by a Ruth L. Kirschstein National Research Service Award institutional research training grant T32, the Veterans Affairs Clinical Research Center of Excellence, and the Geriatric Research Education and Clinical Center, Veterans Affairs, Tennessee Valley Healthcare, Nashville, TN.

The authors have reported to the ACCP that no significant conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Manuscript received October 30, 2008; revision accepted March 24, 2009.

Correspondence to: Eduard E. Vasilevskis, MD, Vanderbilt University Medical Center; 1215 Twenty-First Ave South, 6006 Medical Center East, NT, Nashville, TN 37232-8300; e-mail: eduard.vasilevskis@vanderbilt.edu

DOI: 10.1378/chest.08-2591

required collection of variables within 1 h of admission to the ICU. The other models used the most abnormal physiologic values in the first day after ICU admission. A list of diagnoses organized by system and condition was used to code the reason for ICU admission.[21] Data collection methods and interrater reliability have been previously described.[20]

Statistical Analysis

We compared CALICO hospital characteristics with all California hospitals that had > 50 hospital beds using the 2004 American Hospital Association survey.[22] Next, we divided data into development (60%) and validation (40%) samples, and used the χ^2 test, Student *t* test, and Mann-Whitney test, where appropriate, to compare characteristics of the samples.

Due to the hierarchical nature of the data (patients clustered within hospitals), we then used mixed-effects, multilevel modeling to generate ICU LOS prediction models for APACHE IV, MPM_0 III, and SAPS II using all variables in the original models. Due to known calibration limitations arising from using estimates of predictive performance on populations other than the one on which a risk model was developed,[23,24] we also reestimated the APACHE IV coefficients on the CALICO data set. This was necessary given the different time period, as well as reports of regional variations in health-care utilization patterns,[25,26] demographic mix,[27,28] and quality of care.[29,30] Our recalibration procedure maintained the original variable weights in the APACHE acute physiology score, as well as the spline knot values. The final models are APACHE IV models using coefficients described by the original publication of the APACHE IV LOS model (APACHE IVorig), APACHE IV with coefficients recalibrated to LOS values of the study cohort (APACHE IVrecal), MPM_0 III LOS model, and SAPS II LOS model.

Multiple methods were used to assess model performance in the validation sample. First, we used the paired Student *t* test to compare mean observed ICU LOS to mean predicted ICU LOS for the entire validation population and for specific subgroups (age groups, medical vs surgical patients, and patients grouped by primary clinical system deranged). Second, we divided the sample into deciles of predicted LOS and used the paired Student *t* test and calibration curves to compare mean predicted LOS to observed LOS for each model. Third, to measure the variance in LOS explained by the models, we calculated coefficients of determination (R^2) equal to the square of the correlation coefficient between the individual predicted LOS and the observed LOS. To assess the proportion of variation across hospitals explained by the models, we performed bivariate regressions of the mean observed LOS against the mean predicted LOS (grouped R^2) for hospitals with > 100 admissions, which was consistent with the intent of the developers of the original APACHE LOS model.[15]

Finally, we compared the assessments by the three models of the performance of the ICU of each hospital. The hospital LOS predictions were standardized by calculating a standardized LOS ratio (SLOSR) that was equal to the mean observed LOS divided by the mean predicted LOS for each hospital. Confidence intervals (CIs) were calculated by the Fieller method.[31] SLOSRs were limited to hospitals with > 100 admissions, which was consistent with prior studies.[15,32] We then assessed intraclass correlations between SLOSRs produced by the models.

Additional Risk Factors and Sensitivity Analyses

Due to the potential relationship of demographic and hospital factors with LOS, we developed additional models using data from the 2004 American Hospital Association survey and the California Office of Statewide Health Planning and Development. We adjusted for "do not resuscitate" (DNR) orders at hospital admission, payor status (Medicare, Medicaid, private, other), and hospital bed size.[33,34] We also used Spearman rank correlations to assess the relationship between demographic patient mix (*eg*, percentage of Medicaid patients) and hospital SLOSR performance assessed by the APACHE IVrecal.

Next, to determine whether hospital SLOSR was sensitive to hospital admission thresholds or the availability of step-down units,[35,36] we developed models after excluding patients with very short (< 24 h) LOSs. In addition, to assess the impact of case mix on performance, we assessed the Spearman correlation between the hospital mean severity of illness and the SLOSR.

Finally, we tested an additional SAPS II model treating each variable as an independent predictor, rather than a summed score, to evaluate for differences in model accuracy. The institutional review boards of the University of California, San Francisco, and the state of California approved the study. All analyses were performed using a statistical software package (STATA, version 9.2; Stata Corp; College Station, TX).

Results

Hospital Characteristics

The 35 participating hospitals included 57% not-for-profit institutions, 29% teaching hospitals, 9% hospitals with < 100 beds, 51% with 100 to 300 beds, and 41% with > 300 beds. Additional information on the CALICO hospitals has been previously published.[20]

Patient Characteristics

A total of 11,366 patients met our inclusion criteria. Of those, 71 patients (0.6%) had missing or indeterminate ICU LOS data, leaving a final data set of 11,295 patients. The overall mean and median LOSs were 4.0 and 2.0 days, respectively. The characteristics between the estimation and validation data sets were statistically similar across all characteristics (Table 1).

Predictive Performance of Four Models

The development sample (n = 6,684) was used to estimate coefficients for each model. Coefficients for MPM_0 III LOS and SAPS II LOS models are given in Table 2. Original coefficients for APACHE IV LOS are publicly available,[12] and reestimated coefficients are given in the Appendix.

Model performance was assessed in the 40% validation sample (n = 4,611). The difference between the mean observed LOS and the predicted ICU LOS for the validation sample was 4.6 h for APACHE IVorig ($p = 0.006$), 1.7 h for APACHE IVrecal ($p = 0.32$), 0.2 h for MPM_0 III LOS ($p = 0.90$), and 0.4 h for SAPS II LOS ($p = 0.82$). Observed LOS vs predicted LOS for strata of age,

Table 1—*Demographic and Clinical Characteristics*

Characteristics	Total Sample (n = 11,295)	Estimation Sample (n = 6,684)	Validation Sample (n = 4,611)	p Value*
Age,† yr	62.2 (17.4)	62.2 (17.6)	62.2 (17.3)	0.94
Age categories‡				0.33
18–44 yr	1,919 (17.0)	1,150 (17.2)	769 (16.7)	
45–64 yr	3,852 (34.1)	2,244 (33.6)	1,608 (34.9)	
65–84 yr	4,578 (40.5)	2,711 (40.6)	1,867 (40.5)	
> 85 yr	946 (8.4)	579 (8.7)	367 (8.0)	
Race‡				0.17
White	6,510 (57.6)	3,787 (56.7)	2,723 (59.1)	
Black	669 (5.9)	409 (6.1)	260 (5.6)	
Hispanic	1,960 (17.4)	1,193 (17.9)	767 (16.6)	
Asian/Pacific Islander	630 (5.6)	379 (5.7)	251 (5.4)	
Native American/other	319 (2.8)	184 (2.8)	135 (2.9)	
Unknown	1,207 (10.7)	732 (11.0)	475 (10.3)	
Expected payor‡				0.27
Medicare	5,021 (44.5)	2,989 (44.7)	2,032 (44.1)	
Medicaid	1,605 (14.2)	962 (14.4)	643 (13.9)	
Private coverage	2,597 (23.0)	1,490 (22.3)	1,107 (24.0)	
Other (*eg*, self-pay, workers' compensation, other government)	865 (7.7)	511 (7.7)	354 (7.7)	
Unknown	1,207 (10.7)	732 (11.0)	475 (10.3)	
DNR patients at admission‡	541 (4.8)	313 (4.7)	228 (4.9)	0.52
Operative status‡				0.65
Nonoperative	8,789 (77.8)	5,181 (77.5)	3,608 (78.3)	
Elective surgery	2,016 (17.9)	1,208 (18.1)	808 (17.5)	
Emergency surgery	490 (4.3)	295 (4.4)	195 (4.2)	
Severity of illness†				
APACHE score	44.9 (27.6)	44.7 (27.4)	45.2 (28.0)	0.31
SAPS II score	33.2 (17.6)	33.1 (17.5)	33.4 (17.7)	0.41
Location prior to ICU admission‡				0.51
Emergency department	5,548 (49.1)	3,270 (48.9)	2,278 (49.4)	
Operating room/recovery room	2,506 (22.2)	1,503 (22.5)	1,003 (21.8)	
Floor	2,426 (21.5)	1,421 (21.3)	1,005 (21.8)	
Transfer from another hospital	440 (3.9)	255 (3.8)	185 (4.0)	
Other	375 (3.3)	235 (3.5)	140 (3.0)	
Primary reason for admission: system‡				0.49
Cardiac	4,699 (41.6)	2,759 (41.3)	1,940 (42.1)	
Pulmonary	2,181 (19.3)	1,286 (19.2)	895 (19.4)	
GI	1,480 (13.1)	900 (13.5)	580 (12.6)	
Neurologic	1,582 (14.0)	923 (13.8)	659 (14.3)	
GU	269 (2.4)	172 (2.6)	97 (2.1)	
Overdose/poisoning	379 (3.4)	216 (3.2)	163 (3.5)	
Metabolic	392 (3.5)	232 (3.5)	160 (3.5)	
Hematologic/oncologic	115 (1.0)	71 (1.1)	44 (1.0)	
Other	198 (1.8)	125 (1.9)	73 (1.6)	
LOS				
Prior LOS,§ d	0.3 (0.1–0.8)	0.3 (0.1–0.8)	0.3 (0.1–0.8)	0.98
ICU LOS,† d	4.0 (6.4)	4.0 (6.7)	4.0 (6.2)	0.93
ICU LOS,§ d	2.0 (1.0–4.1)	2.0 (1.0–4.2)	1.9 (1.0–4.1)	0.24
ICU mortality‡	1,279 (11.4)	752 (11.3)	527 (11.4)	0.77
In-hospital mortality‡	1,766 (15.6)	1,036 (15.5)	730 (15.8)	0.63

GU = genitourinary.

*The p values are based on χ^2 test of statistical independence for categorical data, Student *t* test for parametric data, or Mann-Whitney test for nonparametric data. Totals may not add to 100% due to rounding.

†Values are given as the mean (SD).

‡Values are given as the No. (%).

§Values are given as the median (interquartile range).

medical vs surgical admission status, and the primary system affected leading to ICU admission are displayed in Table 3. APACHE IVorig, APACHE IVrecal, and MPM_0 III LOS each had a single age stratum with significant differences between observed and predicted LOS. SAPS II LOS systematically underpredicted LOS for younger patients and overpredicted LOS for older patients. APACHE

Table 2—*Coefficients for MPM_0 III LOS and SAPS II LOS Models*

Variables	Coefficient for Estimation Sample (n = 6,684)	95% CI
MPM_0 III LOS model		
Heart rate ≥ 150 beats/min	1.6517	0.9290 to 2.3744
SBP ≤ 90 mm Hg	0.1442	−1.0821 to 1.3704
Chronic kidney disease	−0.5952	−1.1567 to −0.0337
Cirrhosis	1.3865	−1.4989 to 4.2718
Coma/deep stupor	−1.4622	−3.4426 to 0.5182
Metastatic neoplasm	3.4601	1.1031 to 5.8171
Acute renal failure	0.6548	−0.1365 to 1.4461
Cardiac dysrhythmia	−0.9552	−3.0329 to 1.1225
Cerebrovascular incident	1.1122	0.5227 to 1.7016
GI bleed	−0.7975	−1.3560 to −0.2390
Intracranial mass effect	1.8107	−0.0294 to 3.6508
CPR before ICU admission	1.9279	−0.5657 to 4.4215
Mechanical ventilation	2.4888	2.1530 to 2.8246
Unscheduled surgical admission or medical admission	1.3964	1.0410 to 1.7518
Age (per 10 yr)	0.1369	0.0562 to 0.2176
Full code on ICU admission	0.8537	0.2926 to 1.4147
Zero risk factors (no factors other than age)	−0.6006	−0.9936 to −0.2076
Interaction terms		
Age coma/deep stupor	0.1247	−0.1714 to 0.4208
Age SBP ≤ 90 mm Hg	0.0165	−0.1667 to 0.1997
Age cirrhosis	−0.0546	−0.5703 to 0.4610
Age metastatic neoplasm	−0.4949	−0.8649 to −0.1249
Age cardiac dysrhythmia	−0.0051	−0.2941 to 0.2838
Age intracranial mass effect	−0.3209	−0.6210 to −0.0208
Age CPR prior to admission	−0.2442	−0.6078 to 0.1193
Intercept	0.5566	−0.3409 to 1.4541
SAPS II LOS model		
SAPS score	0.0178	0.0019 to 0.0337
Log (SAPS score)	1.6057	1.1150 to 2.0965
Intercept	−2.2334	−3.4928 to −0.9741

CPR = cardiopulmonary resuscitation; SBP = systolic BP.

IVrecal and MPM_0 III-LOS accurately predicted ICU LOS for medical and elective surgical patients. For more specific diagnostic categories, including emergency surgery, APACHE IVrecal was the most accurate.

For each decile of predicted ICU LOS, the difference between mean observed and predicted LOS differed significantly ($p \leq 0.05$) for 6, 3, 2, and 6 of the 10 deciles, respectively, using APACHE IVorig, APACHE IVrecal, MPM_0 III LOS, and SAPS II LOS (Table 4). This is graphically represented in Figure 1 as calibration curves. The calibration curve of APACHE IVorig demonstrates poor fit at the lowest deciles. APACHE IVrecal demonstrates excellent fit, with the poorest calibration in the lowest decile. MPM_0 III LOS demonstrates an excellent fit as well. SAPS II LOS appears to have a poor fit across multiple deciles.

The coefficients of determination for patient-level ICU LOS predictions were as follows: APACHE IVorig, $R^2 = 0.182$; APACHE IVrecal, $R^2 = 0.202$; MPM_0 III LOS, $R^2 = 0.098$; and SAPS II LOS, $R^2 = 0.049$. Grouped R^2 analysis for the 29 hospitals with > 100 admissions were as follows: APACHE IVorig, $R^2 = 0.439$; APACHE IVrecal, $R^2 = 0.422$; MPM_0 III LOS, $R^2 = 0.279$; and SAPS II LOS, $R^2 = 0.008$. This indicates that 42% and 28%, respectively, of the ICU LOS variations are accounted for by APACHE IVrecal and MPM_0 III-LOS.

Finally, Figure 2 displays a comparison of the predictions of the models for hospital-level SLOSRs, excluding the original APACHE model. Regardless of the model used, there was significant variation in SLOSRs among 29 hospitals with > 100 admissions. There were similar ranges among the SLOSRs of the hospitals for each model as follows: APACHE IVrecal, 0.47 to 1.60; MPM_0 III LOS, 0.40 to 1.68; and SAPS II LOS, 0.38 to 1.69. The intraclass correlations of the SLOSRs between each pair of models were high: APACHE IVrecal and MPM_0 III-LOS, $r = 0.89$ (95% CI, 0.74 to 0.96); APACHE IVrecal and SAPS II-LOS, $r = 0.85$ (95% CI, 0.70 to 0.93); and MPM_0 III-LOS and SAPS II-LOS, $r = 0.96$ (95% CI, 0.92 to 0.98).

Additional Risk Factors and Sensitivity Analyses

The addition of DNR status and Medicaid payment (when compared to private insurance) to APACHE IV models independently predicted shorter LOS (−1.10 days; 95% CI, −0.57 to −1.65) and longer LOS (0.74 days; 95% CI, 0.38 to 1.09), respectively. The number of hospital beds had no effect. Each of these factors did not significantly improve the accuracy, calibration, or agreement of hospital SLOSRs between each model. In addition, there was no statistically significant correlation between percentages of DNR patients ($r = 0.18$; $p = 0.36$) or Medicaid patients ($r = 0.35$; $p = 0.06$) of the hospital and the SLOSR. Likewise, there was no statistically significant correlation between bed size ($r = -0.25$; $p = 0.22$) and SLOSR.

Models developed on the population excluding patients with the short ICU LOS (< 24 h) maintained excellent calibration for APACHE IVrecal and improved calibration for MPM_0 III LOS. The range of SLOSRs for each model when excluding patients with LOS < 24 h (SLOSR range: APACHE

Table 3—*Difference Between Observed and Predicted LOS for Age and Primary Medical/Surgical System Categories on Validation Sample*

Variables	Patients, No.	APACHE IVorig Model		APACHE IVrecal Model		MPM_0 III LOS Model		SAPS II LOS Model	
		Difference of Observed Minus Predicted, d	p Value*	Difference of Observed Minus Predicted, d	p Value*	Difference of Observed Minus Predicted, d	p Value*	Difference of Observed Minus Predicted, d	p Value*
Age									
18–30 yr	224	0.5	0.08	0.2	0.4	0.0	0.93	0.8	0.006
31–45 yr	602	0.1	0.46	0.0	1.0	−0.1	0.61	0.4	0.03
46–60 yr	1,209	0.4	0.003	0.1	0.45	0.3	0.04	0.5	0.002
61–70 yr	864	0.2	0.14	0.0	0.89	0.1	0.39	0.1	0.74
71–80 yr	1,012	−0.1	0.31	−0.3	0.02	−0.2	0.11	−0.6	< 0.001
≥ 81 yr	700	0.2	0.37	−0.2	0.18	−0.2	0.18	−0.8	< 0.001
Medical vs surgical status									
Elective surgery	808	0.3	0.04	−0.1	0.64	0.0	0.8	−0.1	0.27
Emergency surgery	195	0.3	0.45	0.2	0.67	0.8	0.05	1.4	0.002
Medical	3608	0.2	0.04	−0.1	0.29	0.0	0.75	−0.1	0.45
Medical/surgical system									
Cardiac medical	1,670	0.0	0.88	−0.2	0.02	−0.4	< 0.001	−0.8	< 0.001
Cardiac surgical	270	0.5	0.03	0.2	0.46	−0.2	0.48	−0.4	0.1
Pulmonary medical	759	0.7	0.006	0.3	0.28	1.5	< 0.001	1.9	< 0.001
Pulmonary surgical	136	0.7	0.10	−0.2	0.61	0.5	0.21	0.4	0.38
GI medical	297	−0.1	0.53	−0.4	0.12	−0.4	0.1	−0.7	0.002
GI surgical	283	0.2	0.60	0.0	0.88	0.7	0.03	0.8	0.009
Neurologic medical	441	−0.1	0.56	−0.3	0.22	−0.3	0.28	0.3	0.29
Neurologic surgical	218	0.1	0.77	−0.1	0.8	0.0	0.91	0.1	0.63
GU medical	63	0.7	0.37	0.5	0.55	0.4	0.66	0.0	0.97
GU surgical	34	−0.1	0.77	−0.1	0.83	−0.4	0.26	−0.7	0.04
Overdose/poisoning	163	0.21	0.38	0.2	0.5	−1.4	< 0.001	−0.8	0.001
Metabolic	160	0.5	0.03	0.2	0.31	−0.7	0.003	−1.0	< 0.001
Hematology/oncology	44	−0.8	0.03	−1.4	< 0.001	−1.3	0.003	−1.6	< 0.001
Other	73	1.1	0.12	1.0	0.13	0.3	0.7	0.6	0.37

*Based on paired Student *t* tests. See Table 1 for abbreviation not used in the text.

Table 4—*Differences Between Observed and Predicted LOS Across Decile of Predicted LOS for Each Model in Validation Data Set*

Decile of Predicted ICU LOS,* %	APACHE IVorig Model					APACHE IVrecal Model					MPM III_0 LOS Model					SAPS II LOS Model				
	Patients, No.	Mean Observed ICU LOS, d	Mean Predicted ICU LOS, d	Difference of Observed-Predicted LOS, d	p Value†	Patients, No.	Mean Observed ICU LOS, d	Mean Predicted ICU LOS, d	Difference of Observed-Predicted LOS, d	p Value†	Patients, No.	Mean Observed ICU LOS, d	Mean Predicted ICU LOS, d	Difference of Observed-Predicted LOS, d	p Value†	Patients, No.	Mean Observed ICU LOS, d	Mean Predicted ICU LOS, d	Difference of Observed-Predicted LOS, d	p Value†
0–10	462	1.5	0.6	0.9	< 0.001	462	1.5	0.9	0.7	< 0.001	462	2.4	2.0	0.3	0.05	564	2.0	1.8	0.2	0.08
11–20	461	1.8	1.2	0.6	< 0.001	461	1.8	1.6	0.1	0.19	462	2.4	2.5	−0.1	0.45	381	2.4	2.8	−0.4	0.03
21–30	461	2.1	1.6	0.5	< 0.001	461	2.1	2.0	0.1	0.33	461	2.7	2.7	0.0	0.84	525	2.5	3.1	−0.6	< 0.001
31–40	461	2.4	2.1	0.4	0.003	461	2.5	2.4	0.1	0.48	464	3.1	3.0	0.1	0.53	426	2.7	3.4	−0.7	< 0.001
41–50	461	2.9	2.7	−0.2	0.23	461	2.8	3.0	−0.2	0.16	457	3.2	3.1	0.1	0.67	470	3.6	3.7	−0.1	0.77
51–60	461	3.2	3.5	−0.3	0.11	461	3.2	3.6	−0.4	0.03	461	2.8	3.3	−0.5	< 0.001	446	3.6	4.0	−0.4	0.10
61–70	461	3.9	4.3	−0.4	0.05	461	4.2	4.5	−0.3	0.17	461	3.8	3.9	−0.1	0.66	454	4.7	4.3	0.4	0.11
71–80	461	4.9	5.3	−0.4	0.16	461	4.9	5.4	−0.5	0.04	462	4.4	4.8	−0.4	0.15	457	5.4	4.6	0.8	< 0.01
81–90	461	6.1	6.5	−0.4	0.18	461	6.6	6.6	0.0	0.92	464	6.1	5.8	0.3	0.33	456	6.8	5.1	1.7	< 0.001
91–100	461	9.1	8.3	0.8	0.05	461	8.4	8.7	−0.3	0.48	457	7.1	6.7	0.4	0.29	432	4.6	5.9	−1.4	< 0.001

*Population sorted by increasing predicted risk and then split into deciles.
†Based on paired Student *t* test.

IVrecal, 0.58 to 1.49; MPM_0 III LOS, 0.61 to 1.46; and SAPS II LOS, 0.55 to 1.53) was smaller than the range of SLOSRs produced when using all patients in the sample, with comparable agreement. There was no correlation between the mean severity of illness of the hospitals ($r = -0.05$; $p = 0.80$) and the SLOSR. The mean SLOSRs of the five hospitals with the lowest and highest mean severity of illness were 1.0 (SD, 0.2) and 1.0 (SD, 0.3), respectively.

Finally, a model based on the SAPS II LOS independent variables revealed no meaningful differences in accuracy ($R^2 = 0.061$) and calibration between that and the primary SAPS II model used in the analyses just cited. No further data from that model are presented.

Discussion

Our study is the first description of the use of MPM_0 III LOS and SAPS II LOS variables for the additional purpose of predicting risk-adjusted ICU LOS. In addition, our study is the first independent validation of the APACHE IV LOS model. We have shown MPM_0 III LOS, an alternative risk-adjustment model originally developed for mortality prediction, can also be used for predicting LOS in a broad medical and surgical population. However, SAPS II LOS did not appear well suited for LOS prediction. The MPM_0 III LOS model explains the lower variation in hospital-level LOSs but requires substantially fewer resources to implement than the APACHE IV LOS model. Individual hospitals received similar rankings with these two models.

Regardless of the model, we observed sizable variations in risk-adjusted LOS performance among hospitals that could not be accounted for by patient risk factors. The apparent variation in ICU LOS after accounting for differences in patient severity of illness supports the need to assess risk-adjusted ICU LOS as one aspect of performance.

The primary objective of our study was to assess the utility of two established mortality prediction models in predicting an alternative outcome, ICU LOS, and to compare these models to the APACHE IVorig and APACHE IVrecal models. With regard to model accuracy, APACHE IVrecal has the best predictive accuracy across clinical categories, excellent calibration, and the highest grouped R^2. The APACHE IVrecal model proved more accurate when compared to the APACHE IVorig model. There are many potential reasons for this, as follows: (1) the CALICO cohort had a different patient mix, including more nonsurgical patients and higher

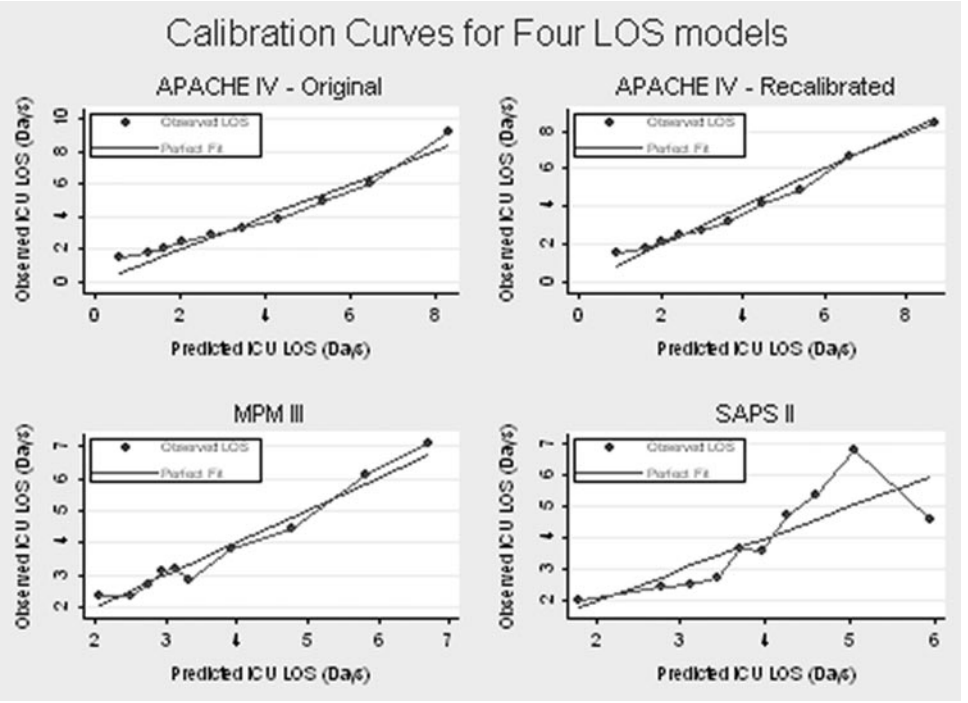

FIGURE 1. Calibration curves comparing mean observed and mean predicted ICU LOS for four ICU LOS models.

mean APACHE score; (2) when compared to APACHE IVorig, the coefficients for individual risk factors differed across many domains, including, but not limited to, acute and chronic diagnoses; (3) patterns in health-care utilization may differ in the CALICO cohort; and (4) in contrast to CALICO hospitals, the APACHE IV cohort hospitals were users of the APACHE system,[15] which could be a marker of increased attention toward quality, efficiency, and information technology.

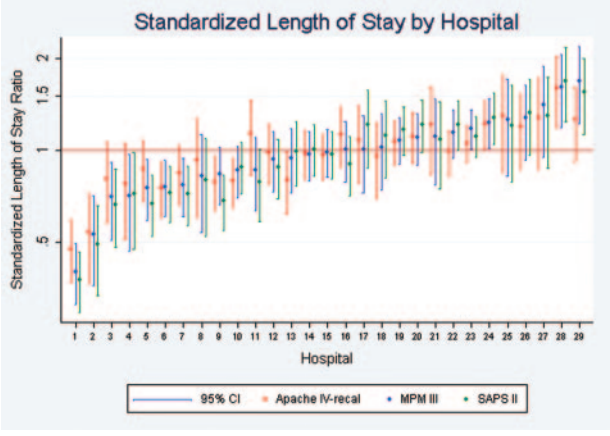

FIGURE 2. Plot of LOS prediction model-specific SLOSRs for each hospital with at least 100 admissions.

The superior predictive accuracy of APACHE IVrecal compared to the other models may be explained by having more variables. Including the ICU admitting diagnosis may be particularly influential because prior research[15] has shown that they account for up to 17% of the explanatory power of the original APACHE IV model. In addition, the use of linear splines to model nonlinearities in predictor response (*eg*, acute physiology score) address the reality that patients with both the lowest and highest acute physiology scores will generally have shorter average LOSs.[15] Alternatively, it may be that part of the additional predictive power comes from including variables that reflect pre-ICU care, such as pre-ICU LOS and admission source, or response to treatment (because the worst physiology values for the first 24 h are included). Further research is needed to define the source of the additional predictive power and to assess whether including these variables is actually desirable. For instance, if the model predicts LOS better because it "risk adjusts" for undertreatment, that may not be desirable.

The poor accuracy of SAPS II LOS suggests that this model is inadequate for predicting LOS. The limited value of the SAPS II LOS model might be improved by reweighting the individual variables that make up the SAPS II LOS score or modeling

their relationships to LOS as nonlinear. Treating the individual variables as independent rather than summarized did not provide significant additional benefit.

With > 100 fewer model coefficients than APACHE IVrecal and without modeling nonlinear relationships, the MPM_0 III LOS model nonetheless displayed fair accuracy and excellent calibration. Despite a low R^2 for predicting an individual patient's LOS, MPM_0 III LOS was effective in predicting LOS across hospital, demographic, and broad clinical groups. The inability of the MPM_0 III LOS model to predict LOS especially well for derangements of an individual physiologic system reflects the absence in the MPM_0 III model of a variable indicating the system involved. This suggests that MPM_0 III LOS may be poorly suited for assessing the performance of individual specialty ICUs. MPM_0 III LOS may also be poorly suited for assessing ICUs that care for a large proportion of emergency surgery patients (*eg*, trauma ICU). Despite being statistically significant, differences between predicted LOS and actual LOS did not always appear to be clinically significant (*eg*, for the medical cardiac system, a difference of < 12 h). Therefore, if predictions for clinical subgroups are an important goal, the MPM_0 III LOS model may be considered, albeit with caution.

MPM_0 III LOS and APACHE IVrecal were also similar in their appraisals of hospital performance. Performance assessments from the two models were highly correlated ($r = 0.89$) and were not significantly affected by additional patient and hospital factors (*eg*, DNR status, payor status, number of hospital beds). Limiting the sample to patients with an ICU stay of at least 24 h maintained high correlation ($r = 0.85$) and improved calibration of the MPM_0 III LOS model. Improvement in calibration may reflect difficulty in predicting LOS for patients with very short ICU stays due to low severity of illness or early mortality. Performance estimates on this reduced sample were more conservative, as evidenced by a narrower range of SLOSRs. Therefore, one would expect fewer performance outliers in the restricted sample.

With respect to model accuracy, the APACHE IV LOS model is a superior tool for LOS risk adjustment. APACHE IV is an excellent tool for hospital mortality risk adjustment and, unlike the MPM_0 III model, has been applied as well to CABG patients. However, there are real-world limitations in data collection, so using MPM_0 III may be a legitimate consideration. First, MPM_0 III is a validated tool for risk-adjusted mortality,[18] and it involves about a third the data collection time of APACHE IV.[20] Few hospitals currently have ICU risk variables available electronically, and the degree to which hospitals face resource and technology barriers may influence the preferences for MPM_0 III LOS vs APACHE IV LOS.[37,38] However, this benefit of the MPM_0 III LOS model may be lessened if hospitals are not currently using a risk-adjustment model for CABG patients and are considering the measurement of ICU and CABG outcomes. Second, because model performance deteriorates over time or when applied to populations that differ from the one used for model development, another factor to consider is the ability to reestimate the model to the study population. With substantially fewer coefficients, reestimation of the MPM_0 III LOS requires a smaller database and, hence, can be performed more often or when the size of the database does not allow for the recalibration of APACHE. This problem with APACHE would be lessened if the Joint Commission was to adopt a national ICU performance set, therefore creating a large national database with which frequent recalibration would be possible with any model. Finally, the MPM_0 III LOS model only uses risk information from the first hour after a patient's ICU admission, whereas the APACHE IV LOS model requires data be collected throughout the first day of ICU care. Limiting the data collection period may decrease the resources needed to collect data and limits the influence of treatment on the predicted LOS. For example, although hypotension that results from sepsis should be included as a risk factor, hypotension caused by failure to treat appropriately (*eg*, not starting appropriate therapy with antibiotics in sepsis patients) should not. Models that use post-hospital admission data cannot distinguish between these cases, so their better predictive ability may not always serve the purpose of identifying the best performing ICUs.

Our study has important limitations to consider. One is that we used a convenience sample of volunteer hospitals from California. Despite this, the sampling strategy is more likely to affect the estimation of individual model coefficients and is less likely to affect the comparisons between the models. We would recommend a reestimation of the coefficients for all models if applied to a national sample. Second, our hospital sample has a limited number of performance outliers. A larger sample of hospitals is needed to draw more reliable conclusions about the validity of the three models for identifying performance outliers. Third, the recently updated SAPS III model[39] became available after our data collection began, so we did not capture all of its required data elements. Finally, although LOS may be a useful

measure, it is likely affected by hospital discharge policies, bed availability, and community resources. Adding information about these factors might improve the predictive capacity of LOS models, although it would require frequently updated hospital-level information (*eg*, the number of stepdown unit or regular ward beds that are available on each hospital day). In addition, adding these factors to LOS models would mask the extent to which the management of these resources by a hospital contributes to its ICU LOS. Because understanding (and eliminating) the impact of such factors is a goal of clinicians and policymakers who seek to assess ICU LOS, their inclusion in predictive models would improve accuracy but might reduce the relevance of the assessments. In any case, risk-adjusted LOS should be used as a complementary measure to a suite of ICU performance measures, including structural, process, and outcomes measures of performance, because these other measures may both help to explain variations in ICU LOS and contribute to efforts to improve performance.[33,40,41]

In summary, the APACHE IVrecal and MPM_0 III LOS model are more accurate than the SAPS II LOS model for the prediction of ICU LOS. APACHE IVrecal is the most accurate LOS prediction model for specific ICU subpopulations. This is in part due to its larger number of variables, but it also likely reflects a longer window of data collection (the first 24 h, instead of the first hour, in the ICU). It is the preferred model when either ample resources are available for data collection or the APACHE IV variables can be generated by an electronic medical record, and there are no concerns about treatment impacting measured severity of illness over the first day of treatment. The MPM_0 III LOS model is less accurate, although it performs well across broad hospital populations, imposes less of a data collection burden, uses a shorter data collection window, and, therefore, is less likely to be influenced by treatment. The final choice of a model by physicians, hospitals, quality-reporting groups, or payers must reflect value judgments regarding the balance between predictive accuracy and data burden. Only with a wider application of risk-adjusted LOS and mortality measures will we understand those factors that account for the large observed differences in hospital outcomes and be able to accelerate improvements in ICU care.

ACKNOWLEDGMENT: We acknowledge Teresa Chipps, BS, Department of Medicine (General Internal Medicine and Public Health), Center for Health Services Research, Vanderbilt University, Nashville, TN, for her administrative and editorial assistance in the preparation of this article.

Appendix

Appendix—*Reestimated Coefficients for APACHE IV LOS Model*

Variables	Coefficient Estimation Sample (n = 6,684)	95% CI
Age	0.0078	−0.0234 to 0.0390
Knot = 27	0.000001	−0.00003 to 0.00003
Knot = 51	−0.000059	−0.0003 to 0.0001
Knot = 64	0.00027	−0.0003 to 0.0009
Knot = 74	−0.00066	−0.0016 to 0.0003
Knot = 86	0.0021	−0.0003 to 0.0045
Comorbidity		
None	Reference	Reference
Cirrhosis	−0.0547	−0.8426 to 0.7334
Immunosuppressed	−0.0917	−0.6706 to 0.4873
Cancer, metastatic	−0.2231	−0.8596 to 0.4134
Lymphoma	0.0901	−1.1180 to 1.2981
Hepatic failure	2.3535	1.2357 to 3.4713
AIDS	−0.4178	−1.8666 to 1.0310
Leukemia, myeloma	0.8278	−0.3980 to 2.0537
APS	0.0411	−0.0204 to 0.1025
Knot = 10	−0.000034	−0.0002 to 0.0001
Knot = 22	0.00016	−0.0002 to 0.0006
Knot = 32	−0.00021	−0.0006 to 0.0001
Knot = 48	0.000085	−0.00002 to 0.0002
Knot = 89	0.000001	−0.00003 to 0.00003
PaO_2/FIO_2 ratio	−0.0052	−0.0063 to −0.0041
Ventilated on ICU day 1	1.8966	1.5566 to 2.2366
Admission source		
Other	Reference	Reference
Floor	0.3217	−0.0208 to 0.6643
Other hospital	1.3000	0.6194 to 1.9807
Operating/recovery room	−1.0302	−2.2836 to 0.2233
Emergency surgery	1.1476	0.5190 to 1.7762
Previous LOS	−0.2760	−1.4315 to 0.8795
Knot = 0.121	1.7218	−1.0812 to 4.5249
Knot = 0.423	−3.3143	−8.8047 to 2.1762
Knot = 0.794	1.6265	−1.1756 to 4.4285
Knot = 2.806	−0.0392	−0.1899 to 0.1114
Thrombolytic therapy for AMI	0.3031	−0.6018 to 1.2080
GCS score	0.0215	−0.0214 to 0.0645
Unable to assess GCS	0.7593	0.3503 to 1.1682
Nonoperative diagnostic groups		
Cardiovascular diagnoses		
AMI		
Anterior	0.0926	−0.8988 to 1.0841
Inferior/lateral	−0.2644	−1.2252 to 0.6964
Non-Q wave	−0.6638	−2.2126 to 0.8849
Other	Reference	Reference
Cardiac arrest	1.8213	0.2694 to 3.3731
Cardiogenic shock	0.8254	−0.5682 to 2.2191
Cardiomyopathy	−0.2542	−2.3527 to 1.8442
Congestive heart failure	−0.1450	−0.9686 to 0.6785
Chest pain, rule out AMI	1.0292	−2.1827 to 4.2410
Hypertension	−0.3278	−1.5456 to 0.8899

(Continued)

Appendix—*(Continued)*

Variables	Coefficient Estimation Sample (n = 6,684)	95% CI
Hypovolemia/dehydration (not shock)	−0.5539	−2.9398 to 1.8320
Hemorrhage (not related to GI bleeding)	−1.8497	−4.8867 to 1.1873
Aortic aneurysm, dissecting	1.5569	−0.9018 to 4.0156
Peripheral vascular disease	0.1520	−1.7145 to 2.0185
Rhythm disturbance	−0.3191	−1.1107 to 0.4725
Sepsis		
Cutaneous	0.2151	−1.8327 to 2.2629
GI	0.3856	−1.3586 to 2.1298
Pulmonary	2.2312	0.8886 to 3.5737
Urinary tract	0.6214	−0.5759 to 1.8188
Other	0.0842	−2.9556 to 3.1241
Unknown	0.4545	−0.5199 to 1.4289
Cardiac drug toxicity	0.4403	−1.9391 to 2.8198
Unstable angina	−0.2866	−1.2664 to 0.6932
Cardiovascular, other	−0.0935	−0.9220 to 0.7351
Respiratory diagnoses		
Airway obstruction	−1.1566	−2.5816 to 0.2683
Asthma	−0.9504	−2.4029 to 0.5021
Aspiration pneumonia	1.8594	0.6822 to 3.0366
Bacterial pneumonia	1.3593	0.5127 to 2.2059
Viral pneumonia	11.9734	7.9610 to 15.9858
Parasitic/fungal pneumonia	−0.3144	−2.4677 to 1.8390
COPD	−0.5337	−1.4327 to 0.3653
Pleural effusion	2.3729	0.2764 to 4.4693
Pulmonary edema (noncardiac, ARDS)	1.8502	0.5768 to 3.1236
Pulmonary embolism	0.0365	−1.3239 to 1.3969
Respiratory arrest	5.5090	2.4528 to 8.5652
Respiratory cancer	1.6241	−0.7706 to 4.0187
Restrictive lung disease	−0.3943	−3.4324 to 2.6439
Respiratory, other	0.6541	−0.2716 to 1.5797
GI diagnoses		
GI bleeding, upper	−0.1162	−1.0717 to 0.8393
GI bleeding, lower	0.0846	−1.2942 to 1.4634
GI bleeding, varices	0.0706	−1.1279 to 1.2691
GI inflammatory disease	2.0000	0.1665 to 3.8335
Neoplasm	−0.1524	−2.4206 to 2.1158
Obstruction	−1.5949	−4.4752 to 1.2853
Perforation	2.3205	−2.1588 to 6.7999
Vascular insufficiency	0.3367	−5.9729 to 6.6464
Hepatic failure	1.3973	−0.8488 to 3.6434
Intra/retroperitoneal hemorrhage	−0.0192	−4.0357 to 3.9974
Pancreatitis	−0.0271	−2.1165 to 2.0623
GI, other	1.0184	−0.7128 to 2.7496
Neurologic diagnoses		
Intracerebral hemorrhage	1.1529	0.2131 to 2.0927
Neurologic neoplasm	0.1640	−1.9908 to 2.3188
Neurologic infection	0.2610	−1.4320 to 1.9541
Neuromuscular disease	−0.3268	−2.9793 to 2.3256
Drug overdose	−0.9729	−1.8955 to −0.0502
Subdural/epidural hematoma	0.4392	−1.0542 to 1.9326
Subarachnoid hemorrhage, intracranial aneurysm	2.9454	1.6706 to 4.2203

(Continued)

Appendix—*(Continued)*

Variables	Coefficient Estimation Sample (n = 6,684)	95% CI
Seizures (no structural disease)	−0.0589	−1.2930 to 1.1753
Stroke	0.9552	−0.3379 to 2.2483
Neurologic, other	2.3299	0.8984 to 3.7615
Metabolic/endocrine diagnoses		
Acid-base, electrolyte disorder	0.1873	−1.6411 to 2.0156
Diabetic ketoacidosis	−0.6338	−1.6196 to 0.3521
Diabetic HHNC	−0.5630	−1.7594 to 0.6334
Metabolic/endocrine, other	−0.2237	−1.6751 to 1.2277
GU diagnoses		
Renal, other	0.1151	−0.9682 to 1.1983
Miscellaneous diagnoses		
General, other	−0.2454	−1.7009 to 1.2101
Operative diagnoses		
Cardiovascular surgery		
Valvular heart surgery	−1.0431	−2.8156 to 0.7295
Aortic aneurysm, elective repair	0.4275	−1.3497 to 2.2047
Aortic aneurysm, ruptured	0.5937	−4.0943 to 5.2817
Aortic aneurysm, dissection	0.3527	−2.8310 to 3.5364
Femoral-popliteal bypass graft	0.3356	−1.5599 to 2.2311
Aortoiliac, aortofemoral bypass graft	0.9262	−2.5131 to 4.3654
Peripheral ischemia (emobolectomy, thrombectomy, dilation)	−0.4225	−4.6282 to 3.7832
Carotid endarterectomy	0.8925	−0.7279 to 2.5129
Cardiovascular surgery, other	0.1896	−1.5406 to 1.9198
Respiratory surgery		
Thoracotomy, malignancy	0.9806	−0.7279 to 2.6892
Neoplasm, mouth, larynx	1.5202	−0.9609 to 4.0013
Thoracotomy, lung biopsy, pleural disease	4.8600	1.7232 to 7.9968
Thoracotomy, respiratory infection	0.2357	−2.3060 to 2.7774
Respiratory surgery, other	1.7429	−0.0452 to 3.5310
GI surgery		
GI malignancy	1.7652	0.0896 to 3.4409
GI bleeding	0.8628	−1.4034 to 3.1291
Fistula, abscess	−0.8891	−3.8190 to 2.0408
Cholecystitis, cholangitis	−0.0360	−2.0664 to 1.9945
GI inflammation	1.8150	−0.9391 to 4.5692
GI obstruction	0.1693	−1.6523 to 1.9909
GI perforation	2.5490	0.6072 to 4.4909
GI vascular ischemia	5.2939	1.8331 to 8.7548
Liver transplant	−3.1338	−7.3945 to 1.1270
GI surgery, other	0.0103	−1.5726 to 1.5932
Neurologic surgery		
Craniotomy or transsphenoidal procedure for neoplasm	0.7337	−0.8877 to 2.3552
Intracranial hemorrhage	1.8154	−1.0389 to 4.6697
Subarachnoid hemorrhage, intracranial aneurysm	2.9454	1.6706 to 4.2203

(Continued)

Appendix—*(Continued)*

Variables	Coefficient Estimation Sample (n = 6,684)	95% CI
Subdural/epidural hematoma	0.4392	−1.0542 to 1.9326
Laminectomy, fusion, spinal cord surgery	0.7094	−1.0999 to 2.5188
Neurologic surgery, other	0.6249	−1.1605 to 2.4102
Genitourinary surgery		
Renal/bladder/prostate neoplasm	0.2622	−3.2134 to 2.4526
Renal transplant	−2.0178	−7.3254 to 3.2897
Hysterectomy	−0.1985	−3.2134 to 2.8164
Genitourinary surgery, other	0.4942	−1.9733 to 2.9617
Miscellaneous surgery		
Amputation, nontraumatic	−0.3057	−9.2631 to 8.6516
Intercept	2.2550	−4.4486 to 8.9587

Knot = numerical cut point for each splined variable; APS = acute physiology score; FIO_2 = fraction of inspired oxygen; GCS = Glasgow coma scale; AMI = acute myocardial infarction; HHNC = hyperglycemic hyperosmolar nonketotic coma. See Table 1 for abbreviations not used in the text.

References

1 Halpern NA, Pastores SM, Greenstein RJ. Critical care medicine in the United States 1985–2000: an analysis of bed numbers, use, and costs. Crit Care Med 2004; 32: 1254–1259

2 Rapoport J, Teres D, Lemeshow S, et al. Explaining variability of cost using a severity-of-illness measure for ICU patients. Med Care 1990; 28:338–348

3 Rapoport J, Teres D, Lemeshow S, et al. A method for assessing the clinical performance and cost-effectiveness of intensive care units: a multicenter inception cohort study. Crit Care Med 1994; 22:1385–1391

4 Render ML, Kim HM, Deddens J, et al. Variation in outcomes in Veterans Affairs intensive care units with a computerized severity measure. Crit Care Med 2005; 33: 930–939

5 Rosenthal GE, Harper DL, Quinn LM, et al. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals: results of a regional study. JAMA 1997; 278:485–490

6 Woods AW, MacKirdy FN, Livingston BM, et al. Evaluation of predicted and actual length of stay in 22 Scottish intensive care units using the APACHE III system. Anaesthesia 2000; 55:1058–1065

7 Lilly CM, Sonna LA, Haley KJ, et al. Intensive communication: four-year follow-up from a clinical practice study. Crit Care Med 2003; 31(suppl):S394–S399

8 Pronovost PJ, Angus DC, Dorman T, et al. Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review. JAMA 2002; 288:2151–2162

9 Pronovost PJ, Jenckes MW, Dorman T, et al. Organizational characteristics of intensive care units related to outcomes of abdominal aortic surgery. JAMA 1999; 281:1310–1317

10 Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 2000; 342:1301–1308

11 Gupta N, Kotler PL, Dudley RA. Considerations in the development of intensive care unit report cards. J Intensive Care Med 2002; 17:211–217

12 Joint Commission on Accreditation of Healthcare Organizations. National hospital quality measures: ICU. Available at: http://www.jointcommission.org/PerformanceMeasurement/MeasureReserveLibrary/Spec+Manual+-+ICU.htm. Accessed May 18, 2009

13 Hospital Association of Southern California. Quality/patient safety resources: 2008 CHART hospital performance measures. Available at: http://www.hasc.org/download.cfm?ID=28358. Accessed June 4, 2009

14 Knaus WA, Wagner DP, Zimmerman JE, et al. Variations in mortality and length of stay in intensive care units. Ann Intern Med 1993; 118:753–761

15 Zimmerman JE, Kramer AA, McNair DS, et al. Intensive care unit length of stay: benchmarking based on Acute Physiology and Chronic Health Evaluation IV. Crit Care Med 2006; 34:2517–2529

16 Higgins TL, Teres D, Copes WS, et al. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). Crit Care Med 2007; 35:827–835

17 Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993; 270:2957–2963

18 Tri-Analytics, Inc. Project IMPACT CCM's Critical Care Data Systems. Available at: http://www.trianalytics.com/programs_pi.html. Accessed June 4, 2009

19 California HealthCare Foundation. Rating hospital quality in California, 2008. Available at: http://www.calhospitalcompare.org. Accessed May 18, 2009

20 Kuzniewicz MW, Vasilevskis EE, Lane R, et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest 2008; 133:1319–1327

21 Young JD, Goldfrad C, Rowan K. Development and testing of a hierarchical method to code the reason for admission to intensive care units: the ICNARC coding method; Intensive Care National Audit & Research Centre. Br J Anaesth 2001; 87:543–548

22 American Hospital Association. AHA Annual Survey Database. 2004 ed. Chicago, IL: American Hospital Association, 2004

23 Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol 2003; 56:441–447

24 van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. Stat Med 2000; 19:3401–3415

25 Fisher ES, Wennberg DE, Stukel TA, et al. The implications of regional variations in Medicare spending: part 1. The content, quality, and accessibility of care. Ann Intern Med 2003; 138:273–287

26 Wennberg JE, Fisher ES, Stukel TA, et al. Use of hospitals, physician visits, and hospice care during last six months of life among cohorts loyal to highly respected hospitals in the United States. BMJ 2004; 328:607

27 US Census Bureau. United States Census 2000, migration by race and Hispanic origin for the population 5 years and over for the United States, regions, states, and Puerto Rico: 2000 (PHC-T-25); 2008. Available at: http://www.census.gov/population/www/cen2000/briefs/phc-t25/tables/tab01.pdf. Accessed May 18, 2009

28 Nelson DE, Bolen J, Wells HE, et al. State trends in uninsurance among individuals aged 18 to 64 years: United States, 1992–2001. Am J Public Health 2004; 94:1992–1997

29 Burwen DR, Galusha DH, Lewis JM, et al. National and state trends in quality of care for acute myocardial infarction between 1994–1995 and 1998–1999: the Medicare health care quality improvement program. Arch Intern Med 2003; 163:1430–1439
30 Jencks SF, Cuerdon T, Burwen DR, et al. Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. JAMA 2000; 284:1670–1676
31 Fieller EC. A fundamental formula in the statistics of biological assay, and some applications. Q J Pharm Pharmacol 1944; 17:117–123
32 Nathanson BH, Higgins TL, Teres D, et al. A revised method to assess intensive care unit clinical performance and resource utilization. Crit Care Med 2007; 35:1853–1862
33 Angus DC, Linde-Zwirble WT, Sirio CA, et al. The effect of managed care on ICU length of stay: implications for Medicare. JAMA 1996; 276:1075–1082
34 Jayes RL, Zimmerman JE, Wagner DP, et al. Variations in the use of do-not-resuscitate orders in ICUs: findings from a national study. Chest 1996; 110:1332–1339
35 Arabi Y, Venkatesh S, Haddad S, et al. The characteristics of very short stay ICU admissions and implications for optimizing ICU resource utilization: the Saudi experience. Int J Qual Health Care 2004; 16:149–155
36 Rosenthal GE, Sirio CA, Shepardson LB, et al. Use of intensive care units for patients with low severity of illness. Arch Intern Med 1998; 158:1144–1151
37 Ash J, Gorman P, Seshadri V, et al. Computerized physician order entry in U.S. hospitals: results of a 2002 survey. J Am Med Inform Assoc 2004; 11:95–99
38 Poon E, Jha A, Christino M, et al. Assessing the level of healthcare information technology adoption in the United States: a snapshot. BMC Med Inform Decis Mak 2006; 6:1
39 Moreno R, Metnitz P, Almeida E, et al. From evaluation of the patient to evaluation of the intensive care unit: part 2. Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med 2005; 31:1345–1355
40 Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. BMJ 1995; 311:793–796
41 Wagner DP, Knaus WA, Harrell FE, et al. Daily prognostic estimates for critically ill adults in intensive care units: results from a prospective, multicenter, inception cohort analysis. Crit Care Med 1994; 22:1359–1372