Performance of Simplified Acute Physiology Score 3 in Predicting Hospital Mortality in Emergency Intensive Care Unit

Qing-Bian Ma, Yuan-Wei Fu, Lu Feng, Qiang-Rong Zhai, Yang Liang, Meng Wu, Ya-An Zheng

Department of Emergency Medicine, Peking University Third Hospital, Beijing 100091, China

Abstract

Background: Since the 1980s, severity of illness scoring systems has gained increasing popularity in Intensive Care Units (ICUs). Physicians used them for predicting mortality and assessing illness severity in clinical trials. The objective of this study was to assess the performance of Simplified Acute Physiology Score 3 (SAPS 3) and its customized equation for Australasia (Australasia SAPS 3, SAPS 3 [AUS]) in predicting clinical prognosis and hospital mortality in emergency ICU (EICU).

Methods: A retrospective analysis of the EICU including 463 patients was conducted between January 2013 and December 2015 in the EICU of Peking University Third Hospital. The worst physiological data of enrolled patients were collected within 24 h after admission to calculate SAPS 3 score and predicted mortality by regression equation. Discrimination between survivals and deaths was assessed by the area under the receiver operator characteristic curve (AUC). Calibration was evaluated by Hosmer-Lemeshow goodness-of-fit test through calculating the ratio of observed-to-expected numbers of deaths which is known as the standardized mortality ratio (SMR).

Results: A total of 463 patients were enrolled in the study, and the observed hospital mortality was 26.1% (121/463). The patients enrolled were divided into survivors and nonsurvivors. Age, SAPS 3 score, Acute Physiology and Chronic Health Evaluation Score II (APACHE II), and predicted mortality were significantly higher in nonsurvivors than survivors ($P < 0.05$ or $P < 0.01$). The AUC (95% confidence intervals [CIs]) for SAPS 3 score was 0.836 (0.796–0.876). The maximum of Youden's index, cutoff, sensitivity, and specificity of SAPS 3 score were 0.526%, 70.5 points, 66.9%, and 85.7%, respectively. The Hosmer-Lemeshow goodness-of-fit test for SAPS 3 demonstrated a Chi-square test score of 10.25, $P = 0.33$, SMR (95% CI) = 0.63 (0.52–0.76). The Hosmer-Lemeshow goodness-of-fit test for SAPS 3 (AUS) demonstrated a Chi-square test score of 9.55, $P = 0.38$, SMR (95% CI) = 0.68 (0.57–0.81). Univariate and multivariate analyses were conducted for biochemical variables that were probably correlated to prognosis. Eventually, blood urea nitrogen (BUN), albumin, lactate and free triiodothyronine (FT3) were selected as independent risk factors for predicting prognosis.

Conclusions: The SAPS 3 score system exhibited satisfactory performance even superior to APACHE II in discrimination. In predicting hospital mortality, SAPS 3 did not exhibit good calibration and overestimated hospital mortality, which demonstrated that SAPS 3 needs improvement in the future.

Key words: Hosmer-Lemeshow Good-of-fit Test; Independent Risk Factor; Probability of Hospital Mortality; Prognosis; Receiver Operating Characteristic; Simplified Acute Physiology Score 3; Standardized Mortality Ratio

INTRODUCTION

The Acute Physiology and Chronic Health Evaluation Score (APACHE[1]) came into use after its publication in the early 1980s. Since then, severity of illness scoring systems has gained increasing popularity in Intensive Care Units (ICUs). Physicians used them for predicting mortality and for assessing illness severity in clinical trials. Among these systems, the APACHE, Simplified Acute Physiology

Access this article online

Quick Response Code:

Website:
www.cmj.org

DOI:
10.4103/0366-6999.208250

Score (SAPS[2]), and mortality prediction model[3] are most widely used scoring systems.

Le Gall *et al*.[2] developed the SAPS by simplifying the APACHE. To adapt to the clinical setting, the SAPS evolves into its third version (SAPS 3[4,5]). The successor provides physicians with customized equations for different regions of the world, which would improve the accuracy for predicting mortality probability. SAPS 3 is widely used both in Europe and America in the ICUs while relevant evidence is absent in China as in clinical practice. Thus, the objective of the study was to assess the performance of SAPS 3 in predicting clinical prognosis and hospital mortality in emergency ICU (EICU).

Participants were graded by SAPS 3 admission score, the approach of which is complemented by the development of specific customized equations for major areas of the world. The overall discriminatory capability of the model was measured by receiver operating characteristic (ROC) curve. Hosmer-Lemeshow goodness-of-fit test was adopted to assess the quality of predictions in the validation sets.

METHODS

Ethical approval
The study was conducted in accordance with the *Declaration of Helsinki* and was approved by the local ethics committee of the institute. Informed written consent was obtained from all patients prior to their enrollment in this study.

Participants
This study was performed at the EICU of Peking University Third Hospital. Four hundred and sixty-three patients admitted to EICU from January 2013 to December 2015 were retrospectively studied, including 257 male and 206 female patients. The average age of participants is 72.1 ± 15.5 years, and all of them suffered from internal diseases. Patients <18 years old, with ICU stay <24 h or missing components for SAPS 3 analysis, were excluded from the study. Other exclusion criteria include treatment of mild hypothermia therapy. Only the first ICU admission of patients with multiple ICU admissions during a single hospital stay was considered.

Data collection
Data were collected by reviewing medical records of eligible patients, which include demographic data, surgical interventions, previous comorbidities, main diagnosis for ICU admission, length of stay, and the prognosis. The worst physiological data within 24 h of admission were recorded, such as vital signs, blood routine tests, and Glasgow Coma Scale score. Severity scores (APACHE IV, SAPS 3) were calculated based on clinical and laboratory data collected.

Predicting hospital mortality
Actual mortality rates of the sample were calculated. Predicted hospital mortalities were calculated using the general SAPS 3 and the customized AUS-SAPS3 equations as follows: logit = −32.6659 + ln (SAPS 3 score + 20.5958) × 7.3068, logit = 22.5717 + ln (SAPS 3 score + 1) × 5.3163

for Australasia SAPS 3 admission scores (AUS-SAPS3) and the probability of death = elogit/(1 + elogit).[4]

Statistical analysis
We used SPSS software version 19.0 (IBM corp., NY, USA) to perform the statistical analysis. Continuous variables were presented as mean ± standard deviation (SD) or median (interquartile range), relationships between groups were analyzed with Student's *t*-test or rank test. Kolmogorov-Smirnov tests were used for normality test. Chi-squared tests were adopted for comparing ratios. Multivariate logistic regression was performed to identify independent factors associated with hospital mortality. A two-tailed $P < 0.05$ was considered statistically significant. To evaluate prognostic performance of different models, the following tools were used: discrimination - capability of the model to distinguish between patients who die from patients who survive was evaluated, which was tested by measuring the area (area under the curve, [AUC]) under the ROC curve.[6] ROC curve is constructed by plotting the 1-specificity against sensitivity. Youden index, along with sensitivity and specificity, was obtained from the curve. *Z*-test was used to compare the AUC. The sensitivity is the proportion of patients who died that was predicted correctly by the model. Specificity refers to the proportion of true negatives. Youden index is a single statistic that captures the performance of a diagnostic test. J = Sensitivity + Specificity − 1. Its value ranges from −1 to 1. A value of 1 indicates that there are no false positives or false negatives, i.e., the test is perfect.

Calibration
Reliability of the system can be quantified in terms of calibration, which represents the level of accordance between observed and predicted probabilities of the outcome. This is usually derived from Hosmer-Lemeshow goodness-of-fit test, which is a statistical test for logistic regression models. The H-statistic is based on fixed cut points on the predictions such as the deciles of risk whereas the C-statistic is based on equally sized groups, based on probability of death. Notably, the Hosmer-Lemeshow test indicates that goodness-of-fit is fairly good when $P > 0.05$.[7,8]

Standardized mortality ratio
Standardized mortality ratio (SMR) corresponding to the ratio predicted mortality/mortality observed[9] was calculated for scoring systems. Confidence intervals (*CI*s) for the SMRs were calculated using a 95% confidence limit.[10] If the SMR is equal to 1.0, this means the number of observed deaths equals that of expected cases. If higher than 1.0, there is a lower number of deaths than is expected. The equation for the lower limit is as follows:

$$SMR_L = \frac{O\left(1 - \dfrac{1}{9O} - \dfrac{1.96}{3\sqrt{O}}\right)^3}{E}$$

The equation for the upper limit is as follows:

$$SMR_U = \frac{(O+1)\left(1 - \dfrac{1}{9(O+1)} + \dfrac{1.96}{3\sqrt{O+1}}\right)^3}{E}$$

RESULTS

Main patients' demographic and clinical characteristics are shown in Table 1. During the study period, there were 463 patients admitted to the ICU. The median age was 72.1 ± 15.5 years. Median length of ICU stay was 16 (8.8–29.0) days. Respiratory and cardiovascular diseases were the most frequent organ-specific indications for ICU admission. In-hospital mortality was 26.1% (121/463). Table 2 presents the performance of survivors and nonsurvivors. The age, SAPS 3 score, APACHE II score, and predicted mortality were all significantly greater in nonsurvivors than survivors ($P < 0.05$ or $P < 0.01$). There were significant difference in severe pneumonia, acute exacerbation of chronic obstructive pulmonary disease, interstitial lung disease, sudden death and septic shock between two groups, which indicated that respiratory and cardiovascular failure might account for most of the deaths.

Performances of the models are summarized in Table 3. The discriminative power, assessed using the AUC, was significantly lower ($Z = 4.172$, $P < 0.0001$) for the APACHE II model (0.741) as compared with SAPS 3 (0.836). The maximum Youden index of SAPS 3 was 0.526, at which the optimal cutoff was 70.5 points. APACHE II had the Youden index of 0.327 at its most when the optimal cutoff was 17.5 points. Hospital mortality reached 93.1% when SAPS 3 was no <70.5 points. The ROC curves for scores are shown in Figure 1.

The calibration of the SAPS 3 and APACHE II prognostic model: Hosmer-Lemeshow goodness-of-fit test for APACHE II, SAPS 3, and AUS-SAPS3 is shown in Table 4. The calculated Chi-square statistics are 4.13 ($P = 0.91$) for APACHE II, 10.25 ($P = 0.33$) for SAPS 3, and 9.55 ($P = 0.38$) for AUS-SAPS3. All P values > 0.05, which indicated goodness-of-fit of the three scores are fairly good. APACHE II models tended to underestimate the hospital mortality (SMR >1); however, SAPS 3 and AUS-SAPS 3 models tended to overestimate the hospital mortality (both SMR <1). Calibration curves for APACHE II, SAPS 3, and AUS-SAPS3 are shown in Figures 2–4. It is shown from the curves that SAPS 3 and AUS-SAPS 3 models overestimate the hospital mortality at nearly all deciles.

Analyses were conducted to assess the prognostic scores adjusting for respiratory disease, severe pneumonia, circulation system disease, sudden death, septic shock, and the use of mechanical ventilation on hospital mortality. Table 5 demonstrates the performance of subgroups sorted by disease. It shows that the SMR was <1 in all the subgroups, which indicates an overestimated hospital mortality across all subgroups.

Table 6 presents the variables that were collected but not included in the SAPS 3 score. Hemoglobin (Hb), hematocrit (HCT), and albumin followed the normal distribution when using the Kolmogorov-Smirnov test. Univariate analysis was performed to identify factors

associated with hospital mortality. There was significant difference in procalcitonin, brain natriuretic peptide, D-dimer, troponin I, Hb, blood urea nitrogen (BUN), albumin, lactate, glucose, free triiodothyronine (FT3), free thyroxine (FT4), HCT, and 24 h urine volume between two groups ($P < 0.05$).

Variables were analyzed using multiple logistic regression analysis. Results are listed in Table 7. BUN, albumin, lactate, and FT3 were selected as independent risk factors for prognostic value. The regression coefficients of albumin and FT3 were −0.062 and −0.918, respectively. The odds ratios of albumin and FT3 were 0.940 (95% *CI*:

Table 1: Basic demographic and clinical characteristics

Variables	Characteristics
Age (years)	72.1 ± 15.5
Male	257 (55.5)
Main diagnosis for ICU admission	
Respiratory diseases	228 (49.2)
Severe pneumonia	115 (24.8)
AECOPD	56 (12.1)
Bronchiectasis	15 (3.2)
Bronchial asthma	10 (2.2)
Interstitial lung disease	8 (1.7)
Pulmonary embolism	7 (1.5)
Lung cancer	5 (1.1)
Other	12 (2.6)
Circulation system diseases	141 (30.5)
Heart failure	46 (9.9)
Sudden death	34 (7.3)
Coronary heart disease	28 (6.0)
Septic shock	24 (5.2)
Other	9 (1.9)
Digestive system diseases	31 (6.7)
Gastrointestinal bleeding	13 (2.8)
Acute pancreatitis	11 (2.4)
Gastrointestinal infections	5 (1.1)
Other	2 (0.4)
Urinary system diseases	24 (5.2)
Kidney failure	20 (4.3)
Urinary infection	4 (0.9)
Drug intoxication	15 (3.2)
Nervous system disease	11 (2.4)
Acute complications of diabetes	9 (1.9)
Other	4 (0.9)
Mechanical ventilation on admission	
Invasive mechanical ventilation	191 (41.3)
Noninvasive mechanical ventilation	109 (23.5)
No mechanical ventilation	163 (35.2)
Hospital length of stay (days)	16 (8.8–29.0)
SAPS 3 scores	61 (53–72)
APACHE II scores	13 (9–18)
Hospital mortality	121 (26.1)

Data are presented as mean ± SD or *n* (%) or median (interquartile range). AECOPD: Acute exacerbation of chronic obstructive pulmonary disease; APACHE II: Acute Physiology and Chronic Health Evaluation Score II; SAPS 3: Simplified Acute Physiology Score 3; ICU: Intensive Care Unit; SD: Standard deviation.

Table 2: Performance of survivors and nonsurvivors

Items	Survivors (n = 342)	Nonsurvivors (n = 121)	P
Age, (years)	69.8 ± 16.2	78.6 ± 10.8	<0.001
Male	185	72	0.303
Length of stay (days)	16 (9–28)	18 (7–35)	0.960
Main diagnosis for ICU admission			
Respiratory disease			
Severe pneumonia	74 (21.6)	41 (33.9)	0.007
AECOPD	49 (14.3)	7 (5.8)	0.013
Bronchiectasis	14 (4.1)	1 (0.8)	0.148
Bronchial asthma	10 (2.9)	0	0.124
Interstitial lung disease	0	7 (5.8)	<0.001
Pulmonary embolism	7 (2.0)	0	0.249
Lung cancer	2 (0.6)	3 (2.5)	0.222
Circulation system diseases			
Heart failure	35 (10.2)	11 (9.1)	0.718
Sudden death	18 (5.3)	16 (13.2)	0.004
Coronary heart disease	21 (6.1)	7 (5.8)	1.000
Septic shock	11 (3.2)	16 (13.2)	<0.001
Digestive system diseases			
Gastrointestinal bleeding	10 (2.9)	3 (2.5)	1.000
Acute pancreatitis	10 (2.9)	1 (0.8)	0.340
Gastrointestinal infections	5 (1.5)	0	0.409
Urinary system diseases			
Kidney failure	18 (5.3)	2 (1.7)	0.093
Urinary infection	2 (0.6)	2 (1.7)	0.603
Drug intoxication	14 (4.1)	1 (0.8)	0.148
Nervous system disease	8 (2.3)	3 (2.5)	1.000
Acute complications of diabetes	9 (2.6)	0	0.156
SAPS 3 scores	57 (51.0–65.3)	75 (66.5–83.5)	<0.001
APACHE II scores	11 (8–16)	18 (13–25)	<0.001
APACHE II predicted mortality (%)	10.6 (5.9–20.8)	32.2 (16.5–56.9)	<0.001
SAPS 3 predicted mortality (%)	30 (19.0–46.5)	66 (49–79)	<0.001
AUS-SAPS 3 predicted mortality (%)	27.2 (17.3–43.1)	61.1 (45.5–73.4)	<0.001

Data are presented as mean ± SD or n (%) or median (interquartile range). SAPS 3: Simplified Acute Physiology Score 3; AUS-SAPS3: Australasia-SAPS 3; APACHE II: Acute Physiology and Chronic Health Evaluation Score II; AECOPD: Acute exacerbation of chronic obstructive pulmonary disease; SD: Standard deviation; ICU: Intensive Care Unit.

Table 3: Comparing the AUC of the SAPS 3 and APACHE II models

Variables	Score	AUC	95% CI	Cutoff	Sensitivity (%)	Specificity (%)	Z-test	P
SAPS 3	61 (53–72)	0.836	0.796–0.876	70.5	66.9	85.7	4.172	<0.0001
APACHE II	13 (9–18)	0.741	0.691–0.791	13.5	52.9	79.8		

AUC: Area under the curve; SAPS 3: Simplified Acute Physiology Score 3; APACHE II: Acute Physiology and Chronic Health Evaluation Score II; CI: Confidence interval.

Table 4: Comparison of calibration of the SAPS 3 and AUS-SAPS3

Version	APACHE II	SAPS 3	AUS-SAPS3
Hosmer-Lemeshow (χ^2, P)	4.13 (0.91)	10.25 (0.33)	9.55 (0.38)
SMR (95% CI)	1.21 (1.00–1.44)	0.63 (0.52–0.76)	0.68 (0.57–0.81)
Observed mortality rate (%)	26.1	26.1	26.1
Predicted mortality rate (%)	21.7	41.3	38.3

SMR: Standardized mortality ratio; SAPS 3: Simplified Acute Physiology Score 3; AUS-SAPS3: Australasia SAPS 3; APACHE II: Acute Physiology and Chronic Health Evaluation Score II; CI: Confidence interval.

0.896–0.986) and 0.399 (95% CI: 0.232–0.687), respectively, which manifested their protective effect for prognosis. In contrast, the regression coefficients of BUN and lactate were 0.036 and 0.282, with the odds ratios of 1.037 (95% CI: 1.010–1.064) and 1.326 (95% CI: 1.157–1.519), respectively. BUN and lactate were therefore demonstrated

Table 5: Performance of SAPS 3 for subgroups

Disease group	n	AUC (95% CI)	Hosmer-Lemeshow		SMR (95% CI)
			χ^2	P	
Total	463	0.836 (0.796–0.876)	10.25	0.33	0.63 (0.52–0.76)
Respiratory disease	228	0.831 (0.771–0.892)	5.77	0.80	0.67 (0.51–0.86)
Severe pneumonia	115	0.835 (0.757–0.912)	2.98	0.96	0.80 (0.57–1.08)
Circulation system disease	141	0.851 (0.789–0.913)	4.38	0.85	0.73 (0.54–0.96)
Sudden death	34	0.852 (0.727–0.978)	2.32	0.88	0.79 (0.45–1.28)
Septic shock	24	0.854 (0.706–1.000)	1.90	1.00	0.82 (0.42–1.43)
Mechanical ventilation	300	0.785 (0.719–0.851)	5.14	0.85	0.68 (0.55–0.83)
No mechanical ventilation	163	0.842 (0.765–0.920)	5.86	0.69	0.49 (0.31–0.73)

SMR: Standardized mortality ratio; AUC: Area under the curve; SAPS 3: Simplified Acute Physiology Score 3; CI: Confidence interval.

Table 6: Biomarkers between survivors and nonsurvivors

Variables	Survivors (n = 342)	Nonsurvivors (n = 121)	P
PCT (µg/L)	0.45 (0.12–1.90)	1.48 (0.42–6.57)	<0.001
BNP (ng/L)	2320.00 (844.00–9150.00)	5960.00 (1452.00–15,375.00)	<0.001
D-dimer (mg/L)	0.93 (0.45–2.03)	1.30 (0.70–3.47)	<0.001
TnI (µg/L)	0.01 (0.01–0.07)	0.03 (0.01–0.28)	<0.001
Hemoglobin (g/L)	109.09 ± 28.20	98.62 ± 24.97	<0.001
BUN (mmol/L)	8.00 (5.20–13.45)	12.00 (7.25–21.10)	<0.001
Albumin (g/L)	31.64 ± 6.90	28.12 ± 5.83	<0.001
Na (mmol/L)	142.00 (138.00–146.00)	141.20 (137.80–148.00)	0.833
K (mmol/L)	3.94 (3.68–4.40)	4.14 (3.70–4.60)	0.061
Lactate (mmol/L)	1.10 (0.80–1.70)	1.40 (1.10–2.40)	<0.001
Glucose (mmol/L)	6.80 (5.10–9.43)	7.90 (5.40–10.50)	0.020
FT3 (pmol/L)	1.87 (1.53–2.25)	1.55 (1.23–1.80)	<0.001
FT4 (pmol/L)	1.09 (0.93–1.26)	1.01 (0.80–1.18)	0.001
TSH (mIU/L)	0.73 (0.30–1.73)	0.62 (0.28–1.51)	0.539
Hematocrit (%)	33.70 ± 9.87	30.56 ± 7.76	0.003
Urine volume (24 h, ml)	1550.00 (1050.00–2412.50)	1450.00 (850.00–2125.00)	0.029
Length of stay (days)	16.00 (9.00–28.00)	18.00 (7.00–35.00)	0.960

Data are presented as mean ± SD or median (interquartile range). PCT: Procalcitonin; BNP: Brain natriuretic peptide; TnI: Troponin I; BUN: Blood urea nitrogen; FT3: Free triiodothyronine; FT4: Free thyroxine; SD: Standard deviation; TSH: Thyroid-stimulating hormone.

Table 7: Multiple logistic regression for variables

Variables	B	SE	Wals	df	P	OR	95% CI
BUN	0.036	0.013	7.524	1	0.006	1.037	1.010–1.064
Albumin	−0.062	0.024	6.494	1	0.011	0.940	0.896–0.986
Lactate	0.282	0.069	16.475	1	<0.001	1.326	1.157–1.519
FT3	−0.918	0.277	10.991	1	0.001	0.399	0.232–0.687
Constant	1.375	0.764	3.243	1	0.072	3.955	

BUN: Blood urea nitrogen; FT3: Free triiodothyronine; SE: Standard error; CI: Confidence interval; OR: Odds ratio.

as risk factors for patients. Moreover, higher values were associated with a high risk of death.

DISCUSSION

SAPS 3 is used worldwide in ICUs. Some important differences within the patients might affect outcome. These include, for example, genetic makeups, styles of living, and distribution of major diseases in different regions, as well as availability of the health-care system. While relevant evidence is absent in China at present, in this study, we have conducted a retrospective analysis to verify its prognostic value.

Participants admitted to EICU were mostly aged patients, thus respiratory and circulation system diseases accounted for 80% of all causes. We evaluated the performance of survivors and nonsurvivors: SAPS 3 and APACHE II scores were both significantly greater in nonsurvivors than survivors. Higher severity score was associated with higher hospital mortality. The SAPS 3 demonstrated better discrimination with the AUC (0.836) higher than APACHE II (AUC = 0.741). Besides, its sensitivity and specificity were superior to APACHE II at the optimal cutoff, which also indicted better discrimination. However, the two prognostic
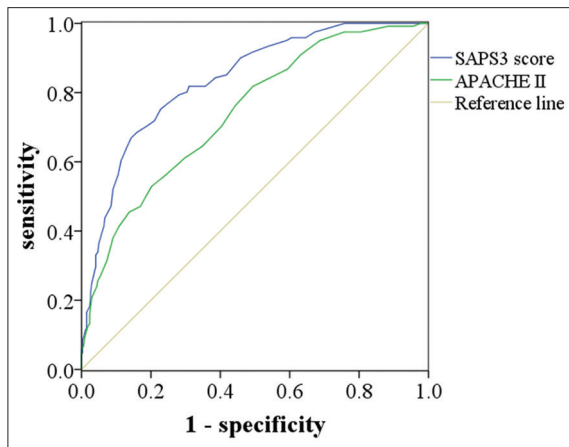
Figure 1: Receiver operating characteristic curve for Simplified Acute Physiology Score 3 (SAPS3) and Acute Physiology and Chronic Health Evaluation Score II (APACHE2).
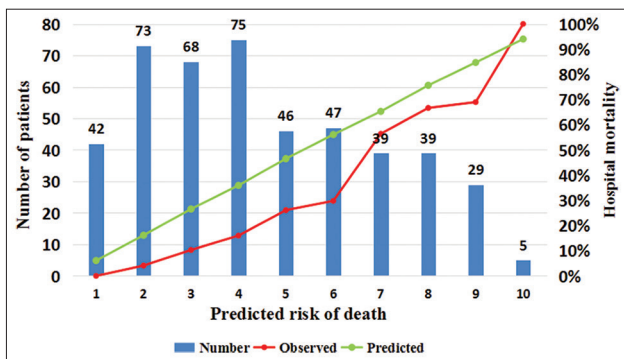

Figure 2: Calibration curves for Acute Physiology and Chronic Health Evaluation Score II.

Figure 3: Calibration curves for Simplified Acute Physiology Score 3.
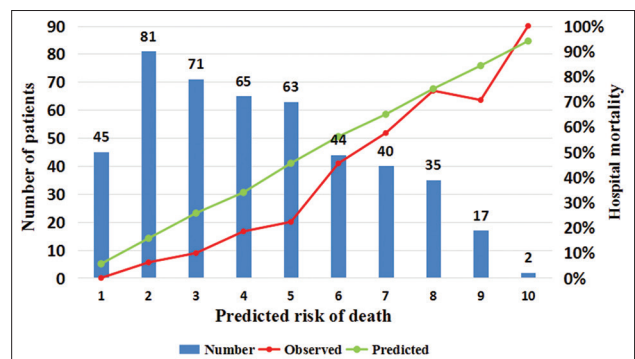
Figure 4: Calibration curves for Simplified Acute Physiology Score 3 (Australasia).

models did not showed excellent calibration, and the SAPS 3 scoring system (SMR = 0.63) was not superior to APACHE II (SMR = 1.21) in calibration.

Moreno *et al.*[5] developed the customized equations for major areas of the world using multilevel logistic regression. AUS-SAPS 3 was developed from a database built from patients in Australasia, India, and Hong Kong.[4,5] According to the study of Lim *et al.*,[11] general and regional Australasia SAPS 3 admission scores showed poor calibration for use in Korean ICU patients. However, the prognostic power of the SAPS 3 was significantly improved after country-specific customization. The present study shows that SAPS 3 and AUS-SAPS 3 had excellent calibration (*P* = 0.33 and *P* = 0.38, respectively), yet overestimated hospital mortality (SMR = 0.63 and SMR = 0.68, respectively). The findings are in accordance with previous studies. Furthermore, the calibration has not improved after customizing the logit of the original equation. Similar conclusion can be drawn from another study by Lim *et al.*[12] We came up with two explanations. First, the AUS-SAPS 3 equation was derived from data collected from 1756 patients from Australia (*n* = 651), India (*n* = 532), and Hong Kong (*n* = 573). There are likely to be vast differences stemming from variability in genetic makeups, styles of living, and distribution
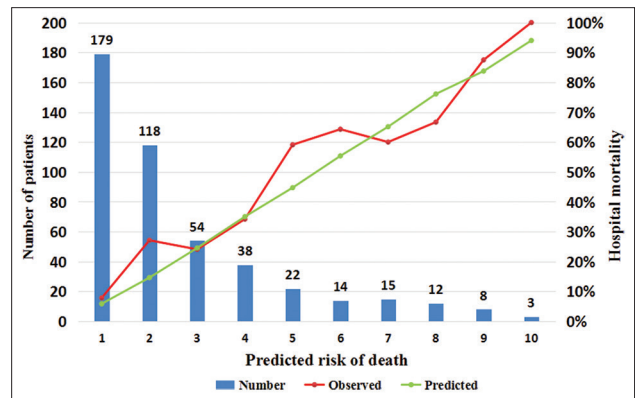
of major diseases. Second, all participants suffered from internal disease at our center, whereas SASP 3 was built on multicenter and multinational cohort study. That means the disease spectrum is more diverse. Moreover, during the retrospective analysis in the EICU, biomarkers were recorded as the worst value in the first 24-h period rather than the value within 1 h on admission, which might contribute to the overestimation of the hospital mortality.

We also analyzed prognostic power of the SAPS 3 for various subgroups, including respiratory disease, severe pneumonia, circulation system disease, sudden death, septic shock, and the use of mechanical ventilation. Except for the patients on mechanical ventilation, all the subgroups demonstrated excellent discrimination (AUC all >0.8). It can be concluded from above that SAPS 3 is good at distinguishing between nonsurvivors and survivors. Moreover, all subgroups showed fair calibration by Hosmer-Lemeshow goodness-of-fit test (all *P* > 0.05). The calibration remains to be improved since SAPS 3 and AUS-SAPS 3 models overestimated the hospital mortality in all subgroups. This systematic overestimation of mortality has been reported in other studies.[13-15] Nonetheless, groups of severe pneumonia, septic shock, and sudden death had SMR of 0.80, 0.79, and 0.82, respectively, which were distinctly higher than other groups. As expected, the use of SAPS 3 resulted in improved calibration for common severe diseases of ICU.

We determined four biomarkers as significant variables by univariate analysis and multiple logistic regression analysis. As shown above, BUN, albumin, lactate, and FT3 were related to prognosis. These biomarkers affected critical patients in certain ways though not included in the SAPS 3 score. BUN and creatinine functions are similar that both are biomarkers for acute or chronic renal damage. Systemic diseases such as serious infection, multiple organ failure, and severe acute pancreatitis can increase BUN levels, followed by a poor prognosis.[16,17] The level of albumin appeared negatively related with the mortality. Serum albumin is produced in the liver. Albumin is usually degraded ubiquitously, in an amount comparable to that synthesized by the liver. Clinical conditions activating an inflammatory process might repress albumin synthesis, i.e., sepsis, trauma, and massive hemorrhage.[18] Besides, two of the most important determinants of acute hypoalbuminemia are hemodilution during fluid resuscitation and capillary leakage into the interstitial space in patients with a systemic inflammatory response.[19] A reduction of albumin concentration usually results in decreasing blood volume, which might even cause multiple organ dysfunction when serious. Furthermore, an essential function of albumin is to neutralize toxic compounds such as oxygen radicals and nitrite peroxides, decreased albumin can make infection control more difficult. Hypoalbuminemia strongly links with mortality of critical patients.[20] Lactate is the end product of anaerobic glycolysis. In situations of hypoperfusion or hypoxia, pyruvate will no longer enter into the mitochondria for aerobic metabolism, but instead, it is preferentially reduced to lactate, resulting in the accumulation of lactate in the blood.[21] The International Guidelines for Management of Severe Sepsis and Septic Shock of 2012 pointed out that the mortality rate is high in septic patients with lactate ≥4 mmol/L alone. Research has shown that patients with high serum lactate levels (>10 mmol/l) showed severe acidosis and their probability of mortality was high.[22] Euthyroid sick syndrome occurs in a variety of nonthyroidal illnesses. In critical illness, many abnormalities in the pituitary–thyroid axis have been demonstrated, including attenuated TRH response, decreased TSH release, decreased level of TBG, decreased total T_4 and T_3 levels, low tissue uptake of thyroid hormones, and altered thyroid hormone metabolism. The syndrome has been reported in starvation, acute and chronic medical illnesses, bone marrow transplantation, surgery, trauma, and, in fact, can be seen in any severe systemic illness.[23] FT3, FT4, and TSH levels were all found to be associated with higher mortality in intensive care patients.[24,25] Generally speaking, measurement of BUN, albumin, lactate, and FT3 levels might be useful as a predictor of mortality in intensive care patients. Those four biomarkers are expected to be included in model in the future.

The present study has potential limitations. One could criticize the study as a retrospective and single-center study, with relatively small sample size. For this reason, it is necessary to validate the SAPS 3 prognostic model in a prospective multicenter study to minimize possible biases. Another potential limitation is the fact that the study overestimated hospital mortality. We might need to customize SAPS 3 model for China using the method that other studies applied.[10,26]

The SAPS 3 score system exhibited satisfactory performance, even superior to APACHE II in discrimination. In predicting hospital mortality, SAPS 3 did not exhibit good calibration and overestimated hospital mortality. To improve the predictive value of the model, it might be necessary to improve the models on a regional basis.

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: A physiologically based classification system. Crit Care Med 1981;9:591-7. doi: 10.1097/00003246-198108000-00008.
2. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. Crit Care Med 1984;12:975-7. doi: 10.1097/00003246-198411000-00012.
3. Teres D, Lemeshow S, Avrunin JS, Pastides H. Validation of the mortality prediction model for ICU patients. Crit Care Med 1987;15:208-13. doi: 10.1097/00003246-198703000-00005.
4. Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3 – From evaluation of the patient to evaluation of the Intensive Care Unit. Part 1: Objectives, methods and cohort description. Intensive Care Med 2005;31:1336-44. doi: 10.1007/s00134-005-2762-6.
5. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3 – 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;31:1345-55. doi: 10.1007/s00134-005-2763-5.
6. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36. doi: 10.1148/radiology.143.1.7063747.
7. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982;115:92-106. doi: 10.1093/oxfordjournals.aje.a113284.
8. Chunxiao W. Performance of Three Mortality Prediction Scores in Term Neonates (in Chinese) [D]. College of Medicine, Zhejiang University; 2010.
9. Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance based on logistic regression models. Stat Med 1995;14:2161-72. doi: 10.1002/sim.4780141909.
10. Haaland OA, Lindemark F, Flaatten H, Kvåle R, Johansson KA. A calibration study of SAPS II with Norwegian intensive care registry data. Acta Anaesthesiol Scand 2014;58:701-8. doi: 10.1111/aas.12327.
11. Lim SY, Koh SO, Jeon K, Na S, Lim CM, Choi WI, et al. Validation of SAPS3 admission score and its customization for use in Korean Intensive Care Unit patients: A prospective multicentre study. Respirology 2013;18:989-95. doi: 10.1111/resp.12115.
12. Lim SY, Ham CR, Park SY, Kim S, Park MR, Jeon K, et al. Validation of the simplified acute physiology score 3 scoring system in a Korean Intensive Care Unit. Yonsei Med J 2011;52:59-64. doi: 10.3349/ymj.2011.52.1.59.
13. Maccariello E, Valente C, Nogueira L, Bonomo H, Ismael M, Machado JE, et al. SAPS 3 scores at the start of renal replacement therapy predict mortality in critically ill patients with acute kidney injury. Kidney Int 2010;77:51-6. doi: 10.1038/ki.2009.385.
14. Costa e Silva VT, de Castro I, Liaño F, Muriel A, Rodríguez-Palomares JR, Yu L. Performance of the third-generation

models of severity scoring systems (APACHE IV, SAPS 3 and MPM-III) in acute kidney injury critically ill patients. Nephrol Dial Transplant 2011;26:3894-901. doi: 10.1093/ndt/qfR201.

15. Hernandez AM, Palo JE. Performance of the SAPS 3 admission score as a predictor of ICU mortality in a Philippine private tertiary medical center Intensive Care Unit. J Intensive Care 2014;2:29. doi: 10.1186/2052-0492-2-29.

16. Lu RH, Fang Y, Gao JY, Cai H, Zhu ML, Zhang MF, *et al*. Analysis of incidence and risk factor in hospitalized patients with acute kidney injury (in Chinese). Chin Crit Care Med 2011;23:413-7. doi: 10.3760/cma.j.issn.1003-0603.2011.07.010.

17. Han YM, Xin-Hong WU, Xiao DP. Etiology and risk factors of acute kidney injury (in Chinese). Chin Gen Pract 2011;14:3696-8. doi: 10.3969/j.issn.1007-9572.2011.32.013.

18. Caironi P, Gattinoni L. The clinical use of albumin: The point of view of a specialist in intensive care. Blood Transfus 2009;7:259-67. doi: 10.2450/2009.0002-09.

19. McCluskey A, Thomas AN, Bowles BJ, Kishen R. The prognostic value of serial measurements of serum albumin concentration in patients admitted to an Intensive Care Unit. Anaesthesia 1996;51:724-7. doi: 10.1111/j.1365-2044.1996.tb07883.x.

20. Wang Q, Zhou LN, Zhang CB, Cao Z. Analysis on factors related to deaths in emergency Intensive Care Unit (in Chinese). Chin J Emerg Med 2008;17:306-9. doi: 10.3760/j.issn: 1671-0282.2008.03.019.

21. Zhang Z, Xu X, Chen K. Lactate clearance as a useful biomarker for the prediction of all-cause mortality in critically ill patients: A systematic review study protocol. BMJ Open 2014;4:e004752. doi: 10.1136/bmjopen-2013-004752.

22. Lee SW, Hong YS, Park DW, Choi SH, Moon SW, Park JS, *et al*. Lactic acidosis not hyperlactatemia as a predictor of in hospital mortality in septic emergency patients. Emerg Med J 2008;25:659-65. doi: 10.1136/emj.2007.055558.

23. Umpierrez GE. Euthyroid sick syndrome. South Med J 2002;95:506-13. doi: 10.1097/00007611-200205000-00006.

24. Zhang HY, Dong L, Wu CJ, Zhao F. Clinical significance of low triiodothyronine impact on evaluation of prognosis (in Chinese). Chin J Emerg Med 2013;22:1132-5. doi: 10.3760/cma.j.issn.1671-0282.2013.10.014.

25. Tas A, Tetiker T, Beyazit Y, Celik H, Yesil Y. Thyroid hormone levels as a predictor of mortality in intensive care patients: A comparative prospective study. Wien Klin Wochenschr 2012;124:154-9. doi: 10.1007/s00508-012-0125-1.

26. Moreno R, Apolone G. Impact of different customization strategies in the performance of a general severity score. Crit Care Med 1997;25:2001-8. doi: 10.1097/00003246-199712000-00017.