

Study of the role of different severity scores in respiratory ICU

Mona Mansour^a, Iman Galal^a, Enas Kassem^b

Background Scoring systems are increasingly used in the ICUs in an attempt to accurately predict the mortality outcome in critically ill patients.

Objective The performance of the Acute Physiology and Chronic Health Evaluation II (APACHE II) score, the Sequential Organ Failure Assessment (SOFA) score, and the Simplified Acute Physiology Score (SAPS) II was compared in terms of calibration and discrimination in critically ill patients admitted to the respiratory ICU.

Materials and methods Mean admission APACHE II, SAPS II, and SOFA scores were compared in 105 patients. The outcome measure was ICU mortality. The discriminatory ability of the scores was evaluated using the area under the receiver operating characteristic curve. Calibration was tested using the Hosmer–Lemeshow goodness-of-fit test.

Results The mean admission APACHE II, SAPS II, and SOFA scores were higher in nonsurvivors compared with survivors; yet, only admission SOFA score differed significantly. There was highly significant positive correlation between the three scores. The cutoffs obtained by the receiver operating characteristic curve were 11 for APACHE II, 7.5 for SOFA, and 40 for SAPS II score. Discrimination power of the three scores was poor; yet, in the order of best

discrimination, SOFA [area under the curve (AUC) = 0.63] was followed by APACHE II (AUC = 0.60) and then SAPS II (AUC = 0.59). In terms of calibration, SAPS II ($\chi^2 = 4.82$; $P = 0.78$) had the best calibration and APACHE II ($\chi^2 = 7.34$; $P = 0.39$) had the worst. Logistic regression analysis showed that, of the three scores, only the SOFA score was an independent predictor of mortality among the respiratory ICU patients; with a unit increase in the SOFA score, there was a 1.2 times higher risk for mortality.

Conclusion The SOFA score performed well in terms of calibration, whereas the SAPS II score performed well in terms of discrimination. The APACHE II score did not perform well in terms of calibration and had poor discrimination power. *Egypt J Broncho* 2013 7:55–59 © 2013 Egyptian Journal of Bronchology.

Egyptian Journal of Bronchology 2013 7:55–59

Keywords: APACHE II, ICU, respiratory ICU, SAPS II, SOFA

^aDepartment of Pulmonary Medicine, Faculty of Medicine, Ain Shams University and ^bAbbassia Chest Hospital, Cairo, Egypt

Correspondence to Iman Galal, Department of Pulmonary Medicine, Faculty of Medicine, Ain Shams University, Egypt, Post/zip code: 11341; Tel: 01001502179
e-mail: dr.imangalal@gmail.com

Received 1 October 2013 **Accepted** 27 October 2013

Introduction

The high-complexity features of the ICU services and the clinical situation of patients themselves render correct prognosis, fundamentally important not only for patients, their families, and physicians, but also for hospital administrators, fund providers, and controllers [1].

The severity scoring systems were first introduced for critically ill patients in ICUs in 1980. The basis for their development was the intention to provide information on the prognosis of patients, on the efficacy of therapeutic interventions, on stratification for clinical studies, on workload, and on benchmarking of ICUs [2].

Over the last three decades, several scoring systems have been developed; the Acute Physiology and Chronic Health Evaluation (APACHE) [3,4] and the Simplified Acute Physiology Score (SAPS) [5] scores are the most widely used scoring systems in the ICU.

The APACHE II score consists of patient's age, chronic health condition, and physiological variables. Although APACHE II was one of the first systems described, it

is still the most widely used of them, insofar as the data required for its calculation are simple, well defined, reproducible, and can be collected on a routine basis during intensive care service provision [1].

The development of the Sepsis-Related Organ Failure Assessment (SOFA) score was an attempt to objectively and quantitatively describe the degree of organ dysfunction over time and to evaluate morbidity in ICU patients with sepsis [6]. Later, when it was realized that it could be applied equally well in nonseptic patients, the acronym 'SOFA' was taken to refer to Sequential Organ Failure Assessment [7]. The SOFA scoring scheme daily assigns 1–4 points to each of the following six organ systems depending on the level of dysfunction: Respiratory, circulatory, renal, hematology, hepatic, and central nervous system.

SAPS II is a standardized and internationally accepted system to assess the severity and prognosis in patients hospitalized in the ICU. Twelve acute physiological variables are scored, besides age, admission type, and the presence of a chronic disease. The final score, converted through a logistic regression equation into

probability of hospital mortality, results from the sum of the variable scores, with higher scores corresponding to more severe patient conditions [5].

This study aimed at evaluating the performance of different scoring systems, in terms of calibration and discrimination, and predicting the patients' outcome in the respiratory ICU (RICU).

Materials and methods

In this prospective observational study, all consecutive patients admitted to the RICU of Abbassia Chest Hospital were enrolled. In all, 105 patients were studied prospectively. Demographic data, admission diagnosis of the patients, comorbidities, and outcome were recorded. For all patients, APACHE II, SOFA, and SAPS II scores were determined on the day of admission to the RICU. The study was approved by the institutional ethics committee.

Statistical analysis

Parametric data were expressed as minimum, maximum, and mean \pm SD, and nonparametric data were expressed as number and percentage of the total. Student's *t*-test was used for comparing the averages of continuous measurements. Correlation between two studied parameters was determined using Pearson's correlation coefficient. The predictive capability of the three scores at the best cutoffs was assessed using the receiver operating characteristic (ROC) curve. Discrimination was tested using the ROC curves and by evaluating areas under the curve (AUC). Observed and predicted mortality was compared using the Hosmer–Lemeshow goodness-of-fit test, in which lower χ^2 values and higher *P* values (>0.05) indicate good fit. Stepwise logistic regression analysis was used to estimate the predictive ability of the APACHE II, SOFA, and SAPS II scoring systems in assessing outcome. The dependent variable was mortality and the potential independent variables were APACHE II, SOFA, and SAPS II. Statistical significance was set at *P* value less than 0.05. Statistical analyses were performed using Statistical Package for Social Sciences software (SPSS for Windows, version 16.0; SPSS Inc., Chicago, Illinois, USA).

Results

The study enrolled a total of 105 patients; the mean \pm SD age was 54.59 ± 15.75 years with a range of 20–88 years, 71 (68%) were male patients and 34 (32%) were female patients. On admission to the RICU, the mean \pm SD of APACHE II, SOFA, and SAPS II scores was 16.07 ± 7.31 , 4.95 ± 2.49 , and 41.17 ± 11.93 , respectively. Descriptive data of the included patients are displayed in Table 1. The mean \pm SD admission SOFA score

was the only score that differed significantly between the survivors and nonsurvivors (4.95 ± 2.49 , 6.11 ± 2.76 ; *P* = 0.028, respectively). Although the mean \pm SD admission scores were not significantly different between the survivors and nonsurvivors for APACHE II (16.07 ± 7.31 , 18.77 ± 7.55 ; *P* = 0.07, respectively) and SAPS II (41.17 ± 11.93 , 46.23 ± 15.37 ; *P* = 0.068, respectively), they were higher in nonsurvivors compared with survivors (Table 2). ROC curve was constructed for each score with respect to the outcome, and accuracy and measure of the AUC were obtained. The efficacy of various scores to discriminate between the survivors and nonsurvivors, as assessed by the AUC, is given in Table 3 and Fig. 1. All the scores tested

Table 1 Demographics and characteristics of the patients

Age ^a (years)	54.59 \pm 15.75 (20–88)
Sex (M/F) (%)	68/32
Admission APACHE II ^a score	16.07 \pm 7.31 (4–42)
Admission SOFA ^a score	4.95 \pm 2.49 (1–15)
Admission SAPS II ^a score	41.17 \pm 11.93 (18–90)
Diagnosis on admission	
COPD exacerbation	23
Community acquired pneumonia	16
IPF	8
Pleural disease	10
Bronchiectasis	16
Tuberculosis	19
Others	13
Total	105
Mortality [<i>n</i> (%)]	58 (55.2)

APACHE, Acute Physiology and Chronic Health Evaluation; COPD, chronic obstructive pulmonary disease; IPF, interstitial pulmonary fibrosis; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure Assessment; ^aData in parentheses represent range.

Table 2 Comparison between the admission scores of survivors and nonsurvivors

Scoring system	Survivors (<i>N</i> = 47)	Nonsurvivors (<i>N</i> = 58)	<i>P</i>
	Mean \pm SD	Mean \pm SD	
APACHE II	16.07 \pm 7.31	18.77 \pm 7.55	0.07
SOFA	4.95 \pm 2.49	6.11 \pm 2.76	0.028
SAPS II	41.17 \pm 11.93	46.23 \pm 15.37	0.068

APACHE, Acute Physiology and Chronic Health Evaluation; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure Assessment.

Table 3 Predictive probability of APACHE II, SOFA, and SAPS II scores

Parameters	APACHE II	SOFA	SAPS II
Cutoff	11	7.5	40
Sensitivity (%)	93	34	53.4
Specificity (%)	24.1	90	61.7
Accuracy (%)	55.2	64.7	57.1
AUC	0.60	0.63	0.59
<i>P</i>	0.091	0.028	0.11

APACHE, Acute Physiology and Chronic Health Evaluation; AUC, area under the curve; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure Assessment.

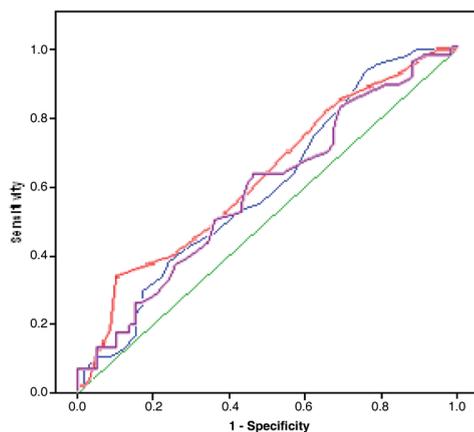
had poor discrimination power with AUC less than 0.7; yet, SOFA score performed better (AUC = 0.63) than APACHE II (AUC = 0.60) and SAPS II scores (AUC = 0.59). The cutoff points obtained by the ROC curve simultaneously considered the best sensitivity and specificity with respect to the addressed variable (Table 3 and Fig. 1). On comparing the actual and expected hospital mortality of the three scores, it was found that APACHE II correctly predicted 83% of survivors and 30% of nonsurvivors with overall 59% predictability, SOFA correctly predicted 74% of survivors and 40.1% of nonsurvivors with overall 59% predictability, and SAPS II correctly predicted 83% of survivors and 28% of nonsurvivors with overall 58% predictability. Logistic regression analysis showed that, of the three scores, only SOFA score was an independent predictor of mortality among the RICU patients; with a unit increase in the SOFA score, there was a 1.2 times higher risk for mortality. Using the Hosmer–Lemeshow goodness-of-fit test for evaluating the calibration of the various scoring systems, it was found that SAPS II ($\chi^2 = 4.82$), with $P = 0.78$, had the best calibration and APACHE II ($\chi^2 = 7.34$), with $P = 0.39$, had the worst (Table 4), suggesting that SAPS II score had the least statistically significant discrepancy between predicted and observed mortality. There was highly significant

Table 4 Hosmer–Lemeshow goodness-of-fit tests for evaluating the calibration of the scoring systems

Scoring system	χ^2	P
APACHE II	7.34	0.39
SOFA	7.2	0.3
SAPS II	4.82	0.78

APACHE, Acute Physiology and Chronic Health Evaluation; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure Assessment.

Fig. 1



ROC curve of APACHE II, SOFA, and SAPS II scores. APACHE II, blue line; SOFA, red line; and SAPS II, violet line. APACHE II, Acute Physiology and Chronic Health Evaluation II; ROC, receiver operating characteristic; SAPS II, Simplified Acute Physiology Score II; SOFA, Sequential Organ Failure Assessment.

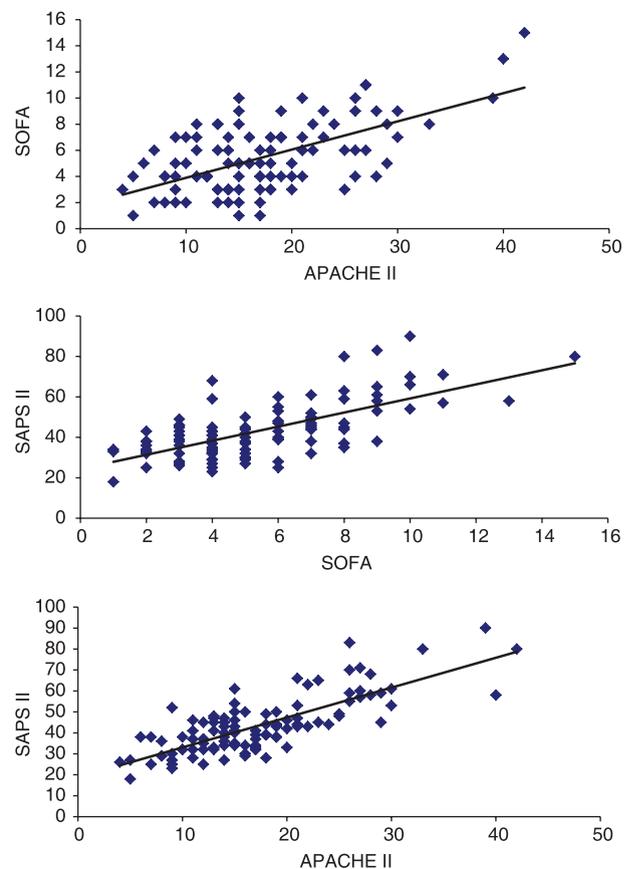
positive correlation between the various scoring systems ($P < 0.01$) as assessed by linear regression analysis. The closest correlation was observed between APACHE II and SAPS II scores ($r^2 = 0.78$) followed by SOFA and SAPS II scores ($r^2 = 0.68$), whereas the least correlation was observed between APACHE II and SOFA scores ($r^2 = 0.61$) (Fig. 2).

Discussion

In our study, we determined the initial scores of APACHE II, SOFA, and SAPS II during the first 24 h of admission to the RICU. We further compared the performance of these three scores with respect to their calibration and discrimination. The outcome measure was ICU mortality. This study has the advantage of evaluating these scores in the RICU, which was rarely tested in previous studies; instead, general and surgical ICUs were mostly the environment under test.

APACHE II [8] and SAPS II [5] are certainly among the most commonly used and validated tools for predicting outcome in the ICUs. An ideal scoring system should be able to predict mortality rate correctly,

Fig. 2



Correlation between the three scores. APACHE II, Acute Physiology and Chronic Health Evaluation II; SAPS II, Simplified Acute Physiology Score II; SOFA, Sequential Organ Failure Assessment.

that is predicted mortality rate should be close to the actual mortality rate; should be well calibrated, that is it should be able to provide risk estimate corresponding to the observed mortality; should have high levels of discrimination, that is it should be able to identify the patients who are at higher risk of dying; should be easy to compute; and should be based on easily available patient parameters. Furthermore, it also has to be dynamic, reflecting the change in management and case mix over time [9].

In our study, the mean admission APACHE II, SAPS II, and SOFA scores were higher in nonsurvivors compared with survivors; yet, only admission SOFA score differed significantly between survivors and nonsurvivors. Similarly, several studies reported that admission APACHE II [10–12] score and SAPS II [11] score did not differ significantly between survivors and nonsurvivors and did not have an influence on the risk for mortality in ICU patients. Furthermore, some studies [13–15] were in agreement with our findings with respect to admission SOFA score that differed significantly between survivors and nonsurvivors. Yet, owing to the difference in the study population with respect to the type of ICU under study and the diagnosis on admission besides the knowledge that in some studies the repetitive daily scores were used instead of the admission scores, some studies reported that APACHE II [16,17] and SAPS II [18] score differed significantly between survivors and nonsurvivors.

The performance of the prognostic models encompasses two objective measures: Calibration and discrimination [19]. Calibration refers to how closely the estimated probabilities of mortality correlate with the observed mortality over the entire range of probabilities. Discrimination refers to how well the model discriminates between individuals who will live and those who will die. The study showed that APACHE II score had the poorest calibration power, SAPS II score showed good calibration, and the calibration power of SOFA score was intermediate between APACHE II and SAPS II scores. The overall discriminatory capability of all the three scoring models, as measured by the AUC of ROC, was generally poor; yet, it was better for the SOFA score compared with the APACHE II and SAPS II scores. Similarly, in a study by Halim *et al.* [20], the discrimination power of SOFA score was better than that of APACHE II score. In a recent study by Sakr *et al.* [21], calibration was worst for APACHE II score compared with SAPS II score, which showed good calibration. Although the APACHE II score carried the highest sensitivity at the selected cutoff, the specificity of this score was very low, whereas the

specificity and the overall accuracy of the SOFA score was the highest among the three scores. Although the APACHE II index was not developed for assessing individual prognoses, ICU physicians and medicine as a whole have yearned for such predictive ability. Thus, many studies have attempted to assess the use of this index with this purpose in mind [22]. Accordingly, if the utility of APACHE II score for assessing patients' outcome in ICU is mandatory, it would be more reliable to combine this score with other scores for more accurate results.

When the three scores were tested, only SOFA score was found to be an independent predictor of mortality among the RICU patients; with a unit increase in the SOFA score, there was a 1.2 times higher risk for mortality.

The correlation between the three scores was significantly positive. A similar strong positive correlation was found between admission SOFA and APACHE II scores in other studies [15,23]. This significant positive correlation observed in our study might suggest that the overall performance of combining these scores can improve the accuracy of individual scores.

Our study tested these scores in the RICU; differences in the ICU types, ethnicity, pattern of disease, critical care offered to patients, and admission criteria might lead to different results.

The study has some limitations. The small sample size is the most important limitation, as it might influence the evaluation of calibration and discrimination of the scores. Furthermore, repetitive scores were lacking in this study. Finally, no follow-up data of the patients discharged from the RICU were available. It can be concluded that SOFA score has better discriminatory power, whereas SAPS II score has better calibration. These findings were not surprising on the basis of the understanding that it is impossible for any model to have perfect calibration and discrimination at the same time [24]. Yet, more studies are needed on a larger number of patients to support our findings.

Acknowledgements

Conflicts of interest

None declared.

References

- 1 Chiavone PA, dos Santos Sens YA. Evaluation of APACHE II system among intensive care patients at a teaching hospital. *Sao Paulo Med J* 2003; **121**:53–57.
- 2 Schusterschitz N, Joannidis M. Predictive capacity of severity scoring systems in the ICU. *Contrib Nephrol* 2007; **156**:92–100.

- 3 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**:591–597.
- 4 Vassar MJ, Lewis FR Jr, Chambers JA, Mullins RJ, O'Brien PE, Weigelt JA, *et al.* Prediction of outcome in intensive care unit trauma patients: A multicenter study of Acute Physiology and Chronic Health Evaluation (APACHE), Trauma and Injury Severity Score (TRISS), and a 24-h intensive care unit (ICU) point system. *J Trauma* 1999; **47**:324–329.
- 5 Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *J Am Med Assoc* 1993; **270**:2957–2962.
- 6 Vincent J, De Mendonça A, Cantraine F, Moreno R, Takala J, Suter P, Sprung C. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. Working group on 'sepsis related problems' of the European Society of Intensive Care Medicine. *Crit Care Med* 1998; **26**:1793–1800.
- 7 Vincent J, Ferreira F, Moreno R. Scoring systems for assessing organ dysfunction and survival. *Crit Care Clin* 2000; **16**:353–366.
- 8 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; **13**:818–829.
- 9 Juneja D, Singh O, Nasa P, Dang R. Comparison of newer scoring systems with the conventional scoring systems in general intensive care population. *Minerva Anestesiol* 2012; **78**:194–200.
- 10 Zilberberg MD, Epstein SK. Acute lung injury in the medical ICU: Comorbid conditions, age, etiology and hospital outcome. *Am J Respir Crit Care Med* 1998; **157**:1159–1164.
- 11 Arabi Y, Venkatesh S, Haddad S, Al Shimemeri A, Al Malik S. A prospective study of prolonged stay in the intensive care unit: Predictors and impact on resource utilization, prospective, mixed ICU, >14 days. *Int J Qual Health Care* 2002; **14**:403–410.
- 12 Apostolopoulou E, Nikoloudi P, Georgoudi E. Outcome in ICU patients with nosocomial infections. *ICUs Health Sci J* 2003; **14**:1–9.
- 13 Pettilä V, Pettilä M, Sarna S, Voutilainen P, Takkunen O. Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. *Crit Care Med* 2002; **30**:1705–1711.
- 14 Yıldız T, Gündoğuş B, Ateş G, Akyıldız L, Çelik Y, Topçu F, *et al.* The effectiveness of scoring systems and various biochemical parameters in predicting survival in a respiratory intensive care unit. *Turk Biochem* 2010; **35**:128–132.
- 15 Shrestha GS, Gurung R, Amatya R. Comparison of acute physiology, age, chronic health evaluation III score with initial sequential organ failure assessment score to predict ICU mortality. *Nepal Med Coll J* 2011; **13**:50–54.
- 16 Rajnish G. Performance evaluation of APACHE II scores for an Indian patient with respiratory problems. *Indian J Med Res* 2004; **119**:273–282.
- 17 Han-chung HU, Chung-chi H, Ying-Huang T. Outcome analysis of patients requiring mechanical ventilation with severe community acquired pneumonia and identified bacterial pathogens. *Chang Gung Med J* 2005; **28**:229–236.
- 18 Gupta D, Ramanathan P, Aggarwal N, Jindal K. Assessment of factors predicting outcome of acute respiratory distress syndrome in North India. *Respirology* 2001; **6**:125–130.
- 19 Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000; **19**:453–473.
- 20 Halim DA, Murni TW, Redjeki IS. Comparison of Apache II, SOFA, and Modified SOFA scores in predicting mortality of surgical patients in intensive care unit at Dr Hasan Sadiki General Hospital. *Crit Care Shock* 2009; **12**:157–169.
- 21 Sakr Y, Krauss C, Amaral AC, Réa-Neto A, Specht M, Reinhart K, Marx G. Comparison of the performance of SAPS II, SAPS 3, APACHE II, and their customized prognostic models in a surgical intensive care unit. *Br J Anaesth* 2008; **101**:798–803.
- 22 Wong DT, Crofts SL, Gomez M, McGuire GP, Byrick RJ. Evaluation of predictive ability of APACHE II system and hospital outcome in Canadian intensive care unit patients. *Crit Care Med* 1995; **23**:1177–1183.
- 23 Chen YC, Tian YC, Liu NJ, Ho YP, Yang C, Chu YY, *et al.* Prospective cohort study comparing sequential organ failure assessment and acute physiology, age, chronic health evaluation III scoring systems for hospital mortality prediction in critically ill cirrhotic patients. *Int J Clin Pract* 2006; **60**:160–166.
- 24 Diamond GA. What price perfection? Calibration and discrimination of clinical prediction models. *J Clin Epidemiol* 1992; **45**:85–89.