Evaluation of Acute Physiology and Chronic Health Evaluation II and sequential organ failure assessment scoring systems for prognostication of outcomes among Intensive Care Unit’s patients

ABSTRACT

Context: Acute Physiology and Chronic Health Evaluation II (APACHE II) and sequential organ failure assessment (SOFA) are of the most validated and prevalent general scoring systems over the world.

Aims: The aim of the current study was to evaluate APACHE II and SOFA ability in predicting the outcomes (survivors, nonsurvivors) in surgical and medical Intensive Care Unit (ICU).

Setting and Design: This was an observational and prospective study of 300 consecutive patients admitted in surgical and medical ICU during a 6-month period.

Materials and Methods: APACHE II and SOFA scores and demographic characteristics were recorded for each patient separately in the first admission 24 h.

Statistical Analysis Used: Receiver operator characteristic (ROC) curves, Hosmer–Lemeshow test, and logistic regression were used in the statistical analysis (95% confidence interval).

Results: Data analysis showed a significant statistical difference in APACHE II and SOFA scores between survivor and nonsurvivor patients ($P < 0.0001$, $P = 0.001$; respectively). The discrimination power was acceptable for APACHE II and poor for SOFA (area under ROC [AUC] curve: 73.7% (standard error [SE]: 3.2%), 63.4% [SE: 3.6%]; respectively). The acceptable calibration was seen just for SOFA ($\chi^2 = 11.018$, $P = 0.051$).

Conclusions: Both APACHE II and SOFA showed good predictive accuracy for results in surgical and medical ICUs; however, the SOFA is the choice to select, because of being simpler and easier to record data.

Key words: Acute Physiology and Chronic Health Evaluation II; Intensive Care Unit; nonsurvivor; sequential organ failure assessment; survivor

Introduction

Severities of illness scoring systems are used in Intensive Care Units (ICU) for about 35 years. These models have different strengths and weaknesses, and currently, there is no ideal scoring system. The advantages of using Acute...
Physiology and Chronic Health Evaluation II (APACHE II) score (introduced in 1985) are its ease of use and the fact that it has been used for a long period of time, which enables different comparisons to be made both within and between units.[32] This system consists of three components: Twelve physiological variables along with previous state of patient’s health and age. The maximum score is 71, points of 25 or less denote <50% mortality rate, while points of 35 or more denote more than 80% mortality.[31] Sequential organ failure assessment (SOFA) is one of the most widely used scoring systems, introduced in 1996 and is based on six different scores, one each for respiratory, cardiovascular, hepatic, coagulation, renal and neurological systems. Scores <9 give predictive mortality at 33% while above 99 can be close to or above 95%.[4] Reliability and validity of two models are obtained in several studies.[5,6] In spite of that, there are still conflicting data concerning which of this two scoring systems is the best predictor model. External validation is essential before application of scoring systems in the group of patients who are different from that group originally used for model development.[7]

Adam et al.[8] evaluated the predictive power of the APACHE II and SOFA scoring system in 39 patients who were suffering from acute pancreatitis admitted to the ICU. SOFA scores correlated significantly with mortality rate. All patients with SOFA score ≥11 at any time during ICU stay had higher mortality rate (80% sensitivity, 79% specificity, receiver operator characteristic [ROC] = 0.837). There was not any statistically significant association between APACHE II scores and mortality, but higher SOFA score predicted higher ICU/hospital mortality. Chen et al.[9] in a retrospective study, assessed the effectiveness of SOFA and APACHE II scores at the onset of bacteremia in predicting the mortality of patients with Acinetobacter baumannii bacteremia. Goodness-of-fit (GOF) was good for APACHE II and SOFA, and both models displayed excellent area under ROC (AUROC) curves (APACHE II: 0.8 ± 0.08, SOFA: 0.83 ± 0.06 in predicting 14-day mortality; APACHE II: 0.81 ± 0.04, SOFA: 0.85 ± 0.04 in predicting in-hospital mortality).

Gursel and Demirtas.[10] in a prospective observational cohort study examined the prognosis of 63 patients with ventilator-acquired pneumonia by APACHE II and SOFA. Mortality rate was 54%. For APACHE II, discrimination was excellent (ROC area under curve [AUC]: 0.81; P = 0.001) and for SOFA acceptable (ROC AUC: 0.71; P = 0.005). Just APACHE II >16 was an independent predictor of the mortality (odds ratio: 5; 95% confidence interval: 1.3-18; P = 0.019) in the logistic regression analysis. Qiao et al.[11] assessed APACHE II and SOFA in predicting mortality outcome in 106 critically ill elderly patients (aged >65 years). The area under the ROC curve was 0.76 for the APACHE II score and ranged from 0.74 for the initial SOFA score to 0.98 for the maximum SOFA score. Hosmer-Lemeshow values for the APACHE II score and various SOFA scores indicated that predictions based on these scores closely fit the observed outcomes. However, Cerro et al.[11] after performing two cohort study, noted there is no consistent performance for calibration and discrimination of these two models.

Therefore, it is recommended that regular re-calibration of scoring systems should be undertaken to provide a well-validated model to predict mortality. The aim of this study was to evaluate the prognostic accuracy of APACHE II and SOFA in predicting outcomes in surgical/medical ICUs.

Materials and Methods

Design

It was a prospective observational cohort study of patients from July 2014 to January 2015.

Population

The study population included 300 consecutive patients admitted to surgical and medical ICUs. Excluded from the study population were patients with a length of ICU stay <24 h as APACHE II cannot be calculated in these group of patients.

Data collection

Data collection included demographic information (including gender and age), Glasgow Coma Scale (GCS), preexisting underlying disease, 12 common physiological laboratory values, and six different scores for body systems, necessary for computing severity of illness as assessed by APACHE II and SOFA score. Patients’ privacy maintained by not publishing identifying information. Based on the worst data from the first 24 h after admission to surgical/medical ICU, a mark adjusting for chronic health problems and a mark for age, APACHE II was calculated. APACHE II included 12 physiologic variables (mean arterial blood pressure, heart rate, temperature, oxygenation, respiratory rate, arterial PH, serum sodium, potassium and creatinine, hematocrit, white blood cell count, and GCS), chronic health evaluation, and age adjustment score. Each variable is weighted from 0 to 4 score, with higher scores denoting an increasing deviation from normal. It is measured during the first 24 h of ICU admission. SOFA score consists of six parts scores including: Respiratory, cardiovascular, renal, liver, coagulation, and neurological. Each system takes 1-4 scores, so total score for SOFA will be 6 to 24. The higher APACHE II or SOFA scores indicate the higher probability of mortality rate. APACHE II and SOFA scores calculated after the first 24 h of admission, and then the relationship between
patient outcomes and these scores studied. Data were recorded initially on a standardized data collection form for APACHE II and SOFA and then transferred to SPSS statistical software (IBM Corp., Released 2013, IBM SPSS Statistics for Windows, Version 22.0, Armonk, NY).

**Intervention**

There was no intervention in this study.

**Outcome measures**

The primary outcomes for this investigation were survivors and nonsurvivors.

**Data analysis**

In this study, patients who were transferred from ICU to other wards of the hospital were included in the survivor group. After encoding the data using SPSS statistical software version 22 (Copyright IBM corporation and other(s) 1989-2013), simple descriptive statistics were used to summarize the study population characteristics. Data for continuous variables are presented as means with standard deviations and Frequencies with percentages are used for categorical data. The association between APACHE II and SOFA with patient’s outcomes was assessed by logistic regression. Hence, APACHE II and SOFA were considered as independent continuous variables. £ 0.05 was considered to be significant. Validations of two models were performed using standard tests to measure discrimination and calibration. The power to distinguish between survivors and nonsurvivors (discrimination) was assessed by calculating the AUC of the ROC curve. An AUC of 0.5 is equivalent to random chance (a diagonal line), AUC >0.7 indicates a moderate prognostic model, and AUC >0.8 (a bulbous curve) indicates a good prognostic model. An agreement between individual probabilities and actual outcomes (calibration) was assessed using the Hosmer–Lemeshow GOF test and P > 0.05 was considered as well-calibrated.

**Results**

A total of 300 patients admitted on surgical/medical ICUs were evaluated. The median age of the cohort was 52.74 ± 26.14 years (range 2-91 years), which 185 (61.7%) were men, and 115 (38.3%) were women. The overall mortality rate for all subjects was 27.3% (82 patients). The mean age of patients was 14.26 ± 74.52, which were 185 men (7.61%) and 115 females (3.38%), respectively. The characteristics of the study population are shown in Table 1.

For the entire cohort of patients, APACHE II and SOFA scores, age and sex were significantly different between the survivors and nonsurvivors. The mean age of the survivors (47.59 ± 25.60) was significantly different from the nonsurvivors (66.41 ± 22.50), P < 0.0001; nonsurvivors showed significantly higher APACHE II scores than survivors. The mean APACHE II score for nonsurvivors was 21.02 ± 6.71 compared with 14.93 ± 6.02 for survivors, P < 0.0001; also nonsurvivors had significantly higher SOFA scores (6.18 ± 2.04) compared to survivors (5.28 ± 2.10), P = 0.001. 79.5% of men and 61.7% of women included in the survivor group, it was statistically significant (P = 0.001).

The performance of two scoring systems has compared in Table 2.

The discrimination power was weak for SOFA but it was acceptable for APACHE II (AUC = 0.634 vs. AUC = 0.737, respectively). The best Youden index (sensitivity + specificity −1) was used to determine the best cut-off score point for both scoring models. By cut-off score 13.5, APACHE II predicted ICU mortality with a sensitivity of 89%, a specificity of 45%, and accuracy of 57%, with an AUROC curve of 0.737 ± 0.032 standard error (SE) (95%; 0.675-0.800, P < 0.0001). For SOFA, a cut-off score

**Table 2: Comparison of APACHE II and SOFA scores between survivors and nonsurvivors**

| Variables   | Cut-off score | Sensitivity (%) | Specificity (%) | Youden index | Accuracy (%) | Area under ROC curve | SE | Significant value (%) | Positive predictive value (%) | Negative predictive value (%) |
|-------------|---------------|-----------------|-----------------|--------------|--------------|-----------------------|----|-----------------------|-------------------------------|------------------------------|
| APACHE II   | 13.5          | 89              | 45              | 0.34         | 57           | 0.737 (0.675-0.800)   | 0.032| 0.000                | 37.82                         | 91.59                        |
| SOFA        | 5.5           | 57.3            | 67              | 0.24         | 64.3         | 0.634 (0.565-0.704)   | 0.036| 0.000                | 39.50                         | 80.66                        |

© Copyright IBM corporation and other(s) 1989-2013
5.5 showed a sensitivity of 57.3%, a specificity of 67%, and accuracy of 64.3%, also the area under the ROC curve was 0.634 ± 0.036 SE (95%; 0.565-0.704, P < 0.0001) [Table 2]. By using the Hosmer–Lemeshow Chi-Square statistic, the SOFA showed good calibration (χ² = 11.08, P = 0.051) but calibration power for APACHE II was weak (χ² = 66.633, P = 0.000). ROC curves were drawn for the APACHE II and SOFA to assess predictive accuracy [Figure 1].

**Discussion**

In this study, two predictive scoring systems have been evaluated in the surgical/medical ICUs. The mean APACHE II and SOFA scores were significantly higher in nonsurvivors when compared to survivors (P < 0.0001 and P = 0.001, respectively). The acceptable capability of discriminating survivors from nonsurvivors obtained by APACHE II, with AUROC curve of 0.737, the discrimination power of SOFA was slightly weaker than APACHE II (ROC = 0.634). The cut-off score for APACHE II and SOFA was 13.5 and 5.5, respectively, and both models showed acceptable overall accuracy. Based on AUCs, there was a difference between discrimination powers of two models, it may arise from case-mix and need for short-term or long-term cares. Better calibration obtained for SOFA compared to APACHE II based on Hosmer–Lemeshow test (χ² = 11.018, P = 0.051; χ² = 66.633. P < 0.0001, respectively); It might be explained by the suitability of SOFA in long-term ICU cares. Based on findings in this study predictive accuracy of SOFA was slightly better than APACHE II also younger and male patients had more chance to be in the survivors group.

The findings of our study are in agreement with several studies have been cited that higher APACHE II and SOFA scores were significantly associated with higher mortality rate or poor prognosis. According to Balci et al., study results, both models were found to be significant predictors of outcomes and might be used in the prediction of mortality in septic patients. The mortality rate in their study was 25% (near to ours). By using univariate logistic regression in Sawicka et al., study, SOFA (P = 0.00009) and APACHE II (P = 0.0007) identified as risk factors of death in patients with hematological malignancies. Türe et al., in a prospective, observational study analyzed and compared the prognostic accuracy of APACHE II and SOFA scoring systems in predicting ICU mortality in 206 patients who had acute respiratory distress syndrome. Mortality rate was 52.4%. The survivors had a lower APACHE II score (11.50 vs. 15.82, P < 0.0005), a lower SOFA score (6.06 vs. 9.42, P < 0.0005), and a younger age (57 vs. 70 years, P = 0.008); in their study, a cut-off score for APACHE II and SOFA was about our values, also similar to our results a younger age was independent risk factor for mortality. In Milic et al., study, APACHE II score on admission had no value for predicting length of stay (LOS) in the cardio surgical ICU (similar to our results), but SOFA significantly correlated with LOS in ICU (r = 0.258).

There are many studies have indicated to correlation between high scores in APACHE II and SOFA and mortality rates in ICUs. However, some studies have pointed to the inability of these tools in predicting outcomes. For example: Desai and Lakhani, conducted a prospective study on patients with sepsis. They concluded Serial measurement of SOFA score during 1st week is very useful tool in predicting the outcome but the APACHE II score on day of admission was not reliable in predicting the mortality rate in this study and believed that it may need modification in set up like theirs.

In this study, the discrimination power of APACHE II and SOFA based on AUC-ROC was acceptable and weak, respectively; and these models based on Hosmer–Lemeshow test had weak and good calibration power, respectively. Inconsistent with our results, the discrimination power of APACHE II and SOFA in most studies was acceptable or excellent, and calibration power of these models was varied. In Chen et al., study calibration was good for APACHE II and SOFA, and both models displayed excellent AUROCs. They found that SOFA >8, APACHE II >29 and SOFA >7, APACHE II >23 are associated with significantly higher 14-day and in-hospital mortality rates, respectively. Also they noted for ease of calculation, the use of SOFA rather than APACHE II score to predict mortality of A. baumannii bacteremia might have clinical application. In another study, Hantke et al., studied 874 surgical ICU patients to compare APACHE II and SOFA by ROC analyses. The ROC analyses of APACHE II and
SOFA scores were comparable (0.73 and 0.71, respectively). The discrimination power of APACHE II in their study was similar but for SOFA, it was better than our results; also they cited the SOFA score is reliable and might be useful in the daily routine of an ICU. Kellner et al. in their prospective, observation study determined that the mean APACHE II ($P = 0.035$), and SOFA ($P = 0.042$) scores were significantly higher in nonsurvivors than in survivors. In consistent with our results, the AUC for APACHE II was 0.726, but SOFA did not yield valuable results at the maximum score.

Cerro et al. to validate the APACHE II and SOFA scores in patients with suspected infection in clinical settings other than ICUs conducted a two cohort studies on 2530 adult patients. In the first cohort, the AUC-ROC values for mortality at discharge and on day 28 were around 0.50 for the APACHE II and SOFA scores; whereas for the second cohort, the discrimination value was around 0.70. The calibration of both scoring systems for primary outcomes, according to Hosmer–Lemeshow test, showed $P > 0.05$ in the first cohort; while in the second cohort calibration, it only showed a $P > 0.05$ in the case of the SOFA for mortality at hospital discharge. Their studies showed no consistent performance for calibration and discrimination and discrimination for these scoring systems. These discrepancies can be explained by the fact that a scoring system based on validation and testing set from one population when transferred to another population without modification will often lose its predictive accuracy. Therefore, even if initially the model discriminates well, it is possible that following improvement or deterioration in quality of cares, the performance of scoring model would change and would result in reducing applicability of the severity of illness scoring systems to the different settings. By recalibrating these models frequently, we may be overcome these problems with taking into account the changes in quality of care and improved survival.

The present study has several limitations: First, the sample size is known to have a major influence on calibration power of model. Second, there may be bias with regard to case mix (surgical/medical ICUs), quality of care and policies. A big multicenter studies would mitigate the concerns over case mix and benefit larger sample size. The nature of the population being evaluated and the quality of care can influence on discrimination power of models, so customizing the models or perhaps utilizing scoring systems specific to particular settings can improve the mortality estimations. Ethical considerations have been considered in this study.

**Conclusion**

The APACHE II and SOFA scoring systems showed acceptable and weak discrimination power, respectively; and calibration for these models was weak and good, respectively. The SOFA predicting accuracy was slightly better than APACHE II, and because of simplicity to calculation, it is an advisable scoring system to predict the outcomes of patients in surgical/medical ICUs.

**Financial support and sponsorship**

Financial support by Islamic Azad University, Bojnord Branch, Bojnord, North Khorasan Province, Iran.

**Conflicts of interest**

There are no conflicts of interest.

**References**

1. Raj R, Siironen J, Kivisaari R, Hernesniemi J, Skrifvars MB. Predicting outcome after traumatic brain injury: Development of prognostic scores based on the IMPACT and the APACHE II. J Neurotrauma 2014;31:1721-32.
2. Mbongo CL, Monedero P, Guillon-Grima F, Yepes MJ, Vives M, Echarri G. Performance of SAPS3, compared with APACHE II and SOFA, to predict hospital mortality in a general ICU in Southern Europe. Eur J Anaesthesiol 2009;26:940-5.
3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. Crit Care Med 1985;13:818-29.
4. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22:707-10.
5. Balci C, Sungurtekin H, Güres E, Sungurtekin U. APACHE II, APACHE III, SOFA scoring systems, platelet counts and mortality in septic and nonseptic patients. Ulus Travma Acil Cerrahi Derg 2005;11:29-34.
6. Gursel G, Demirtas S. Value of APACHE II, SOFA and CPIS scores in predicting prognosis in patients with ventilator-associated pneumonia. Respiration 2006;73:503-8.
7. Hosseini M, Ramazani J. Comparison of acute physiology and chronic health evaluation II and Glasgow Coma Score in predicting the outcomes of Post Anesthesia Care Unit's patients. Saudi J Anesth 2015;9:136-41.
8. Adam F, Bor C, Uyar M, Demirag K, Çankayali I. Severe acute pancreatitis admitted to intensive care unit: SOFA is superior to Ranson's criteria and APACHE II in determining prognosis. Turk J Gastroenterol 2013;24:430-5.
9. Chen SJ, Chao TF, Chiang MC, Kuo SC, Chen LY, Yin T, et al. Prediction of patient outcome from Acinetobacter baumannii bacteremia with Sequential Organ Failure Assessment (SOFA) and Acute Physiology and Chronic Health Evaluation (APACHE) II scores. Intern Med 2011;50:871-7.
10. Qiao Q, Lu G, Li M, Shen Y, Xu D. Prediction of outcome in critically ill elderly patients using APACHE II and SOFA scores. J Int Med Res 2012;40:1114-21.
11. Cerro L, Valencia J, Calle P, León A, Jaimes F. Validation of APACHE II and SOFA scores in 2 cohorts of patients with suspected infection and
sepsis, not admitted to critical care units. Rev Esp Anestesiol Reanim 2014;61:125-32.
12. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8:283-98.
13. Sawicka W, Owczuk R, Wujtewicz MA, Wujtewicz M. The effectiveness of the APACHE II, SAPS II and SOFA prognostic scoring systems in patients with haematological malignancies in the intensive care unit. Anesthesiol Intensive Ther 2014;46:166-70.
14. Türe M, Memis D, Kurt I, Pamukçu Z. Predictive value of thyroid hormones on the first day in adult respiratory distress syndrome patients admitted to ICU: Comparison with SOFA and APACHE II scores. Ann Saudi Med 2005;25:466-72.
15. Milic M, Goranovic T, Holjevac JK. Correlation of APACHE II and SOFA scores with length of stay in various surgical intensive care units. Coll Antropol 2009;33:831-5.
16. Desai S, Lakhani JD. Utility of SOFA and APACHE II score in sepsis in rural set up MICU. J Assoc Physicians India 2013;61:608-11.
17. Kim YH, Yeo JH, Kang MJ, Lee JH, Cho KW, Hwang S, et al. Performance assessment of the SOFA, APACHE II scoring system, and SAPS II in intensive care unit organophosphate poisoned patients. J Korean Med Sci 2013;28:1822-6
18. Hantke M, Holzer K, Thöne S, Schmandra T, Hanisch E. The SOFA score in evaluating septic illnesses. Correlations with the MOD and APACHE II score. Chirurg 2000;71:1270-6.
19. Kellner P, Prondzinsky R, Pallmann L, Siegmann S, Unverzagt S, Lemm H, et al. Predictive value of outcome scores in patients suffering from cardiogenic shock complicating AMI: APACHE II, APACHE III, Elebute-Stoner, SOFA, and SAPS II. Med Klin Intensivmed Notfmed 2013;108:666-74.

New features on the journal’s website

Optimized content for mobile and hand-held devices
HTML pages have been optimized of mobile and other hand-held devices (such as iPad, Kindle, iPod) for faster browsing speed.
Click on [Mobile Full text] from Table of Contents page.
This is simple HTML version for faster download on mobiles (if viewed on desktop, it will be automatically redirected to full HTML version)

E-Pub for hand-held devices
EPUB is an open e-book standard recommended by The International Digital Publishing Forum which is designed for reflowable content i.e. the text display can be optimized for a particular display device.
Click on [EPub] from Table of Contents page.
There are various e-Pub readers such as for Windows: Digital Editions, OS X: Calibre/Bookworm, iPhone/iPod Touch/iPad: Stanza, and Linux: Calibre/Bookworm.

E-Book for desktop
One can also see the entire issue as printed here in a ‘flip book’ version on desktops.
Links are available from Current Issue as well as Archives pages.
Click on View as eBook