Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients*

Jack E. Zimmerman, MD, FCCM; Andrew A. Kramer, PhD; Douglas S. McNair, MD, PhD; Fern M. Malila, RN, MS

Objective: **To improve the accuracy of the Acute Physiology and Chronic Health Evaluation (APACHE) method for predicting hospital mortality among critically ill adults and to evaluate changes in the accuracy of earlier APACHE models.**

Design: **Observational cohort study.**

Setting: **A total of 104 intensive care units (ICUs) in 45 U.S. hospitals.**

Patients: **A total of 131,618 consecutive ICU admissions during 2002 and 2003, of which 110,558 met inclusion criteria and had complete data.**

Interventions: **None.**

Measurements and Main Results: **We developed APACHE IV using ICU day 1 information and a multivariate logistic regression procedure to estimate the probability of hospital death for randomly selected patients who comprised 60% of the database. Predictor variables were similar to those in APACHE III, but new variables were added and different statistical modeling used. We assessed the accuracy of APACHE IV predictions by comparing observed and predicted hospital mortality for the excluded patients (validation set). We tested discrimination and used multiple tests of calibration in aggregate and for patient subgroups. APACHE IV had good discrimination (area under the receiver operating characteristic curve = 0.88) and calibration (Hosmer-Lemeshow C statistic = 16.9, $p = .08$). For 90% of 116 ICU admission diagnoses, the ratio of observed to predicted mortality was not significantly different from 1.0. We also used the validation data set to compare the accuracy of APACHE IV predictions to those using APACHE III versions developed 7 and 14 yrs previously. There was little change in discrimination, but aggregate mortality was systematically overestimated as model age increased. When examined across disease, predictive accuracy was maintained for some diagnoses but for others seemed to reflect changes in practice or therapy.**

Conclusions: **APACHE IV predictions of hospital mortality have good discrimination and calibration and should be useful for benchmarking performance in U.S. ICUs. The accuracy of predictive models is dynamic and should be periodically retested. When accuracy deteriorates they should be revised and updated. (Crit Care Med 2006; 34:1297–1310)**

KEY WORDS: intensive care unit; patient outcome assessment; severity of illness index; prognostication; health outcome research; informatics

Scoring systems based on physiologic abnormalities have been successful in measuring severity of illness among critically ill patients. Severity scores are used in clinical research to demonstrate equivalency of study and control patients, as inclusion criteria for study enrollment, and for risk stratification in outcome comparisons. Examples include Acute Physiology and Chronic Health Evaluation (APACHE) III (1), Simplified Acute Physiology Score (SAPS) II (2), and the Mortality Probability Model (MPM) II (3). Each of these prognostic models has been used to compare observed outcomes to a case-mix-adjusted benchmark for hospital mortality based on outcomes that reflect the efficacy of treatment during 1988–1992.

Studies using APACHE III, SAPS II, and MPM II within independent intensive care unit (ICU) databases have reported a predicted mortality that was significantly different from observed (4, 5). These differences between observed and expected mortality might have been caused by poor model design, variations in quality of care, or inadequate case-mix-related adjustment. Proposed reasons for inadequate case mix adjustment have included inadequate diagnostic data (6), unreliable Glasgow Coma Scale (GCS) score assessment (7, 8), international and regional differences (9, 10), variations in patient referral patterns (11, 12), and differing selection for and timing of ICU admission (13). In addition, predicted outcomes are likely to be influenced by changes in the effectiveness of therapy over time (5, 6), the frequency of decisions to forgo life-sustaining therapy (14), care before and after ICU admission (13, 15), and the frequency of early discharge to skilled nursing facilities (16).

Strategies suggested to improve the accuracy of prognostic models include a) reestimating the coefficients for each original variable for a specific population; or b) adding a population-specific variable customized to the original model (17). Both types of customization have resulted in improved calibration for aggregate patient samples but have not adequately adjusted for poor uniformity of

*See also p. 1538.

From the George Washington University, Washington, DC (JEZ); and Cerner Corporation, Vienna, VA (JEZ, AAK, FMM) and Kansas City, MO (DSM).

Supported by Cerner Corporation, Kansas City, MO.

Dr. Zimmerman provides consulting services to Cerner Corp.; Dr. Kramer is an employee of and has stock ownership and stock options in Cerner Corp.; Dr. McNair is an employee of, holds equity interest, stock ownership, and stock options in, and has patents for Cerner Corp.; Ms. Malila is an employee of and has stock ownership and stock options in Cerner Corp.

Address requests for reprints to: Andrew A. Kramer, PhD, 1953 Gallows Road, Suite 570, Vienna, VA 22182.

DOI: 10.1097/01.CCM.0000215112.84523.F0

fit within patient subgroups (5, 18–20). These findings suggest that recalibrated models fail to account for important patient risk factors and that remodeling is needed.

Since APACHE III was published (1, 21), we have repeatedly assessed its calibration using patient data collected during subsequent time periods. For patient data collected between 1993 and 1996, the original APACHE III model (version H) had a standardized mortality ratio (SMR) of 1.01 in the entire patient sample, but calibration was not perfect (Hosmer-Lemeshow statistic = 48.71, $p < .01$) (22). In 1998, APACHE III version H was revised and updated using data collected between 1993 and 1996. The revised model (APACHE III version I) had an SMR = 1.00 and a Hosmer-Lemeshow statistic = 24.2 ($p < .01$) in the validation set (unpublished data). In 2004 we reassessed the accuracy of APACHE III version I using patient data collected between 2002 and 2003. The SMR was 0.93 and the Hosmer-Lemeshow statistic = 273.6 ($p < .001$). These results indicated that this equation needed to be remodeled.

We had a different experience in assessing the calibration of the equation that predicts hospital mortality for patients admitted after coronary artery bypass graft (CABG) surgery. The original APACHE III model for predicting hospital mortality after CABG surgery (23) was poorly calibrated when tested using 1993–1996 patient data and was therefore remodeled using data for patients admitted between 1993 and 1996. The SMR was 1.00 and the Hosmer-Lemeshow statistic = 11.0 ($p > .10$) for the complete data set (unpublished data). When this equation was revalidated in 2004 using data collected during 2002 and 2003, the SMR was 0.997.

In this report, we describe APACHE IV, an improved and updated model for predicting group hospital mortality among critically ill patients and the previously remodeled equation for predicting hospital mortality after CABG surgery. Instead of developing a new APACHE IV score, our remodeling efforts were built on the success of severity scoring using the physiologic variables and weights of APACHE III. Our efforts, however, went beyond fitting a new logistic regression equation and developing new coefficients for each variable in a new ICU patient sample. Instead, using recent patient data we developed new predictor variables and used refined statistical methods to develop and validate an improved predictive model that is available in the public domain.

METHODS

The data for this study consisted of a nonrandomized observational cohort of 131,618 consecutive ICU admissions. Data were collected for patients admitted between January 1, 2002, and December 31, 2003, at 104 intensive or coronary care units in 45 hospitals. The 104 units were selected because each had installed an APACHE III computerized data collection and analysis system. Each hospital provided information about the type of ICU where data were collected, hospital bed size, geographic location, and teaching status. Patient data were entered on site using a software program that included computerized pick lists, automated error checks, and calculation of physiologic means and gradients. Some units entered data via electronic interfaces with laboratory and clinical information systems. The procedures used for data collection were based on prior reliability studies (1, 24) and field experience (16). Our methods for training data collectors and for ensuring accuracy have been previously described (22).

Patient Information and Exclusions. Patient data generated as a result of patient care and recorded in the medical record were collected concurrently or in some cases retrospectively for consecutive unselected intensive or coronary care unit admissions. Informed consent was not obtained because of Institutional Review Board waivers during prior studies (1). We excluded and did not collect data for patients who had been admitted for <4 hrs, patients with burns, patients <16 yrs of age, and except for hepatic and renal transplantation, patients admitted after transplant operations. We excluded patients missing an acute physiology score on ICU day 1 and those remaining in hospital for >365 days. To avoid counting more than one hospital outcome for the same patient, analysis included only a patient's first ICU admission. We also excluded from analysis patients admitted from another ICU during the same hospitalization. We did this because extensive life support before ICU admission biases the prognostic implications of first ICU day physiologic measures. Outcomes recorded for each patient included the exact duration (in hours) of ICU and hospital length of stay and mortality at ICU and hospital discharge.

The data collected for each patient are summarized in Table 1. The data included age, chronic health conditions, and physiologic data required to calculate an acute physiology score (APS) of APACHE III. The APS was based on the worst measurement for each component on ICU day 1. Each of the three variables rather than the composite APACHE III score was used in the mortality model. We also re-

Table 1. Data items collected and used for predicting hospital mortality among patients admitted to intensive care unit (ICU) who did not have coronary artery bypass graft (CABG) surgery

Age	Continuous Measure Plus Five Spline Terms
APS variables	Weight determined by most abnormal value within first APACHE day; sum of weights equals the APS, which ranges from 0 to 252. Five spline terms added. Variables include pulse rate, mean blood pressure, temperature, respiratory rate, Pao_2/Fio_2 ratio (or $P(A\text{-}a)o_2$ for intubated patients with $Fio_2 \geq 0.5$), hematocrit, white blood cell count, creatinine, urine output, blood urea nitrogen, sodium, albumin, bilirubin, glucose, acid base abnormalities, and neurological abnormalities based on Glasgow Coma Score
Chronic health variables	AIDS, cirrhosis, hepatic failure, immunosupression, lymphoma, leukemia or myeloma, metastatic tumor. Not used for elective surgery patients
ICU admission diagnosis	116 categories (see Appendix Tables 1 and 2)
ICU admission source	Floor, emergency room, operating/recovery room, stepdown unit, direct admission, other ICU, other hospital, other admission source
Length of stay before ICU admission	Square root plus four spline terms
Emergency surgery	Y/N
Unable to assess Glasgow Coma Scale score	Y/N
Thrombolytic therapy	For patients with acute myocardial infarction (Y/N)
Glasgow Coma Scale score rescaled	15 minus measured Glasgow Coma Scale score
Pao_2/Fio_2 ratio	
Mechanical ventilation	Y/N

APS, acute physiology score; APACHE, Acute Physiology and Chronic Health Evaluation.
All predictor variables use ICU day 1 data.

corded the primary reason for ICU admission (ICU admission diagnosis), the patient's location (admission source), length of stay before ICU admission, whether a patient received mechanical ventilation or underwent emergency surgery, Pao_2/Fio_2 ratio, and whether sedation or paralysis resulted in an inability to assess GCS score.

Each patient's primary reason for ICU admission (ICU admission diagnosis) was recorded by selecting one of 430 diseases, injuries, surgical procedures, or events that was most immediately threatening to the patient and required the services of the intensive or coronary care unit. Each ICU admission diagnosis was first classified as nonoperative or postoperative, next by body system or a transplant or trauma-related category, and then by diagnosis. A residual "other" category was used for unlisted diagnoses within each body system, transplant, and trauma category. For patients with acute myocardial infarction we recorded infarct location (anterior, inferior, lateral, non-Q-wave, posterior, other, unknown) and whether the patient received thrombolytic therapy within 24 hrs pre- or post-ICU admission.

Data collected for patients admitted after CABG surgery are summarized in Table 2. Measurements that were unique for patients admitted after CABG surgery included female gender, number of grafts, whether an internal mammary graft was used, and whether the patient had diabetes, prior CABG surgery, or myocardial infarction during the current hospitalization. A detailed description of these demographic, clinical, and physiologic variables has been previously reported (1, 21, 22) and is available along with instructions for their calculation and use in predicting hospital mortality at www.criticaloutcomes.cerner.com.

Development of the APACHE IV Model for Predicting Hospital Mortality. For each patient we estimated the probability of hospital mortality using a multivariate logistic regression procedure. The predictor variables used by the APACHE IV model for patients who did not have CABG surgery are shown in Table 1. The age, APS, and prior length of stay variables were allowed to have nonlinear relationships with outcomes using restricted cubic regression splines (25). Splines allow estimation of a nonlinear relationship between a variable and an outcome and replace less accurate techniques that assume the relationship is linear. Cut points (knots) are chosen and a separate coefficient is included for each interval between knots. A restricted cubic spline transformation (26) was used to expand age and the APS to five nonlinear terms and previous length of stay to four nonlinear terms. The prior length of stay variable was measured as a continuous rather than an integer variable as was done in APACHE III. Among the items listed in Table 1, four were tested as new predictor variables: a) whether a patient was mechanically ventilated; b) whether a patient with acute myocardial infarction received thrombolytic therapy; c) an adjustment for the differing prognostic implications of the GCS and Pao_2/Fio_2; and d) the impact of inability to assess GCS due to sedation or paralysis. The later variable was tested as a means for reducing predictive inaccuracies caused by defaulting GCS to normal when assessment was not possible (8).

We examined the 430 ICU admission diagnoses to determine whether there were a sufficient number of patients and deaths to enable us to create a greater number of specific diagnostic categories than included in APACHE III. New categories were selected based on their frequency, their clinical homogeneity, and the impact of each diagnosis on hospital mortality. Unlisted and infrequent diagnoses were placed into residual body system related "other" categories. This resulted in an increase in the number of diagnostic categories from 78 in the 1991 APACHE III, version H model (21, 22) to 116 in APACHE IV. A list of the 430 ICU admission diagnoses and the 116 mapped categories used for APACHE IV predictions are available at www.criticaloutcomes.cerner.com.

Table 2 displays the variables used in a multivariate logistic regression procedure for estimating the probability of hospital mortality for patients admitted after CABG surgery. Among the ten predictor variables shown, eight have been previously reported (23) and two, length of stay before ICU admission and whether a patient had diabetes, were added during remodeling in 1998. This analysis also used a restricted cubic spline transformation similar to that used for non-CABG patients.

Analysis. The APACHE IV hospital mortality equation was estimated using a randomly selected patient group that comprised 60% of the database (training set). We then compared observed and predicted hospital mortality for individual patients within the excluded group (validation set): The default categories with a zero coefficient (reference patient) had the following characteristics. Admission diagnosis of acute myocardial infarction (other), no emergency surgery, admitted to ICU directly or from emergency room or a stepdown unit, no chronic health item, APS = 0, able to have GCS measured, GCS of 15, and not on a ventilator. All calculations were performed using SAS version 9.1 (SAS Institute, Cary, NC) and S-Plus 6.2 (Insightful Corporation, Seattle, WA). To compare observed and predicted mortality for all ICU admissions and across patient subgroups, we calculated an SMR by dividing observed by mean predicted hospital mortality.

To assess the accuracy of APACHE IV hospital mortality predictions within the validation set, we used multiple analytic methods. To test the ability of the model to distinguish patients who die from patients who live (discrimination), we used the area under a receiver operating characteristic curve (AU-ROC) (27). Values above 0.80 indicate good discrimination. To test the degree of correspondence between observed and predicted mortality over the entire range of risk (calibration), we used three methods. First, we graphically displayed calibration by plotting observed and predicted mortality across all risk ranges (28). Second, we used goodness-of-fit testing to evaluate calibration across deciles of risk using the Hosmer-Lemeshow C statistic (29). Third, we used a Cox chi-square test to evaluate the equivalence of observed and predicted hospital mortality in aggregate and within subgroups defined by each predictor variable. A significance criterion of $p < .01$ was chosen because of the large sample size and the large number of statistical tests across multiple subgroups. Odds ratios were computed by taking $\exp^X$, where X is the value of a variable's coefficient in the model. The contribution of each risk component was based on its reduction to the full model's log likelihood. To do this we assessed the difference in log likelihood of the intercept-only model and the full model, which is the explainable variation. The difference in log likelihood for models excluding each risk component was built and compared vs. the full model. The difference was then divided by the explainable variation. We have compared observed and predicted hospital mortality across ICUs but will report these results in a future analysis.

To assess the impact of excess length of hospital stay on the mortality model, we developed an equation that included a term that was used in previous APACHE mortality mod-

Table 2. Data items collected and used for predicting hospital mortality among patients admitted to intensive care unit (ICU) after coronary artery bypass graft (CABG) surgery

Age; yrs	0–45, 46–55, 56–60, 61–65, 66–70, 71–75, 76–85, 85+
Acute physiology score variables	Four spline terms added (described in Table 1).
Emergency surgery	Y/N
Prior CABG surgery	Y/N
Female gender	Y/N
Number of grafts	Y/N for six variables for 2, 3, 4, 5, 6, and 7 grafts, respectively.
Internal mammary artery graft	Y/N
Myocardial infarction during current hospitalization	Y/N
Length of stay before ICU admission	Square root
Diabetes	Y/N

All predictor variables use ICU day 1 data.

els (21, 22). This term was derived by calculating the expected hospital length of stay for survivors, calculating the residual (observed − expected) for each individual, and then obtaining the mean residual for each ICU. This variable was used to address institutional factors that keep patients in the hospital longer and thus increase their likelihood of hospital death. The accuracy of mortality prediction using the excess length of stay term was compared with the model not including the added terms by comparing −2 log likelihood for each model.

Mortality Predictions Over Time. To evaluate the need to reassess and update models for predicting hospital mortality, we applied the APACHE IV model and two earlier versions of APACHE III to the current (2002–2003) validation data set. The two earlier versions were a) APACHE III, version H, which was developed and validated using patient data from 1988 to 1989 and included statistically independent equations for 78 disease categories (1, 21); and b) APACHE III, version I, which was developed and validated using patient data from 1993 to 1996 and included statistically independent equations for 94 disease categories (unpublished data). For each model we used the 2002–2003 validation set to assess aggregate discrimination using the AUROC, calibration using a Cox chi-square test to evaluate the equivalence of observed and predicted hospital mortality, and goodness-of-fit using the Hosmer-Lemeshow C statistic (29). We used the same models and validation set to assess the equivalence of observed and predicted mortality across disease subgroups that were selected according to clinical importance, frequency, and uniformity of definition.

RESULTS

Characteristics of Hospitals and Critical Care Units. Of the 45 hospitals studied, 15 (33%) were members of the Council of Teaching Hospitals, 13 (29%) were teaching hospitals that were not members of the Council of Teaching Hospitals, and 17 (38%) were nonteaching hospitals. The mean number of hospital beds was 478 (range 50–1,030 beds). Among the 45 hospitals, 11 (24%) had <300 beds, 11 (24%) had 300–399 beds, eight (18%) had 400–524 beds, seven (16%) had 525–799 beds, and eight (18%) had ≥800 beds. Geographically, 17 (38%) were located in the Southeast, 14 (31%) in the West, nine (20%) in the Midwest, and five (11%) in the Northeast. Among the 104 units that participated in the study, 40 were mixed medical-surgical, 17 coronary care, 14 surgical, 12 cardiothoracic, 11 medical, seven neurologic, and three trauma.

Exclusions and Patient Characteristics. To develop and validate the APACHE IV model for predicting hospital mortality among patients without CABG surgery, 21,137 patients were excluded for the following reasons: a) missing APS on day 1 (n = 450); b) remaining in the hospital >365 days (n = 112); c) second or multiple ICU admissions (n = 8,735); d) admission from another ICU (n = 2,760); and d) undergoing CABG surgery (n = 9,180). These exclusions left 110,558 patients for model development and validation.

The demographic and clinical characteristics of the 110,558 patients are shown in Table 3, and Figure 1 shows the distribution of first ICU day acute physiology scores. Appendix Table 1 lists the 75 nonoperative and Appendix Table 2 the 41 postoperative ICU admission diagnoses used by the APACHE IV model. In aggregate, the 12 most frequent ICU admission diagnoses accounted for 29.8% of all admissions. The 13 residual "other" categories (eight nonoperative and five postoperative) accounted for 12.2% of nonoperative and 17.8% of postoperative admissions.

The APACHE IV Model for Predicting Hospital Mortality. Variables included in the predictive model are shown in Table 1. The variables include the APS plus additional spline terms, age plus additional spline terms, square root of the previous length of stay plus additional spline terms, emergency surgery, whether a patient was ventilated on ICU day 1, a rescaled GCS score, inability to assess GCS due to sedation or paralysis, thrombolytic therapy for patients with acute myocardial infarction, seven chronic health items, three variables for location before ICU admission, Pao_2/Fio_2 ratio, and 115 categorical variables for admission diagnoses ("acute myocardial infarction, other" was the default category with no separate coefficient). Thus a total of 142 variables were included in the model resulting in an average of 42 deaths per variable in the validation data set. This indicates that the full model did not contain too many variables relative to the number of outcomes.

Analysis of Aggregate Discrimination and Calibration. The results from applying the APACHE IV model for non-CABG surgery patients to the validation data set are given in Table 4. The observed and mean predicted mortality were 13.51% and 13.55%, respectively, for an SMR = 0.997 ($p = .76$). AU-ROC was 0.88 indicating that the model had excellent discrimination. The Hosmer-Lemeshow chi-square was 16.8 (10 *df*, $p = .08$), which is not significant (on the validation set 10 *df* was used; see Ref. 30, page 188). The model had an adjusted $R^2 = .44$, indicating that it accounted for 44% of the variation in outcomes. The relative explanatory power for the predictor variables is shown in Figure 2. The most important variable was the APS, followed by disease group and age. Table 5 shows the observed and mean predicted mortality rates by risk deciles. The largest difference by decile was 1.6%, which occurred in the >90–100% decile. Figure 3 graphically displays the close relationship between observed and mean predicted mortality across risk deciles.

Analysis of Predictive Accuracy Within Patient Subgroups. Because splined variables generate predictions based on nonlinear terms, their contribution as APACHE IV predictor variables is best viewed by examining risk over ranges of the variables. The coefficients and mortality probabilities for differing

Table 3. Demographic and clinical characteristics of 110,558 patients admitted to 104 U.S. intensive care units between January 1, 2002, and December 31, 2003

	Training Data Set (n = 66,270)	Validation Data Set (n = 44,288)
Age, yrs	61.51 ± 0.07	61.45 ± 0.08
Acute physiology score	38.83 ± 0.10	38.72 ± 0.12
Prior length of stay, square root days	0.786 ± 0.004	0.777 ± 0.005
Pao_2/Fio_2 ratio	332.70 ± 0.41	332.46 ± 0.50
Died in hospital, %	13.6	13.5
One or more APACHE comorbidities, %	10.4	10.6
Emergency surgery	5.7	5.5
Unable to assess GCS due to sedation, %	8.0	8.1
Ventilated on day 1, %	35.1	35.1
Postoperative patient, %	30.9	30.8
Gender, % male	54.2	54.2
Race, % white	69.3	69.5

APACHE, Acute Physiology and Chronic Health Evaluation; GCS, Glasgow Coma Score.

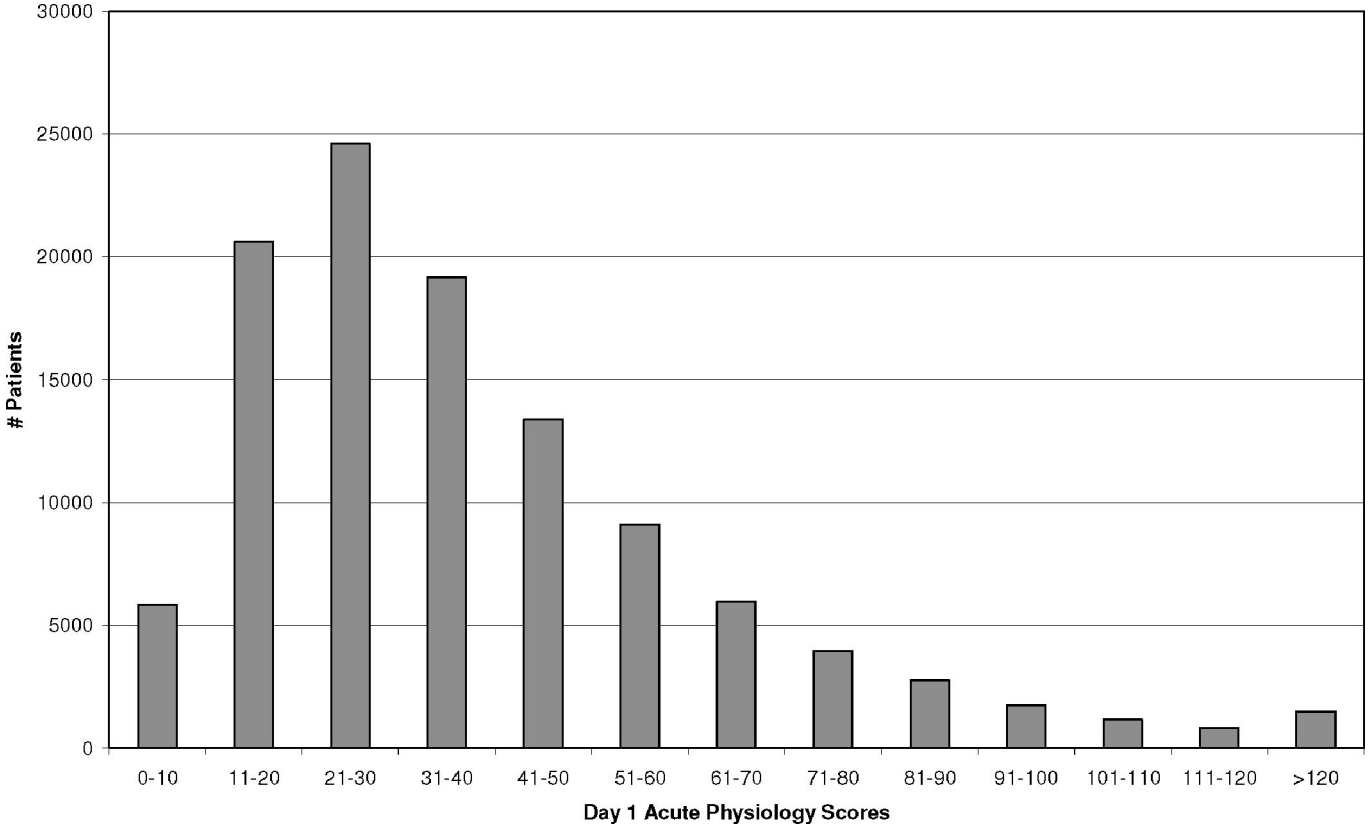

Figure 1. Distribution of intensive care unit day 1 acute physiology score.

Table 4. Comparison of discrimination and calibration of the Acute Physiology and Chronic Health Evaluation (APACHE) IV mortality model and earlier APACHE III versions when applied to the same 2002–2003 validation data set (n = 44,288)

Version	APACHE IV[a]	APACHE III—Version I[b]	APACHE III—Version H[c]
Observed mortality rate, %	13.51	13.51	13.51
Predicted mortality rate, %	13.55	14.64	16.90
SMR, observed/predicted	0.997 ($p = .79$)	0.923 ($p < .001$)	0.799 ($p < .001$)
Area under the ROC curve	0.880	0.870	0.868
Hosmer-Lemeshow χ^2	16.8 ($p = .08$)	124.6 ($p < .001$)	635.4 ($p < .001$)

ROC, receiver operating characteristic curve.

[a]APACHE IV was validated using data for 44,288 admissions to 104 intensive care units during 2002 and 2003; [b]APACHE III version I was developed using data for 40,264 admissions to 188 intensive care units during 1993 and 1996 (unpublished data); [c]APACHE III version H was developed using data for 16,662 admissions to 42 intensive care units during 1988 and 1989 (Refs. 2, 23).

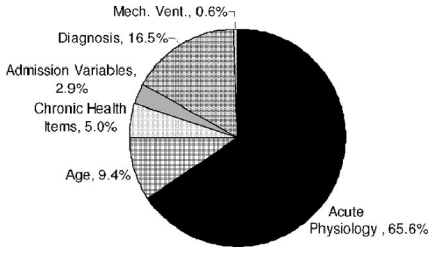

Figure 2. Unique relative contribution of each risk factor to hospital mortality prediction. The acute physiology category includes the acute physiology score of Acute Physiology and Chronic Health Evaluation III, inability to assess Glasgow Coma Scale, and rescaled Glasgow Coma Scale and Pao_2/Fio_2 ratio. The admission variables include patient admission source and length of stay before intensive care unit admission. The diagnosis category includes 116 mutually exclusive primary reasons for intensive care unit admission, emergency surgery, and thrombolytic therapy for patients with acute myocardial infarction.

values of the APS, for age, and for previous length of stay are shown in Table 6. The coefficients for each of these spline terms are automatically calculated at www.criticaloutcomes.cerner.com. The influence of the APS was more pronounced than that for age, and the latter was more influential than previous length of stay. Figure 4 shows how risk increased with increasing APS within four diagnostic groups.

Among the 116 ICU admission diagnoses used in the APACHE IV model, 29 (25%) had significant coefficients and 54 (47%) had p values <.10. The diagnoses for non-CABG surgery patients and the observed and mean predicted mortality, SMR, regression coefficient, and odds ratio for each diagnosis in the validation set are shown in Appendix Tables 1 and 2. Among the 116 ICU admission diagnoses, there was no significant difference ($p < .01$) between observed and mean predicted hospital mortality for 65 (87%) of the 75 nonoperative diagnoses and 40 (98%) of the 41 postoperative diagnoses. Overall, there was no significant difference ($p < .01$) between observed and mean predicted mortality for 105 (91%) of the 116 diagnostic groups. Of the 11 significantly different diagnostic groups, seven had SMRs <1.00 and four had SMRs >1.00. For the residual "other" diagnostic categories listed in Appendix Tables 1 and 2, there

Table 5. Observed and predicted hospital mortality rates across risk deciles within the validation data set (n = 44,288)

Risk Decile[a]	Observed Deaths No. (%)	Predicted Deaths No. (%)	Difference, %
1	24 (0.5)	21 (0.5)	0.1
2	30 (0.7)	44 (1.0)	−0.3
3	70 (1.6)	73 (1.6)	−0.1
4	100 (2.3)	113 (2.6)	−0.3
5	182 (4.1)	172 (3.9)	0.2
6	264 (6.0)	263 (5.9)	0.0
7	420 (9.5)	417 (9.4)	0.1
8	751 (17.0)	700 (15.8)	1.1
9	1,331 (30.1)	1,314 (29.7)	0.4
10	2,813 (63.5)	2,883 (65.1)	−1.6

[a]Risk decile: population sorted by increasing predicted risk and then split into deciles. Sum of (observed − expected)^2/expected = χ^2 = 16.8, df = 10, p = .08.

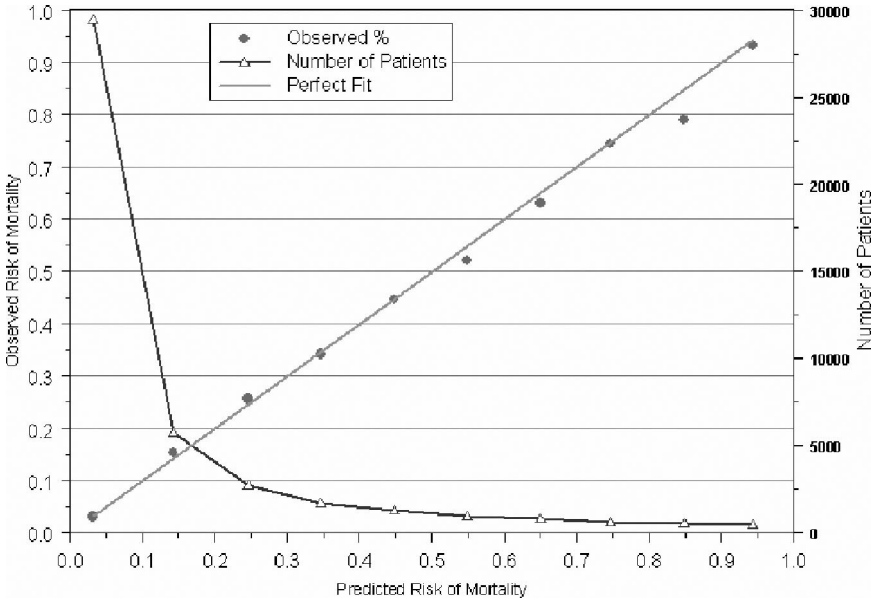

Figure 3. Calibration curve comparing observed and predicted hospital mortality rates across 10% intervals of predicted risk and distribution of risk for the 44,288 intensive care unit admissions in the validation data set. The 45° line indicates perfect predictive ability. The line connecting triangles plots the number of patients in each risk group.

was no significant difference between observed and mean predicted hospital mortality for 12 (92%) of the 13 categories.

Appendix Table 3 shows observed and mean predicted mortality, SMR, odds ratio, and the coefficients for each nonsplined variable. The odds ratios for patients having emergency surgery, unable to have the GCS assessed, and ventilated on day 1 were all significantly greater than 1.00. The observed and predicted mortality rates for these patients were within 1%. The odds ratio for patients with acute myocardial infarction who received thrombolytic therapy was 0.56. However, the SMR for this group was 1.54, which indicates poor calibration among these patients. The odds ratio for rescaled GCS was 1.04 for each 1-point increase in score, apart from the increase in the APS that would also occur. Thus a patient with a rescaled GCS = 12 (measured GCS = 3) had an odds ratio of 1.60 compared with a patient with a rescaled GCS = 0 (measured GCS = 15). Increasing Pao_2/Fio_2 ratios decreased the mortality probability slightly. Except for high values of Pao_2/Fio_2, the model's predictions were close to the observed mortality. Each of the chronic health items had significant odds ratios, which ranged from 1.55 to 2.96. Except for patients with lymphoma, the SMRs by chronic health item were close to 1.0. The various admission sources had a nonsignificant odds ratio, except for admission from the operating/recovery room, which had an odds ratio of 0.56. The SMR for each admission sources was close to 1.0.

Each nonsplined APACHE IV predictor variable and its regression coefficient are also available at www.criticaloutcomes.cerner.com. This Web site also provides the ability to automatically calculate spline terms for APS, age, and previous length of stay and to apply both splined and nonsplined terms to arrive at an individual's prediction of mortality.

Impact of Excess Hospital Length of Stay on Mortality Prediction. When the variable reflecting excess length of hospital stay was added to the APACHE IV model, it was highly significant (likelihood ratio chi-square = 73.8, $p < .001$). When compared with the model without that term, there was a decline in goodness-of-fit (Hosmer-Lemeshow chi-square went from 16.8 to 25.9) but no change in discrimination (AUROC = 0.88 for both models) or SMR (0.997 for both models). Because the excess length of hospital stay variable did not improve model performance and is logistically difficult to calculate, it was not included in the APACHE IV model.

Mortality Prediction for Patients Admitted After CABG Surgery. The original model for predicting hospital mortality after CABG surgery (23) was remodeled in 1998 using data for 3,689 patients admitted during 1993 through 1996. The remodeled equation (APACHE III–CABG, version I) was revalidated using our current data for 9,180 patients admitted during 2002 and 2003. Observed mortality was 2.16% and predicted mortality was 2.16% for an SMR of 0.997 (chi-square = 0.002, p = .96). Given the low mortality rate and excellent SMR, we did not remodel this equation and have included it as a part of APACHE IV.

APACHE Mortality Predictions Over Time. Aggregate hospital mortality predictions using APACHE IV and APACHE III, versions H and I, in the current validation data set are shown in Table 4. Discrimination did not decline for the three models, but calibration as measured by the SMR and Hosmer-Lemeshow statistic declined for version I and deteriorated further for version H. Table 7 shows how applying the three models affected the SMR for diagnostic groups chosen because their frequency and definitions were constant across versions. APACHE IV performed significantly better than APACHE III version I for patients with sepsis (nonurinary tract), cardiac arrest, and noncardiac pulmonary edema (acute respiratory distress syn-

Table 6. Mortality probabilities for various values of the acute physiology score (APS),[a] age,[b] and length of stay[c] given default or mean values for all other variables[d]

	Mortality Probability, %	95% Confidence Interval	
		Lower	Upper
Acute physiology score			
20	2.7	2.0	3.8
30	4.9	3.6	6.6
40	7.8	5.8	10.6
50	11.5	8.5	15.2
60	16.1	12.1	21.0
70	21.8	16.7	28.0
80	28.8	22.4	36.0
90	36.7	29.3	44.8
100	45.5	37.2	54.0
Age, yrs			
30	4.1	2.9	5.6
45	5.6	4.1	7.7
60	7.2	5.3	9.7
70	9.8	7.2	13.0
80	13.2	9.9	17.4
90	16.4	12.3	21.5
Previous length of stay			
0	7.1	5.2	9.7
4 hrs	6.5	4.8	8.8
12 hrs	7.2	5.3	9.7
1 day	8.2	6.0	11.0
2 days	9.4	6.9	12.6
4 days	10.5	7.8	14.2

[a]The coefficients for APS and the spline terms were 0.556, 0.000008719, −0.00005, 0.00005, −0.0000131, and −0.000000865. The last five coefficients emanate from using knots at APS = 10, 22, 32, 48, 89, respectively; [b]the coefficients for age and the spline terms were 0.0242, −0.00000439, 0.00005, −0.00013, 0.000109606, and −0.000027572. The last five coefficients emanate from using knots at age = 27, 51, 64, 74, 86, respectively; [c]the coefficients for previous length of stay and the spline terms were −0.3105, 1.4747, −2.8619, 1.4216590103, and −0.034445822. The last four coefficients emanate from using knots at previous length of stay = 3 hrs, 10 hrs, 19 hrs, and 2.8 days, respectively; [d]default disease group is acute myocardial infarction (AMI) other, thrombolytic therapy = no Glasgow Coma Scale (GCS) = 15, Pao_2/Fio_2 = 385.7, no chronic health items, direct admission, non-ventilated, unable = no, and emergency surgery = no.

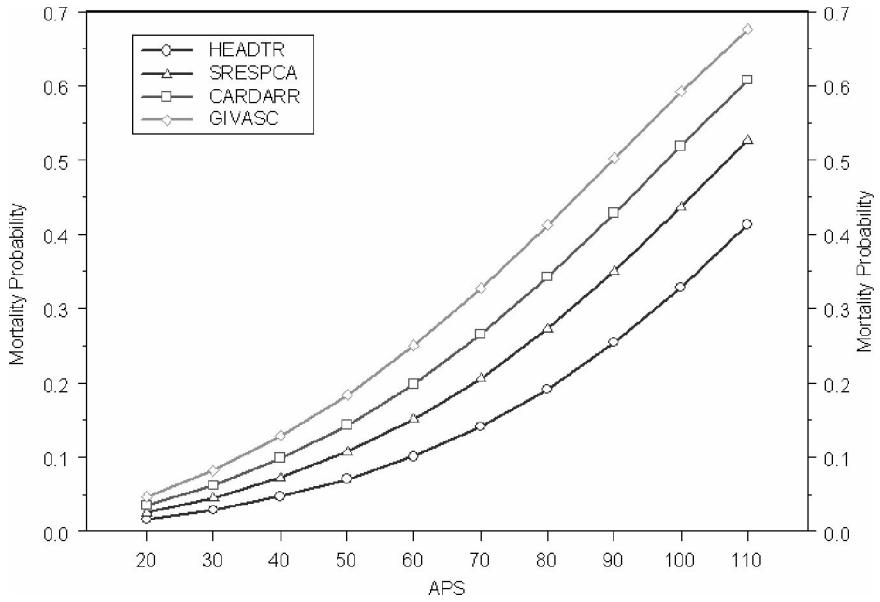

Figure 4. Relationship between first intensive care unit day acute physiology score (APS) and predicted mortality for four disease groups. Disease groups are indicated as follows: *circles*, head trauma (*HEADTR*); *triangles*, surgery for respiratory cancer (*SRESPCA*); *squares*, cardiac arrest (*CARDARR*); *diamonds*, gastrointestinal vascular insufficiency (*GIVASC*).

drome) and significantly better than APACHE III version H for patients with the same diagnoses and for patients with emphysema/bronchitis and hepatic failure.

DISCUSSION

APACHE IV was developed because the accuracy of APACHE III changed significantly over the last decade. The APACHE IV model has excellent discrimination (AUROC = 0.88), and aggregate predicted hospital mortality (13.51%) is statistically identical to observed 13.55% ($p = .76$) for the validation data set. Calibration is excellent with a nonsignificant Hosmer-Lemeshow C statistic (16.8, $p = .08$) despite the large validation sample size. The difference between observed and mean predicted hospital mortality across risk deciles is 0.1–0.4% except for the 70–80% decile (1.1%) and 90–100% decile (1.6%). For most subgroups, the ratio of observed to mean predicted hospital mortality is near 1.0, and 90% of the SMRs within disease groups are not significantly different from 1.0. Although a smaller sample size for some disease groups caused a low power of detection, the relative magnitude of observed mortality correlated well with that of predicted risk. Because of its accuracy for U.S. patients, APACHE IV can be used to benchmark ICU performance using aggregate SMRs to assess quality of care and disease-specific SMRs to evaluate outcomes for patient subgroups. To enhance this application, APACHE IV has been placed in the public domain via a Web site spreadsheet that provides the ability to predict individual patient mortality.

Several factors are likely to account for the accuracy of APACHE IV mortality predictions. First, APACHE IV is based on the successful use of physiologic abnormalities for risk adjustment, which accounted for 66% of the model's explanatory power. Because the APS was splined a simple odds ratio is not possible, but as shown in Table 6 and Figure 3, the predicted mortality consistently increases with an increasing APS. Second, the accuracy of physiologic risk adjustment was improved by adding rescaled Pao_2/Fio_2 and GCS variables and by reducing the impact of defaulting the GCS to a normal value when sedation or paralysis made direct assessment impossible. Third, we improved case-mix adjustment by increasing the precision of disease labeling. ICU admission diagnosis accounts for 17% of the model's explanatory power,

Table 7. Standardized mortality ratio for selected disease groups when Acute Physiology and Chronic Health Evaluation (APACHE) IV, APACHE III version I, and APACHE III version H predictions are used for the 2002–2003 validation data set (n = 44,288)

			APACHE IV[a]		APACHE III Version I[b]		APACHE III Version H[c]	
Disease Group	No.	Observed Mortality, %	Predicted Mortality, %	SMR	Predicted Mortality, %	SMR	Predicted Mortality, %	SMR
Sepsis (nonurinary tract)	1,821	37.3	37.4	1.00	41.8	0.89[d]	45.2	0.83[d]
Cardiac arrest	872	58.3	58.4	1.00	53.1	1.10[d]	54.5	1.07[e]
Emphysema/bronchitis	878	15.1	13.4	1.13	17.4	0.87	19.8	0.76[d]
Noncardiac pulmonary edema (ARDS)	310	27.7	28.2	0.98	36.3	0.76[d]	34.0	0.82[e]
Thoractomy for lung neoplasm	633	4.1	4.3	0.96	3.5	1.16	5.2	0.80
Aortic aneurysm, elective repair	701	5.6	4.7	1.19	3.9	1.41	4.6	1.20
Stroke	860	21.5	20.2	1.06	19.8	1.09	22.6	0.95
Hepatic failure	236	45.8	41.4	1.11	47.4	0.97	59.3	0.77[d]
Respiratory arrest	490	34.1	32.2	1.06	35.3	0.97	37.4	0.91

SMR, standardized mortality ratio; ARDS, acute respiratory distress syndrome.

[a]APACHE IV was validated using data for 44,288 admissions to 104 intensive care units during 2002 and 2003; [b]APACHE III version I was developed using data for 40,264 admissions to 188 intensive care units during 1993 and 1996 (unpublished data); [c]APACHE III version H was developed using data for 1,662 admissions to 42 intensive care units during 1988 and 1989 (Refs. 2, 23); [d]$p < .001$; [e]$p < .01$.

and we believe that expanding the number of diagnostic coefficients to 116 was a major factor in improving predictive accuracy. Fourth, we used advanced statistical methods, particularly the expanded use of splines for age (9% of explanatory power), APS, and prior length of stay variables. Finally, we continued to adjust for the prognostic impact of patient location before ICU admission (11, 12) and incorporated new variables based on data availability and published information about their independent prognostic impact.

The same factors that account for the accuracy of APACHE IV predictions also contribute to its complexity. There are 142 variables in the mortality equation, although most (115) are disease groups. In all nine sets of variables are measured age, APS, chronic health comorbidities, previous length of stay, ventilator status, thrombolytic therapy for patients with acute myocardial infarction, emergency surgery, admission source, and ICU admission diagnosis. Two additional variables, unable to assess GCS and Pao_2/Fio_2 ratio, are assessed during the recording of components of the APS. The heaviest data burden involves collecting the 16 measurements that make up the APS.

We believe the complexity of APACHE IV is best addressed by excellent training and information technology. To ensure thorough training for data collection, there is a Web-based training manual at www.criticaloutcomes.cerner.com. Automated collection of APS variables provides one example of how information technology can reduce data collection effort and improve data reliability (30, 31). Laboratory data are captured electronically, worst values identified, and derived physiologic variables calculated. Each spline term is calculated automatically, and regression coefficients for all splined and nonsplined variables are used to automatically calculate both individual and group mortality predictions. Data collection effort is also reduced and reliability enhanced by the use of computerized pick lists. For example, a "pick list" of 430 ICU admission diagnoses uses a hierarchy of operative status and body system to simplify selection of an ICU admission diagnosis. APACHE IV is also available at www.criticaloutcomes.cerner.com. This Web site supports manual data entry, automatically calculates spline terms, provides regression coefficients, and calculates predicted mortality.

Our analysis of predictive accuracy over time showed that estimates of aggregate hospital mortality deteriorated progressively for older APACHE III versions. Aggregate hospital mortality was overpredicted, SMR fell, and the Hosmer-Lemeshow statistic deteriorated (Table 5). This systematic overestimation of mortality has been reported in other U.S. studies (5, 32) and called "grade inflation" (33). It is overly simplistic, however, to attribute overestimates of mortality to aggregate improvements in ICU therapy. Reductions in mortality from critical illness are typically related to treatment advances that are disease specific. These advances have included new drugs (e.g., drotrecogin alpha [activated]), new technologies (e.g., noninvasive positive pressure ventilation), or new techniques (e.g., low tidal volume ventilation, goal-directed hemodynamic support in sepsis). We speculate that these advances might account for the improved hospital survival and significant overestimation of mortality by APACHE III for patients with sepsis, emphysema or bronchitis, and noncardiac pulmonary edema shown in Table 7. In contrast, for patients admitted for cardiac arrest, hospital survival declined and mortality was progressively underestimated by APACHE III. These changes might be related to recent changes in end-of-life care (34, 35). These findings, together with the marked variations in mortality for specific diagnoses within body system categories (Appendix Tables 1 and 2), support the importance of more precise adjustment for ICU admission diagnosis in prognostic models.

In developing APACHE IV we used published information to improve prognostic accuracy. A consistent policy of recording inability to assess neurologic status due to sedation or paralysis (4, 36) allowed us to reduce the predictive inaccuracies caused by defaulting the GCS to normal values (8). Including an "unable to assess GCS" variable had a significant impact on mortality (odds ratio = 2.19), whereas observed vs. predicted mortality ratios by GCS interval remained in good agreement. We also tested the use of mechanical ventilation (3, 37, 38) and thrombolytic therapy for pa-

tients with acute myocardial infarction (39, 40) as model variables. The increased accuracy of mortality predictions within precisely defined disease categories is also supported by published reports. For example, predicted mortality by infection site in sepsis ranged from 19.4% for urinary tract to 42.3% for gastrointestinal. These differences in outcome by infection site differ in magnitude but are consistent with findings from the PROWESS study (41). Variation in APACHE IV-predicted mortality for patients with different types and locations of acute myocardial infarction is also similar to previous findings (42, 43).

The increased complexity of APACHE IV represents a continued departure from the simplification that characterized APACHE II (44) and other second-generation prognostic scoring systems. A recently developed automated risk adjustment system for Veterans Affairs ICUs (45–47) is similar to APACHE IV in its complexity and emphasis on information technology. We also know that the accuracy of older prognostic models such as APACHE II has deteriorated over time, and they lack predictor variables of proven prognostic significance (1, 21). For these reasons, APACHE II mortality predictions, even when recalibrated in large contemporary databases, are likely to be inaccurate due to the absence of multiple predictor variables. Based on the absence of critical variables and the limits of recalibration, we recommend that APACHE II no longer be used to compare observed and predicted mortality. We believe, however, that the APACHE II score continues to be a useful summary measure of severity of illness.

In the future, prognostic models are likely to become even more complex and dependent on information technology. They may require additional treatment variables (48–52), adjustment for treatment limitations (14, 53, 54), and further diagnostic precision. For example, in this study observed mortality for patients with neurologic infection reflected the average death rate for patients with brain abscess (25%), encephalitis (5%), and meningitis (11%). Predicted mortality (11.4%) was significantly different ($p < .01$) from observed (17.9%), but accuracy would likely improve if patient numbers were sufficient to develop coefficients for each type of neurologic infection.

The APACHE IV model is subject to several limitations. First, it is unlikely that accurate mortality predictions will be possible in other countries. APACHE IV was developed and validated only in U.S. ICUs. International differences in bed availability, ICU structure, patient referral, selection criteria, and care before and after ICU are likely to have an adverse impact on predictive accuracy (7, 9, 13, 55–57). Second, the use of data from ICUs that purchased an APACHE system represents a selection bias. Despite being collected in 45 hospitals with variations in bed size, teaching status, and geographical regions, our data might not be nationally representative. Third, although the sample size in this analysis was large, the results of the logistic regression analysis may have been influenced by the random assignment of patients to training or validation data sets. Although the actual values of the coefficients might change with a different sampling assignment, the small standard errors for the major variables and relatively narrow confidence intervals around the odds ratios suggest that uncertainty is not large. Fourth, prediction for an individual contains variance. For example, for three selected patients where the predicted probabilities were 5.0%, 20.0%, and 40.0% (data not shown), the 95% confidence intervals were (3.9%, 6.5%), (16.8%, 23.7%), and (36.3%, 43.8%), respectively. Thus, a prediction is only an approximate indicator of an individual's probability of mortality. Fifth, the use of aggregate SMR as an ICU performance benchmark is limited by factors that are not directly related to quality of care. These factors include the frequency of treatment limitations, early discharge to skilled nursing facilities, and care before and after ICU admission (14, 16). Sixth, we anticipate that the accuracy of APACHE IV will deteriorate in the future. New knowledge, new therapies, wider use of low tidal volume ventilation and drotrecogin alpha (activated), and changes in end-of-life care make it likely that the model will need to be revised and updated.

CONCLUSIONS

Prognostic estimates derived from predictive models become increasingly inaccurate as the time between their development, updating, and application increases. Mortality tends to be underestimated, but changes are not uniform and may be related to disease specific changes in therapy. Predictive models require periodic retesting, and when accuracy deteriorates they should be reestimated and variables with demonstrated prognostic significance tested for inclusion. APACHE IV predictions of hospital mortality have excellent discrimination and calibration and should be useful for benchmarking ICU performance in U.S. ICUs.

ACKNOWLEDGMENTS

We thank William A. Knaus, MD, for helpful discussions regarding this article. We also thank Douglas P. Wagner, PhD, for technical suggestions and discussions regarding this article.

REFERENCES

1. Knaus WA, Wagner DP, Draper EA, et al: The APACHE III prognostic system. *Chest* 1991; 100:1619–1636
2. Legall 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
3. Lemeshow S, Teres D, Klar J, et al: Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 270:2478–2486
4. Zimmerman JE, Draper EA, Wagner DP: Comparing ICU populations: Background and current methods. *In:* Evaluating Critical Care. Sibbald WJ, Bion JF (Eds). New York, Springer Verlag, 2001, pp 121–139
5. Glance LG, Osler TM, Dick AW: Identifying quality outliers in a large, multiple-institution database by using customized versions of the Simplified Acute Physiology Score II and the Mortality Probability Model II_0. *Crit Care Med* 2002; 30:1995–2002
6. Moreno Rui, Miranda DR, Fidler V, et al: Evaluation of two outcome prediction models on an independent database. *Crit Care Med* 1998; 26:50–61
7. Markgraf R, Deutschinoff G, Pientka L, et al: Comparison of Acute Physiology and Chronic Health Evaluation II and III and Simplified Acute Physiology Score II: A prospective cohort study evaluating these methods to predict outcome in a German interdisciplinary intensive care unit. *Crit Care Med* 2000; 28:26–33
8. Livingston BM, Mackenzie SJ, MacKirdy FN, et al: Should the pre-sedation Glasgow Coma Scale value be used when calculating Acute Physiology and Chronic Health Evaluation scores for sedated patients? *Crit Care Med* 2000; 28:389–394
9. Pappachan JV, Millar B, Bennett D, et al: Comparison of outcome from intensive care admission after adjustment for case mix by the APACHE III prognostic system. *Chest* 1999; 115:802–810
10. Apolone G, Bertolini G, D'Amico R, et al: The performance of SAPS II in a cohort of patients admitted to 99 Italian ICUs: Results from GiViTI. *Intensive Care Med* 1996; 22: 1386–1378
11. Durairaj L, Will JG, Torner JC, et al: Prognostic factors for mortality following interhospital transfers to the medical intensive

care unit of a tertiary referral center. *Crit Care Med* 2003; 31:1981–1986
12. Rosenberg AL, Hofer TP, Strachan C, et al: Accepting critically ill transfer patients: Adverse effect on a referral center's outcome and benchmark measures. *Ann Intern Med* 2003; 138:882–890
13. Goldhill DR, Sumner A: Outcome of intensive care patients in a group of British intensive care units. *Crit Care Med* 1998; 26: 1337–1345
14. Azoulay E, Pochard F, Garrouste-Orgeas M, et al: Decisions to forgo life-sustaining therapy in ICU patients independently predict hospital death. *Intensive Care Medicine* 2003; 29:1895–1901
15. Daly K, Beale R, Chang RWS: Reduction in mortality after inappropriate early discharge from intensive care unit: Logistic regression triage model. *BMJ* 2001; 322:1274–1378
16. Sirio CA, Shepardson LB, Rotondi AJ, et al: Community-wide assessment of intensive care outcomes using a physiologically based prognostic measure. *Chest* 1999; 115:793–801
17. Zhu BP, Lemeshow S, Hosmer DW, et al: Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: A simulation study. *Crit Care Med* 1996; 24: 57–63
18. Moreno R, Apolone G: Impact of different customization strategies in the performance of a general severity score. *Crit Care Med* 1997; 25:2001–2008
19. Metnitz PGH, Valentin A, Vesely H, et al: Prognostic performance and customization of the SAPS II: Results of a multicenter Austrian study. *Intensive Care Med* 1999; 25: 192–197
20. Markgraff R, Deutschinoff G, Pientka L, et al: Performance of the scoring systems Acute Physiology and Chronic Health Evaluation II and III at an interdisciplinary intensive care unit, after customization. *Crit Care* 2001; 5:31–36
21. 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
22. Zimmerman JE, Wagner DP, Draper EA, et al: Evaluation of Acute Physiology and Chronic Health Evaluation III predictions of hospital mortality in an independent database. *Crit Care Med* 1998; 26:1317–1326
23. Becker RB, Zimmerman JE, Knaus WA, et al: The use of APACHE III to evaluate ICU length of stay, resource use, and mortality after coronary artery bypass surgery. *J Cardiovasc Surg* 1995; 36:1–11
24. Damiano AM, Bergner M, Draper EA, et al: Reliability of a measure of severity of illness: Acute Physiology and Chronic Health Evaluation II. *J Clin Epidemiol* 1992; 45:93–101
25. Harrell FE: Regression Modeling Strategies. First Edition. New York, Springer-Verlag, 2001
26. Stone CJ, Koo CY: Additive splines in Statistics. *Proceedings of the Statistical Computing Section, American Statistical Association* 1985; 45–48
27. Hanley J, McNeil B: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36
28. Castella X, Artigas A, Bion J, et al: A comparison of severity of illness scoring systems for intensive care unit patients: Results of a multicenter, multinational study. *Crit Care Med* 1995; 23:1327–1335
29. Hosmer DW, Lemeshow S: Applied Logistic Regression. Second Edition. New York, Wiley, 2000
30. Chen LM, Martin CM, Morrison TL, et al: Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. *Crit Care Med* 1999; 27: 1999–2004
31. Bosman RJ, Oudemans van Straaten HM, Zandstra DF: The use of intensive care information alters outcome prediction. *Intensive Care Med* 1998; 24:953–958
32. Glance LG, Osler TM, Dick A: Rating the quality of intensive care units: Is it a function of the intensive care unit scoring system. *Crit Care Med* 2002; 30:1976–1982
33. Popovich MJ: If most intensive care units are graduating with honors, is it genuine quality or grade inflation? *Crit Care Med* 2002; 30: 2145–2146
34. Prendergast TJ, Luce JM: Increasing incidence of withholding and withdrawal of life support from the critically ill. *Am J Respir Crit Care Med* 1997; 155:15–20
35. Pendergast TJ, Classens MD, Luce JM: A national survey of end-of-life care for critically ill patients. *Am J Respir Crit Care Med* 1998; 158:1163–1167
36. Zimmerman JE, Knaus WA: Glasgow Coma Scale. *In:* Oxford Textbook of Critical Care. First Edition. Webb AR, Shapiro MJ, Singer M, et al. (Eds). New York, Oxford University Press, 1999, pp 350–351
37. Vincent JL, de Mendonca A, Cantraine F, et al: Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. *Crit Care Med* 1998; 26:1793–1800
38. Nierman DM, Schechter CB, Cannon LM, et al: Outcome prediction model for very elderly critically ill patients. *Crit Care Med* 2001; 29:1853–1859
39. The GUSTO Investigators: An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1993; 329:673–682
40. Ohman EM, Granger CB, Harrington RA, et al: Risk stratification and therapeutic decision making in acute coronary syndromes. *JAMA* 2000; 284:876–878
41. Ely EW, Laterre PF, Angus DC, et al: Drotrecogin alfa (activated) administration across clinically important subgroups of patients with severe sepsis. *Crit Care Med* 2003; 31:12–19
42. Haim M, Kornowski RL, Reicher-Reiss H, et al: Comparison of short and long term prognosis in patients with anterior wall versus inferior or lateral wall non-Q wave acute myocardial infarction. *Am J Cardiol* 1997; 79:717–721
43. Behar HM, Boyko V, Hod H, et al: The prognosis of a first Q wave versus non-Q wave myocardial infarction in the reperfusion era. *Am J Med* 2000; 108:381–386
44. Knaus WA, Draper EA, Wagner DP, et al: APACHE II—A severity of disease classification system. *Crit Care Med* 1985; 13:818–829
45. Render ML, Welsh DE, Kollef M, et al: Automated computerized intensive care unit severity of illness measure in the Department of Veterans Affairs: Preliminary results. *Crit Care Med* 2000; 28:3540–3548
46. Render ML, Kim HM, Welsh DE, et al: Automated intensive care unit risk adjustment: Results from a National Veterans Affairs study. *Crit Care Med* 2003; 31: 1638–1646
47. Render ML, Kim HM, Deddens J, et al: Variations in outcomes in Veterans Affairs ICUs with a computerized severity measure. *Crit Care Med* 2005; 33:930–939
48. Nilsson G, Astermark J, Lethagen, et al: The prognostic value of global haemostatic tests in intensive care unit setting. *Acta Anaesthesiol Scand* 2002; 46:1062–1067
49. Kamath PS, Wiesner RH, Malinchoe M, et al: A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001; 33: 464–470
50. Vanderschueren S, De Weerdt A, Malbrain M, et al: Thrombocytopenia and prognosis in intensive care. *Crit Care Med* 2000; 28: 1871–1876
51. Akea S, Haji'Michael P, de Mendonca A, et al: Time course of platelet counts in critically ill patients. *Crit Care Med* 2002; 30: 753–756
52. Hamel MB, Goldman L, Teno J, et al: Identification of comatose patients at high risk for death or severe disability. *JAMA* 1995; 273:1842–1848
53. Shepardson LB, Younger SJ, Speroff T, et al: Increased risk of death in patients with do-not-resuscitate orders. *Med Care* 1999; 37: 727–737
54. Rocker G, Cook D, Sjokvist P, et al: Clinician predictions of intensive care unit mortality. *Crit Care Med* 2004; 32:1149–1154
55. Goldfrad C, Rowan K: Consequences of discharges from intensive care at night. *Lancet* 2000; 355:1138–1142
56. McQuillan P, Pilkington S, Allan A, et al: Confidential inquiry into the quality of care before admission to intensive care. *BMJ* 1998; 316:1853–1858
57. Goldhill DR, White SA, Sumner A: Physiological values and procedures in the 24 hours before ICU admission from the ward. *Anaesthesia* 1999; 54:529–534

APPENDIX TABLE 1. Relationship between predictor variables used in Acute Physiology and Chronic Health Evaluation IV and in-hospital mortality: Nonoperative diagnostic groups

Diagnostic Group	No. of Patients	No. Observed	Hospital Deaths (%) Predicted Rounded to Integer	SMR	χ^2	*p* Value	Coefficient	Odds Ratio
Cardiovascular diagnoses								
AMI								
Anterior	565	54 (9.6)	41 (7.3)	1.31	5.36	.02	0.10295	1.11
Inferior/lateral	863	36 (4.2)	44 (5.1)	0.82	1.99	>.10	−0.15253	0.86
Non-Q	643	34 (5.3)	50 (7.8)	0.68	7.40	<.01	−0.27087	0.76
Other	338	42 (12.4)	34 (10.0)	1.24	2.92	.09	Reference	N/A
Cardiac arrest	872	508 (58.3)	509 (58.4)	1.00	0.01	>.10	0.416919	1.52
Cardiogenic shock	206	89 (43.2)	91 (44.0)	0.98	0.08	>.10	0.239711	1.27
Cardiomyopathy	87	13 (14.9)	13 (14.9)	1.00	0.00	>.10	0.059962	1.06
Congestive heart failure	1,627	220 (13.5)	208 (12.8)	1.06	0.93	>.10	−0.42259	0.66
Chest pain, rule out AMI	347	1 (0.3)	6 (1.8)	0.16	5.11	.02	−1.12235	0.33
Hypertension	417	18 (4.3)	19 (4.5)	0.95	0.06	>.10	−0.81392	0.44
Hypovolemia/dehydration (not shock)	363	44 (12.1)	48 (13.3)	0.91	0.62	>.10	−0.62259	0.54
Hemorrhage (not related to GI bleeding)	88	14 (15.9)	14 (15.6)	1.02	0.01	>.10	−0.65676	0.52
Aortic aneurysm	152	32 (21.1)	28 (18.6)	1.13	0.87	>.10	0.649149	1.91
Peripheral vascular disease	396	28 (7.1)	23 (5.9)	1.20	1.29	>.10	−0.50275	0.60
Rhythm disturbance	1,120	73 (6.5)	98 (8.8)	0.74	9.29	<.01	−0.60306	0.55
Sepsis (by infection site)								
Cutaneous	157	32 (20.4)	46 (29.1)	0.70	8.44	<.01	0.12644	1.13
Gastrointestinal	361	162 (44.9)	153 (42.3)	1.06	1.50	>.10	−0.13011	0.88
Pulmonary	478	192 (40.2)	183 (38.2)	1.05	1.09	>.10	−0.25877	0.77
Urinary tract	573	101 (17.6)	111 (19.4)	0.91	1.59	>.10	−0.73279	0.48
Other location	359	118 (32.9)	126 (35.1)	0.94	1.25	>.10	−0.04234	0.96
Unknown location	466	176 (37.8)	175 (37.5)	1.01	0.02	>.10	−0.09338	0.91
Cardiac drug toxicity	140	11 (7.9)	14 (9.8)	0.80	0.76	>.10	−0.69094	0.50
Unstable angina	1,086	27 (2.5)	17 (1.6)	1.54	5.36	.02	−1.21273	0.30
Cardiovascular, other	758	66 (8.7)	70 (9.2)	0.94	0.31	>.10	−0.36966	0.69
Respiratory diagnoses								
Airway obstruction	189	17 (9.0)	12 (6.2)	1.46	3.32	.07	−0.97767	0.38
Asthma	241	5 (2.1)	5 (2.0)	1.06	0.02	>.10	−1.54068	0.21
Aspiration pneumonia	458	102 (22.3)	129 (28.2)	0.79	10.37	.001	−0.37224	0.69
Bacterial pneumonia	1,289	302 (23.4)	307 (23.8)	0.99	0.12	>.10	−0.04337	0.96
Viral pneumonia	86	21 (24.4)	19 (21.6)	1.13	0.53	>.10	0.254375	1.29
Parasitic/fungal pneumonia	42	12 (28.6)	19 (44.7)	0.64	5.59	.02	1.056187	2.88
COPD (emphysema/bronchitis)	878	133 (15.1)	118 (13.4)	1.13	2.65	>.10	−0.3987	0.67
Pleural effusion	154	42 (27.3)	40 (25.9)	1.05	0.18	>.10	0.189901	1.21
Pulmonary edema (noncardiac)	310	86 (27.7)	87 (28.2)	0.98	0.04	>.10	−0.24169	0.79
Pulmonary embolism	368	48 (13.0)	54 (14.7)	0.89	1.10	>.10	−0.05153	0.95
Respiratory arrest	490	167 (34.1)	158 (32.2)	1.06	1.06	>.10	−0.39063	0.68
Respiratory cancer (oral, larynx, lung, trachea)	132	63 (47.7)	65 (48.9)	0.98	0.10	>.10	0.966314	2.63
Restrictive lung disease (fibrosis, sarcoidosis)	78	31 (39.7)	35 (44.6)	0.89	0.98	>.10	1.555297	4.74
Respiratory Disease, other	1,094	232 (21.2)	226 (20.7)	1.03	0.28	>.10	0.24049	1.27
GI diagnoses								
GI bleeding, upper	1,236	122 (9.9)	128 (10.4)	0.95	0.45	>.10	−0.55183	0.58
GI bleeding lower/diverticulitits	607	42 (6.9)	46 (7.6)	0.91	0.51	>.10	−0.57947	0.56
GI bleeding, varices	192	21 (10.9)	35 (18.4)	0.60	11.27	.001	−0.52772	0.59
GI inflammatory disease	122	19 (15.6)	19 (16.0)	0.98	0.02	>.10	−0.21177	0.81
Neoplasm	40	15 (37.5)	14 (34.9)	1.08	0.18	>.10	0.19513	1.22
Obstruction	90	21 (23.3)	18 (19.5)	1.20	1.16	>.10	−0.36995	0.69
Perforation	67	17 (25.4)	16 (23.6)	1.08	0.16	>.10	−0.32717	0.72
Vascular insufficiency	37	15 (40.5)	18 (48.4)	0.84	1.41	>.10	0.714879	2.04
Hepatic failure	236	108 (45.8)	98 (41.4)	1.11	2.70	>.10	−0.11968	0.89
Intra/retroperitoneal hemorrhage	72	21 (29.2)	15 (21.0)	1.39	4.50	.03	−0.65954	0.52
Pancreatitis	170	30 (17.6)	26 (15.1)	1.17	1.18	>.10	−0.51363	0.60
Gastrointestinal, other	124	8 (6.5)	14 (11.3)	0.57	3.59	.06	−0.25259	0.78
Neurologic diagnoses								
Intracerebral hemorrhage	983	321 (32.7)	315 (32.0)	1.02	0.27	>.10	0.945056	2.57
Neurologic neoplasm	138	15 (10.9)	12 (8.7)	1.24	0.94	>.10	0.018953	1.02
Neurologic infection	145	26 (17.9)	16 (11.4)	1.58	7.51	<.01	−0.53578	0.59
Neuromuscular disease	89	6 (6.7)	6 (7.0)	0.97	0.01	>.10	−0.55065	0.58
Drug overdose	1,063	15 (1.4)	22 (2.0)	0.70	2.29	>.10	−1.55262	0.21
Subdural/epidural hematoma	314	34 (10.8)	41 (13.0)	0.83	1.69	>.10	0.295094	1.34
Subarachnoid hemorrhage, intracranial aneurysm	507	95 (18.7)	96 (19.0)	0.99	0.03	>.10	0.61595	1.85
Seizures (no structural disease)	541	45 (8.3)	47 (8.7)	0.95	0.14	>.10	−0.94217	0.39
Stroke	860	185 (21.5)	174 (20.2)	1.06	1.16	>.10	0.519453	1.68
Neurologic, other	308	13 (4.2)	26 (8.3)	0.51	8.52	<.01	−0.17683	0.84

APPENDIX TABLE 1—(*Continued*)

Diagnostic Group	No. of Patients	No. Observed	Hospital Deaths (%) Predicted Rounded to Integer	SMR	χ^2	*p* Value	Coefficient	Odds Ratio
Trauma diagnoses								
Trauma involving the head								
Head trauma with either chest, abdomen, pelvis, or spine injury	165	26 (15.8)	16 (9.8)	1.60	8.51	<.01	−0.37235	0.69
Head trauma with extremity or facial trauma	188	14 (7.4)	16 (8.4)	0.89	0.25	>.10	−0.36413	0.69
Head trauma only	625	113 (18.1)	105 (16.8)	1.08	1.04	>.10	0.595869	1.81
Head trauma with multiple other injuries	319	51 (16.0)	50 (15.6)	1.02	0.04	>.10	−0.06796	0.93
Trauma, chest and spine trauma	125	16 (12.8)	12 (9.4)	1.36	2.64	>.10	−0.71743	0.49
Trauma, spine only	122	12 (9.8)	9 (7.0)	1.41	1.94	>.10	0.033769	1.03
Multiple trauma (excluding head trauma)	829	41 (4.9)	37 (4.5)	1.10	0.46	>.10	−0.67811	0.51
Metabolic/endocrine diagnoses								
Acid-base, electrolyte disorder	191	22 (11.5)	24 (12.5)	0.92	0.24	>.10	−0.64058	0.53
Diabetic ketoacidosis	605	13 (2.1)	11 (1.8)	1.19	0.44	>.10	−1.7757	0.17
Hyperglycemic hyperosmolar nonketotic coma	352	59 (16.8)	44 (12.5)	1.35	7.65	<.01	−0.92716	0.40
Metabolic/endocrine, other	193	24 (12.4)	15 (7.6)	1.64	8.22	<.01	−0.98644	0.37
Hematologic diagnoses								
Coagulopathy, neutropenia, thrombocytopenia, pancytopenia	59	16 (27.1)	14 (24.3)	1.12	0.37	>.10	0.258172	1.29
Hematologic, other	269	34 (12.6)	28 (10.2)	1.24	2.34	>.10	−0.34235	0.71
Genitourinary diagnoses								
Renal, other	447	83 (18.6)	83 (18.6)	1.00	0.00	>.10	−0.54158	0.58
Miscellaneous diagnoses								
General, other	580	30 (5.2)	29 (5.0)	1.03	0.03	>.10	−0.66758	0.51

SMR, standardized mortality ratio; AMI, acute myocardial infarction; GI, gastrointestinal; COPD, chronic obstructive pulmonary disease.

Appendix is continued on the next page.

APPENDIX TABLE 2. Relationship between predictor variables used in Acute Physiology and Chronic Health Evaluation IV and in-hospital mortality: Postoperative diagnostic groups

Diagnostic Group	No. of Patients	No. Observed	Hospital Deaths (%) Predicted Rounded to Integer	SMR	χ^2	*p* Value	Coefficient	Odds Ratio
Cardiovascular surgery								
Valvular heart surgery	606	18 (3.0)	21 (3.4)	0.88	0.35	>.10	−1.37176	0.25
CABG with double or redo valve surgery	89	11 (12.4)	16 (17.6)	0.70	1.93	>.10	−0.15514	0.86
CABG with single valve surgery	423	28 (6.6)	28 (6.7)	0.99	0.00	>.10	−1.19943	0.30
Aortic aneurysm, elective repair	701	39 (5.6)	33 (4.7)	1.19	1.39	>.10	−0.7607	0.47
Aortic aneurysm, rupture	123	43 (35.0)	39 (32.0)	1.09	0.71	>.10	0.204405	1.23
Aortic aneurysm, dissection	55	5 (9.1)	9 (17.2)	0.53	3.29	.07	−0.17846	0.84
Femoral-popliteal bypass graft	284	8 (2.8)	9 (3.3)	0.86	0.19	>.10	−0.78657	0.46
Aorto-iliac, aorto-femoral bypass graft	256	6 (2.3)	6 (2.5)	0.94	0.03	>.10	−0.83119	0.44
Peripheral ischemia (embolectomy, thrombectomy, dilation)	457	33 (7.2)	27 (6.0)	1.21	1.51	>.10	−0.50421	0.60
Carotid endarterectomy	1,038	12 (1.2)	10 (0.9)	1.25	0.63	>.10	−1.33264	0.26
Cardiovascular surgery, other	784	48 (6.1)	46 (5.9)	1.03	0.06	>.10	−0.59045	0.55
Respiratory surgery								
Thoracotomy, malignancy	633	26 (4.1)	27 (4.3)	0.96	0.05	>.10	0.086934	1.09
Neoplasm, mouth, larynx	248	4 (1.6)	4 (1.5)	1.08	0.02	>.10	−1.15287	0.32
Thoracotomy, lung biopsy, pleural disease	126	12 (9.5)	12 (9.5)	1.01	0.00	>.10	0.405738	1.50
Thoracotomy, respiratory infection	84	2 (2.4)	6 (6.6)	0.36	2.85	.09	−0.00594	0.99
Respiratory surgery, other	440	24 (5.5)	21 (4.7)	1.15	0.60	>.10	−0.24922	0.78
GI surgery								
GI malignancy	745	56 (7.5)	74 (9.9)	0.76	5.66	.02	0.136283	1.15
GI bleeding	63	10 (15.9)	11 (18.0)	0.88	0.29	>.10	−0.32968	0.72
Fistula, abcess	94	10 (10.6)	10 (10.3)	1.03	0.01	>.10	−0.55666	0.57
Cholecystitis, cholangitis	194	9 (4.6)	15 (7.6)	0.61	2.75	.10	−0.59329	0.55
GI inflammation	62	2 (3.2)	9 (14.2)	0.23	8.44	<.01	−0.16559	0.85
GI obstruction	399	57 (14.3)	60 (15.0)	0.95	0.21	>.10	−0.18901	0.83
GI perforation	386	85 (22.0)	81 (20.9)	1.05	0.37	>.10	−0.18996	0.83
GI, vascular ischemia	155	36 (23.2)	48 (30.7)	0.76	5.65	.02	0.498328	1.65
Liver transplant	158	6 (3.8)	7 (4.7)	0.81	0.33	>.10	−1.37028	0.25
GI surgery, other	622	49 (7.9)	48 (7.7)	1.03	0.04	>.10	−0.29589	0.74
Neurologic surgery								
Craniotomy or transphenoidal procedure for neoplasm	836	19 (2.3)	17 (2.0)	1.12	0.26	>.10	−0.43774	0.65
Intracranial hemorrhage	103	24 (23.3)	18 (17.9)	1.30	2.58	>.10	0.526717	1.69
Subarachnoid hemorrhage (aneurysm, arteriovenous malformation)	166	11 (6.6)	8 (5.1)	1.31	0.98	>.10	0.318906	1.38
Subdural/epidural hematoma	213	30 (14.1)	36 (17.0)	0.83	1.68	>.10	0.715683	2.05
Laminectomy, fusion, spinal cord surgery	485	12 (2.5)	14 (2.9)	0.87	0.27	>.10	−0.62861	0.53
Neurologic surgery, other	453	19 (4.2)	17 (3.8)	1.09	0.19	>.10	0.003996	1.00
Trauma surgery								
Head trauma only	111	31 (27.9)	33 (30.1)	0.93	0.35	>.10	1.088819	2.97
Multiple trauma sites including the head	120	20 (16.7)	18 (15.3)	1.09	0.24	>.10	0.357798	1.43
Surgery for extremity trauma	139	7 (5.0)	10 (7.4)	0.68	1.44	>.10	−0.18039	0.83
Multiple trauma (excluding the head)	504	44 (8.7)	38 (7.5)	1.17	1.52	>.10	−0.37781	0.69
Genitourinary surgery								
Renal/bladder/prostate neoplasm	216	4 (1.9)	4 (1.9)	0.96	0.01	>.10	0.086934	1.09
Renal transplant	217	7 (3.2)	3 (1.5)	2.10	4.20	.04	−1.30845	0.27
Hysterectomy	91	7 (7.7)	4 (4.6)	1.66	2.16	>.10	−0.79585	0.45
Genitourinary surgery, other	88	2 (2.3)	5 (5.6)	0.40	2.15	>.10	−0.69357	0.50
Miscellaneous surgery								
Amputation (nontraumatic)	40	9 (22.5)	8 (20.7)	1.09	0.10	>.10	0.60491	1.83

SMR, stardardized mortality ratio; CABG, coronary artery bypass graft; GI, gastrointestinal.

Appendix is continued on the next page.

APPENDIX TABLE 3. Relationship between predictor variables used in Acute Physiology and Chronic Health Evaluation IV and in-hospital mortality: Non-splined variables other than diagnosis

Variable	No.	Observed Mortality, %	Predicted Mortality, %	SMR	Coefficient	*p* Value	Odds Ratio	95% Confidence Interval
Emergency surgery								
Yes	2,431	16.0	15.3	1.05	0.2491	.002	1.28	1.10–1.50
No	41,857	13.4	13.5	0.99				
Unable to assess GCS								
Yes	3,565	21.5	21.0	1.02	0.7858	<.001	2.19	1.99–2.42
No	40,723	12.8	12.9	0.99				
Ventilated on ICU day 1								
Yes	15,543	25.7	25.7	1.00	0.2718	<.001	1.31	1.22–1.41
No	28,745	7.0	6.9	0.99				
Thrombolytic therapy for acute myocardial infarction								
Yes	552	5.3	3.4	1.54	−0.5799	.008	0.56	0.37–0.86
No	1,857	7.4	8.1	0.91				
Rescaled GCS (15-GCS)					0.0391	<.001	1.04	1.03–1.05
15-GCS = 0	27,415	7.9	7.8	1.01			1.00	
15-GCS = 1, 2, 3	9,210	11.4	11.1	1.03			1.04–1.12	
15-GCS = 4, 5, 6	3,375	19.6	22.4	0.88			1.17–1.26	
15-GCS = 7, 8, 9	2,198	31.5	34.6	0.91			1.31–1.42	
15-GCS = 10, 11, 12	2,090	68.2	63.9	1.07			1.48–1.60	
Pao_2/Fio_2 ratio					−0.00040	.003	1.00	0.99–1.00
≤200	7,005	30.1	30.8	0.98			1.00–0.92	
201–300	4,641	20.2	20.6	0.98			0.92–0.89	
301–400	30,014	8.1	8.0	1.02			0.89–0.85	
401–500	1,922	18.2	18.7	0.97			0.85–0.82	
501–600	706	20.1	16.6	1.21			0.82–0.79	
Chronic health items								
AIDS	224	29.9	29.4	1.02	0.9581	<.001	2.61	1.93–3.53
Cirrhosis	498	24.7	27.1	0.91	0.8147	<.001	2.26	1.81–2.82
Hepatic failure	642	37.4	38.2	0.98	1.0374	<.001	2.82	2.32–3.44
Immunosuppressed	1,638	21.3	22.5	0.95	0.4356	<.001	1.55	1.37–1.75
Lymphoma	211	39.8	35.9	1.11	0.7435	<.001	2.10	1.58–2.80
Myeloma	350	39.7	40.2	0.99	0.9693	<.001	2.64	2.10–3.31
Metastatic cancer	1,117	36.3	37.0	0.98	1.0864	<.001	2.96	2.59–3.39
None	39,608	11.6	11.5	1.00				
Admission source								
Floor	6,324	23.6	23.9	0.99	0.0171	.72	1.02	0.93–1.12
Other hospital	2,816	18.0	17.7	1.01	0.0221	.71	1.02	0.91–1.15
Operating/recovery room	13,637	6.7	6.9	0.97	−0.5838	.08	0.56	0.29–1.08
Other	21,511	14.3	14.2	1.01				

SMR, standardized mortality ratio; GCS, Glasgow Coma Scale; ICU, intensive care unit.