Croat Med J 2005;46(6):964-969

Evaluation of Single Intensive Care Unit Performance by Simplified Acute Physiology Score II System

Kristian Deša, Alan Šustić, Željko Župan, Božidar Krstulović, Vesna Golubović

Department of Anesthesiology and Intensive Care, Rijeka University Hospital, Rijeka, Croatia

Aim To evaluate effectiveness and quality of care in a single intensive care unit (ICU) by the Simplified Acute Physiology Score II (SAPS II).

Methods A prospective study included 395 patients from the ICU at Rijeka University Hospital, Croatia. The sum of the SAPS II points was used for calculating predicted mortality for each patient. The observed death rate was compared with predicted mortality calculated by SAPS II system. The ability of the SAPS II prognostic system to predict probability of hospital mortality was assessed with discrimination (receiver operating characteristic [ROC] curve) and calibration (Hosmer-Lemeshow test) measures.

Results The SAPS II score on the first ICU day was low (median, 20; range, 3-83). SAPS II system showed a good ability to separate the patients predicted to live from those predicted to die, as shown by an area under the ROC curve of 0.827. The calibration curve demonstrated under-prediction of the actual death rate (Hosmer-Lemeshow goodness-of-fit test, $C=22.961$; $df=8$; $P=0.003$). The observed mortality was higher than predicted (observed-to-predicted ratio was 1.6).

Conclusions SAPS II system is a useful tool for the assessment of ICU performance. This system demonstrated a good ability of discrimination, but an under-prediction of the actual mortality rate in our ICU.

Over the last three decades, outcome prediction and quantization of the severity of illness has become an irreplaceable tool for the estimation of effectiveness and quality of intensive care as a supplement to structural, procedural, and outcome measurement methods, such as technology availability, staffing patterns, and patient procedures (1). In addition to life and death predictions, other objective common to all investigators is the evaluation of the performance of an individual intensive care unit (ICU) relative to international standards (2). A scoring system defines the severity scores of illness that could be used for the prediction of hospital mortality risk by applying logistic regression equations. Comparisons of actual and predicted outcomes for groups of patients can be used to compare different providers. It is assumed that standardized mortality ratio >1.0 may reflect poor care and, conversely, <1.0 ratio may reflect good care (3). One of the most widely used scoring system for the general severity of illness and prognosis is the Simplified Acute Physiology Score (SAPS) system. The second version of this scoring system was developed and validated for a large group of ICU patients on the basis of an analysis of a large database of physiologic data from critically ill medical and surgical patients (4). The SAPS II scoring system has been shown to accurately stratify risk of death in a wide range of disease states and clinical settings (5-14). This experience has resulted in the widespread use of the SAPS II scoring system as a tool for ICUs audit.

Our objective was to assess the ability of the SAPS II system to predict patient outcome and

to evaluate an individual ICU performance; only few similar studies from developing countries have been reported so far (5,6). We also wanted to compare the performance of our ICU in Croatia with similar data from other studies that used the same methodology.

Patients and Methods

Patients

The study was conducted at a single surgical-medical ICU with 21 beds (5% of acute beds in the hospital), Rijeka University Hospital, Croatia, from January 1 to June 30, 2002. The ICU employed a full-time anesthesiology and intensive care specialist and a resident. There were two nurses per ICU bed, with nurse-to-patient ratio during day and night shifts of approximately 0.5, depending on the number of patients. All physiologic monitors were available. There were several high dependency rooms in the hospital, but they were not equipped with monitors according to the ICU standards and there were no full-time nurses.

We prospectively collected data on 512 patients admitted consecutively to the ICU during the study period. One hundred and seventeen (22.8%) admitted patients were excluded because they met one or more exclusion criteria, as follows: younger than 18 years (n = 12), admitted for less than 4 h (n = 13), and heart surgery patients (n = 100). Only the data on the first admission to the ICU were used for calculating standardized mortality ratio.

We collected the following data on all variables used in the SAPS II system: age, physiologic variables (heart rate, systolic blood pressure, body temperature, partial arterial oxygen pressure/inspiratory oxygen concentration ratio, urinary output, blood urea nitrogen, white blood cell count, serum potassium, sodium, bicarbonate, bilirubin, and Glasgow coma scale), chronic diseases (metastatic cancer, hematologic malignancy and/or acquired immunodeficiency syndrome), and type of admission (scheduled surgical or medical treatment, or unscheduled surgical treatment). Patients operated in the first week before or after ICU admission were identified as surgical patients. For physiological variables, the worst value during the first 24 hours in the ICU was collected. The worse value was defined as the value that would have been assigned the greatest number of points (4). Microsoft® Access 2000 software database was used for data storage. All data were checked by the software for illogical, extreme, or unlikely values and were routinely collected for clinical purposes. The length of ICU stay was the duration of care from admission to discharge from the ICU. Length of stay in the hospital was the duration of care from admission to the ICU to discharge from the hospital. The main outcome was survival status on discharge from the hospital, including deaths in the ICU and hospital wards after discharge from the ICU.

Statistical Analysis

Unvaried comparisons were performed to compare survivors and non-survivors at hospital discharge. All continuous variables were presented as medians with the range and compared by the Mann Whitney U-test. Categorical variables were expressed as numbers and percentages and compared by χ^2 test. All statistical tests were two-sided, with $P < 0.05$ considered statistically significant. Statistica 6.1 (StatSoft. Inc, Tulsa, OK, USA) and SPSS 12.0 (SPSS. Inc, Chicago, IL, USA) statistical packages were used for data analysis.

The observed death rate was compared with the predicted death rate for the study population. We calculated predicted hospital mortality rates for SAPS II, using the logistic regression model suggested by Le Gall et al (4). The predicted risk for each patient was calculated from the SAPS II risk of death equation based on the patient's SAPS II score, which was calculated by summing of points for each variable. The predicted death rate was the sum of SAPS II estimates of hospital mortality risk of individual patients divided by the number of patients in the given groups of ICU admissions. The standardized mortality ratio, obtained by dividing the observed number of deaths for each group by the predicted number, was used to compare observed with predicted mortality.

The accuracy of prediction was tested by Hosmer-Lemeshow C statistic and calibration curves. The admissions were ranked according to the predicted risk of death with an approximately equal number of patients. The records indicated the agreement between the observed and predicted mortality across risk ranges. Large C values and low *P* values (< 0.05) suggested that the model did not correctly reflect the actual outcome. A calibration curve, using 10 equal contiguous risk ranges, presented the observed against the predicted outcomes. The observed death rates were

plotted against the predicted death rates stratified by 10% risk ranges in a calibration curve.

The discrimination power, defined as the ability of the model to discriminate between survivors and non-survivors, was assessed by calculating the area under the receiver operating characteristic (ROC) curve, with estimates of standard error (SE) and 95% confidence intervals (CI). The ROC curve shows the difference between selectivity and sensitivity. Typically, a curve of false positive rate versus true positive rate is plotted, while a sensitivity or threshold parameter is varied. The area under the ROC curve estimated the ability of the model to assign a higher risk of death to patients who die. The predicted and observed outcomes were compared using 2×2-decision matrices at four different decision criteria, as follows: predicted risk of death of 0.1, 0.2, 0.5, and 0.8 (15). The sensitivity, specificity, positive predictive value, negative predictive value, and total correct classification rates derived from classification tables were recorded.

Results

We recorded age, sex, type of admission, severity of illness, and surgical status of 395 patients admitted to the ICU, who were included in the study (Table 1).

There were 57 deaths in the ICU (14.4%) and 87 deaths in the hospital (22.0%). Thirty-four percent of patients stayed in the ICU up to 24h, and 56% up to 4 days. In comparison with the survivors, the non-survivors were older and more severely ill ($P<0.001$ for both), but did not have longer stays in the ICU ($P=0.086$; Mann-Whitney U-test). Hospital mortality was significantly higher for medical patients than for surgical patients (36.2% vs 18.4%, $P<0.001$, χ^2 test). Emergency surgery patients had a significantly higher hospital mortality rate than elective surgery patients (33.3% vs 7.2%, $P<0.001$, χ^2 test). However, there were no significant differences in hospital mortality between medical and emergency surgery patients (36.2% vs 33.3%, $P=0.870$, χ^2 test).

The observed hospital mortality was significantly higher than the SAPS II predicted mortality rate (22.0% vs 13.8%, $P<0.001$, χ^2 test). The standardized mortality ratio for the whole study population was 1.60. Age and sex had no significant influence on the standardized mortality ratio. The standardized mortality ratio was high for medical, elective, and emergency surgery patients (1.46, 1.67, and 1.67, respectively), showing significant under-prediction of the observed mortality.

The majority of admissions had low SAPS II probabilities of death. The predicted risk of hospital mortality was below 0.1 for 64% of the patients, and below 0.3 for 81% of the patients. The accuracy of risk prediction evaluated by Hosmer-Lemeshow statistics failed to confirm adequate calibration with the original SAPS II development database ($C=22.961$; $df=8$; $P=0.003$). The calibration curve for SAPS II equation applied to our data showed a marked deviation from the diagonal for the range 0.5 risk of hospital death (Fig. 1). From the mathematical point of view, this deviation was the main reason for the high standardized mortality ratio because the curve lay close to the diagonal for the groups of predicted risk below and above range of 0.5 risk of hospital death.

Table 1. Characteristics of patients admitted to the intensive care unit (ICU) at Rijeka University Hospital

Characteristic	Patients			*P**
	total (n=395)	survivors (n=308)	non-survivors (n=87)	
Age (years, median, range)	59 (16-90)	57 (16-90)	65 (19-85)	<0.001
Sex (No., %):				
men	247 (62.5)	201 (65.3)	46 (52.9)	0.035
women	148 (37.5)	107 (34.7)	41 (47.1)	
Type of patients (No., %):				
medical	80 (20.3)	51 (16.6)	29 (33.3)	<0.001
surgical	315 (79.7)	257 (83.4)	58 (66.7)	
Surgical status (No., %):				
emergency surgery	135 (42.9)	90 (35.0)	45 (77.6)	<0.001
elective surgery	180 (57.1)	167 (65.0)	13 (22.6)	
SAPS II score (median, range)	20 (3-83)	16 (3-71)	43 (9-83)	<0.001
Length of stay (median, range, days):				
ICU	4 (1-105)	3 (1-105)	6 (1-45)	0.086
hospital	13 (1-159)	14 (1-159)	9 (1-141)	0.026

*Continuous variables are presented as means with the range and compared by the Mann Whitney U-test; categorical variables are expressed as numbers and percentages and compared by χ^2 test.

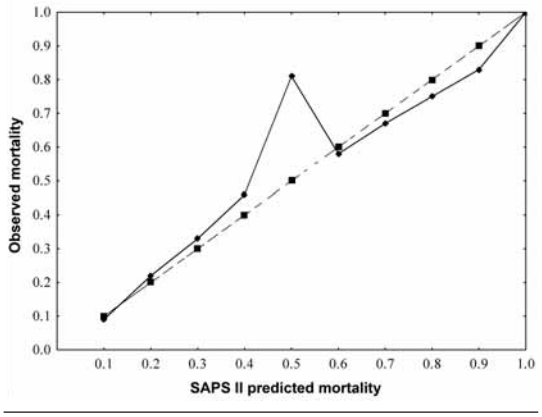

Figure 1. Calibration curve for the Simplified Acute Physiology Score (SAPS) II hospital mortality model, comparing the observed hospital mortality (full line) for the patients grouped by the predicted risk of hospital death. The line of ideal predictive ability (dashed line) is where the number of observed and predicted death is equal.

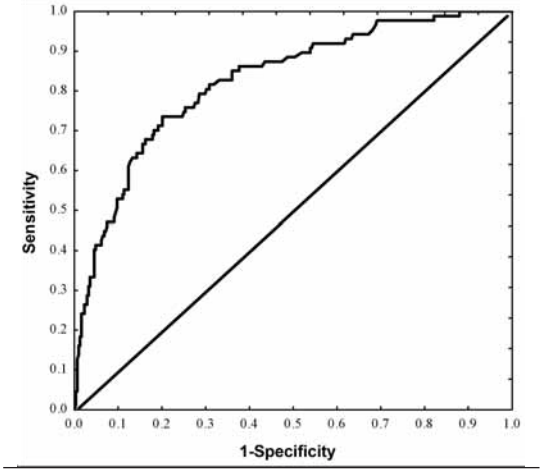

Figure 2. Receiver operating characteristic (ROC) curve for the Simplified Acute Physiology Score (SAPS) II (curved line). The relationship between true-positive sensitivity and false-positive 1 minus specificity is shown. Straight diagonal line is the line of chance performance. The area under the curve is 0.827.

The ability of the SAPS II system to predict prognosis correctly was tested by classification matrices (Table 2). The highest overall correct classification was obtained by a decision criterion of 0.5 with the sensitivity of 45.3%, the specificity of 90.8%, and total correct classification rate of 82.0%. The area under the ROC curve for the SAPS II equation applied to our data was 0.827 (SE, 0.029; 95% CI, 0.785-0.863) and confirmed a good discrimination of the SAPS II system (Fig. 2).

Discussion

This study was an institutional prospective assessment of the SAPS II system in Croatia. The application of severity scoring systems in other countries and comparison with the original database may produce useful information in assessing the state of intensive care medicine. The interpretation of ICU performance and comparisons of patient groups by mortality prediction models to produce a standardized mortality ratio were used in various countries with different social, demographic, economic, and medical environments. The evaluation of the ICU performance based on SAPS II severity of illness scoring has previously been reported from many Western European countries (4,7-14).

Our patients had a mean SAPS II score of 20, ie, the lower severity of illness. Their SAPS II score was lower than that reported in other independent, national and international studies (5-14). Compared to the ICUs in the original SAPS II study (4), we found that patients admitted to our ICU had been more frequently classified as surgical patients and had significantly lower SAPS II scores and death rates than medical patients.

The most commonly used measurement for the assessment of outcome is the ICU and hospital mortality rate. The overall mortality rate is insufficient in describing outcome and comparing groups of critically ill patients treated in different hospitals and countries. The observed hospital mortality rate in this study was within the mortality range limits found in other studies, depending on the case-mix, age, and chronic health status (5-14).

Table 2. Performance characteristics of Simplified Acute Physiology Score (SAPS) II system showed according to the predicted mortality decision criteria

Predicted mortality decision criteria	Performance parameters				
	positive predictive value	negative predictive value	sensitivity	specificity	overall correct classification rate
0.1	44.0	90.2	71.3	74.4	73.7
0.2	56.0	88.2	58.6	87.0	80.8
0.5	68.2	82.1	45.3	90.8	82.0
0.8	68.8	82.1	25.3	96.8	81.0

In comparison with the reports from the Western European countries, the present study showed that the observed death rate was higher than the predicted death rate (standardized mortality ratio of 1.6) when corrected for the severity of illness. However, the standardized mortality ratio for individual ICUs participating in multicenter studies using the SAPS II model shows large variations (5-14).

The analysis of our results did not show the large variations in standardized mortality ratios for the surgical and medical patient groups of ICU patients. This indicated a significant under-prediction of the observed mortality in all groups. The majority of patients with the lower SAPS II scores and risk of death were surgical patients, whereas patients with higher scores were predominantly medical. Also, 64% of the patients had a predicted risk of hospital mortality of 0.1, as most of them were admitted postoperatively to the ICU for invasive monitoring and close observation.

The performance of ICU mortality prediction models can be influenced by both clinical and non-clinical factors (16). The inaccuracy of data interpretation can arise from local differences of clinical practice, case-mix, or data collection. An observed mortality rate higher than predicted can be explained primarily by the quality of prehospital, inhospital, and intensive care. Other important factors are resource limitations, diagnostic diversity, lead-time bias (treatment received before ICU admission), teaching activity, staffing, and technology availability (16-20).

In our ICU, there were two possible causes for high standardized mortality ratio. First, nurse-to-patient ratio at our ICU was 0.5, whereas other studies reported from four to eight nurses per ICU bed and a nurse-to-patient ratio above 1.0, especially for mechanically ventilated patients (19,20). However, there were not differences in nurse-to-patient ratio between patients with 0.5 predicted death rate and other patients in our ICU. The patients with predicted death rate of 0.5 had the highest observed mortality. From a mathematical point of view, the high mortality of patients with predicted mortality of 0.5 was the major factor for high standardized mortality ratio, because the calibration curve for patients with predicted mortality below and above 0.5 lied close to ideal predictive mortality. Second, high dependency units in our hospital operate as regular hospital rooms with acute beds, so the patients who need only basic monitoring and close observation also have to be admitted to ICU. This is the reason why initial SAPS II score in our ICU was low. In addition, there were 30 patients who died after the first discharge from ICU, which may be a consequence of poor performance of high dependency units in our hospital. High mortality in the group of patients with predicted mortality of 0.5 remains unresolved because we could not find any differences in intensive care between them and other groups of patients. Possible reasons could be low-intensity care due to low number of ICU nurses and early discharge from ICU due to the pressure of new admissions.

Whatever severity of illness scoring systems is chosen for hospital mortality prediction and evaluation of the ICU, it is essential to know the goodness of fit in these areas of application, as well as the discriminatory power. In our study, the SAPS II model showed good capability of discriminating survivors from non-survivors, as the area under ROC of 0.827 was very close to 0.823, the value obtained from the original SAPS II model (4). In other studies, the area under ROC for the SAPS II model ranged from 0.744 to 0.888 (5-14). The Hosmer-Lemeshow statistic revealed insufficient calibration. Therefore, the SAPS II model did not fit very well to our ICU population. Many other studies also reported poor calibration (5-14), suggesting differences in performance between ICUs and patient populations.

In conclusion, the findings from the present study confirmed that the SAPS II system was a useful tool for the assessment of ICU outcome in Croatia. The standardized mortality ratio rather than the overall mortality rate or severity of illness score might be an objective measurement of ICU performance. The SAPS II system showed a good ability to distinguish patients who die from patients who live, as presented by ROC curve, but had a low degree of correspondence between the estimated probabilities of mortality and the actual mortality in our ICU, as presented by Hosmer-Lemeshow test. The SAPS II prediction model provided an opportunity to make an international comparison of intensive care. Furthermore, research at the state level that would include large number of ICUs and patients would be useful to show the overall quality of the intensive care in Croatia.

References

1 Teres D, Lemeshow S. When to customize a severity model. Intensive Care Med. 1999;25:140-2.

2 Civetta JM, Hudson-Civetta JA, Nelson LD. Evaluation of APACHE II for cost containment and quality assurance. Ann Surg. 1990;212:266-76.

3 Gunning K, Rowan K. ABC of intensive care: outcome data and scoring systems. BMJ. 1999;319:241-4.

4 Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA. 1993; 270:2957-63.

5 Vosylius S, Sipylaite J, Ivaskevicius J. Evaluation of intensive care unit performance in Lithuania using the SAPS II system. Eur J Anaesthesiol. 2004;21:619-24.

6 Nouira S, Belghith M, Elatrous S, Jaafoura M, Ellouzi M, Boujdaria R, et al. Predictive value of severity scoring systems: comparison of four models in Tunisian adult intensive care units. Crit Care Med. 1998;26:852-9.

7 Moreno R, Morais P. Outcome prediction in intensive care: results of a prospective, multicentre, Portuguese study. Intensive Care Med. 1997;23:177-86.

8 Livingston BM, MacKirdy FN, Howie JC, Jones R, Norrie JD. Assessment of the performance of five intensive care scoring models within a large Scottish database. Crit Care Med. 2000;28:1820-7.

9 Moreno R, Miranda DR, Fidler V, Van Schilfgaarde R. Evaluation of two outcome prediction models on an independent database. Crit Care Med. 1998;26:50-61.

10 Moreno R, Apolone G, Miranda DR. Evaluation of the uniformity of fit of general outcome prediction models. Intensive Care Med. 1998;24:40-7.

11 Bertolini G, D'Amico R, Apolone G, Cattaneo A, Ravizza A, Iapichino G, et al. Predicting outcome in the intensive care unit using scoring systems: is new better? A comparison of SAPS and SAPS II in a cohort of 1,393 patients. Med Care. 1998;36:1371-82.

12 Apolone G, Bertolini G, D'Amico R, Iapichino G, Cattaneo A, De Salvo G, et al. The performance of SAPS II in a cohort of patients admitted to 99 Italian ICUs: results from GiViTI. Gruppo Italiano per la Valutazione degli interventi in Terapia Intensiva. Intensive Care Med. 1996;22:1368-78.

13 Metnitz PG, Valentin A, Vesely H, Alberti C, Lang T, Lenz K, et al. Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Simplified Acute Physiology Score. Intensive Care Med. 1999;25:192-7.

14 Tan IK. APACHE II and SAPS II are poorly calibrated in a Hong Kong intensive care unit. Ann Acad Med Singapore. 1998;27:318-22.

15 Civetta JM, Colton T. How to read a medical article and understand basic statistics. In: Civetta JM, Taylor RW, Kirby RR, editors. Critical care. 3rd edition. Philadelphia (PA): Lippincott-Raven Publishers; 1997. p. 3-20.

16 Zimmerman JE, Wagner DP. Prognostic systems in intensive care: how do you interpret an observed mortality that is higher than expected? Crit Care Med. 2000; 28:258-60.

17 Bastos PG, Knaus WA, Zimmerman JE, Magalhaes A Jr, Sun X, Wagner DP. The importance of technology for achieving superior outcomes from intensive care. Brazil APACHE III Study Group. Intensive Care Med. 1996; 22:664-9.

18 Glance LG, Osler T, Shinozaki T. Effect of varying the case mix on the standardized mortality ratio and W statistic: A simulation study. Chest. 2000;117:1112-7.

19 Krishnagopalan S, Johnson EW, Low LL, Kaufman LJ. Body positioning of intensive care patients: clinical practice versus standards. Crit Care Med. 2002;30: 2588-92.

20 Rapoport J, Teres D, Barnett R, Jacobs P, Shustack A, Lemeshow S, et al. A comparison of intensive care unit utilization in Alberta and western Massachusetts. Crit Care Med. 1995;23:1336-46.

Received: August 2, 2005
Accepted: September 20, 2005

Correspondence to:
Kristian Deša
Department of Anesthesiology and Intensive Care
University Hospital Rijeka
Tome Strižića 3
51000 Rijeka, Croatia
kristi@medri.hr