Comparison of Apache II, SOFA, and Modified SOFA Scores in Predicting Mortality of Surgical Patients in Intensive Care Unit at Dr. Hasan Sadikin General Hospital

Dino Adrian Halim, Tri Wahyu Murni, Ike Sri Redjeki

Abstract

***Introduction*: Scoring systems were developed to assess the severity of organ failures and to predict mortality. The sequential organ failure assessment (SOFA) score and its modification (MSOFA) are gaining popularity through their proven simplicity, validity, and reliability in previous studies.**

***Objective*: To determine and compare the validity of the SOFA and MSOFA scores with the Acute Physiology and Chronic Health Evaluation II (APACHE II) score for predicting mortality in surgical patients treated in ICU in Dr. Hasan Sadikin General Hospital in Bandung, West Java, Indonesia.**

***Patients and Method*s: This was a prospective observational cohort study involving consecutively 144 surgical patients (from January 2008 to December 2008). APACHE II, SOFA, and MSOFA scores were determined on admission. SOFA and MSOFA scores were also repeated every 48-72 hours until ICU discharge or death for determining mean and maximum values of SOFA and MSOFA. Scores validation were determined using Hosmer-Lemeshow goodness-of-fit test and receiver operating characteristic (ROC) curve analyses to determine the area under the curve (AUC).**

***Results*: Mortality rate was 39.8%. The mean APACHE II score (11.63±5.55, 14.95±4.27; p≤0.001), SOFA(3.7±2.23, 5.86±2.88, p≤0.001), and MSOFA(3.98±1.95, 5.79±1.98, p≤0.001) were all higher in non-survivors than in survivors. Discrimination was less satisfactory for APACHE II (AuROC=0.69; p≤0.001) and acceptable for both initial SOFA (AuROC=0.73; p≤0.001) and initial MSOFA (AuROC=0.75; p≤0.001). Mean and maximum values of SOFA and MSOFA showed even better discrimination values with AuROC=0.92; p≤0.001, and AuROC=0.91; p≤0.001 for meanSOFA and maksSOFA respectively, and AuROC=0.90; p≤0.001, AuROC=0.90; p≤0.001 for meanMSOFA and maksMSOFA respectively.**

***Conclusion*: SOFA and MSOFA scoring systems are better than APACHE II system in predicting mortality in ICU surgical patients. Serial measurements of SOFA and MSOFA score significantly improve their predictive accuracy.**

Key words: Mortality, surgical patients, APACHE II, SOFA, MSOFA.

From Department of Surgery, Faculty of Medicine, Padjadjaran University, Dr. Hasan Sadikin General Hospital, Bandung, Indonesia (Dino Adrian Halim, Tri Wahyu Murni, and Ike Sri Redjeki).

Address for correspondence:

Dino Adrian Halim, MD

Department of Surgery, Faculty of Medicine, Padjadjaran University, Dr. Hasan Sadikin General Hospital

Jl. Pasteur 38, Bandung 40161, Indonesia

Email: liem_dino@yahoo.com

Introduction

Intensive Care Units (ICUs) are equiped with many complex life support devices, which are key to monitoring intensive procedures performed by highly trained medical personels 24 hours a day. It is obvious that ICUs consume considerable operational resources and require very high maintenance cost. However, mortality rates of critically ill surgical patients remain very high regardless of technologically advanced centers. Dr. Hasan Sadikin ICU is the provincial referral ICU for West Java province. Unfortunately, due to financial limitations, as physicians, we face daily with the responsibility to select the best candidate to receive treatments. The patients mortality remains high because of the suboptimal medical care our hospital is able to provide. As an illustration, in 2006, our ICU had a total of 554 patients with 119 deaths (21.48%) and for 2007, we had 669 patients with 141 deaths (21.07%). In 2008, there were 144 surgical patients included in this study with 57 deaths (39.58%).
One possible solution to this problem is the application of a good prognostic or predictive scoring system. Good prognosis can help predict outcome or progress in a patient. (1) Scoring systems are composed of degrees of organ dysfunction, organ failure or multiple organ failures, and anatomical derrangements which eventually contribute to morbidity and mortality. (2-4) With the help of such evaluation system, we will be able to distribute the limited resources to more suitable patients.

Scoring systems also provide several advantages like the following:

1) One common language
2) More objective evaluation
3) Better triage system
4) Better therapeutic management and patient response evaluation
5) Better medical administration, medical auditing, ICU performance evaluation, and promotion of scientific research. (2,3,5)

Scoring systems essentially consist of two parts: a severity score, which is a number (generally the higher this is, the more severe the condition) and a calculated probability of mortality. Most commonly, this is the risk of in-hospital mortality, though other outcome measures (e.g. survival to 28 days post-hospital discharge) (6) can also be modelled. In order to develop a scoring system, a database incorporating a large amount of patient data from many ICUs, and ideally from many different countries, is required. The applied variables can be grouped into five categories: age, co-morbidities, physiological abnormalities, acute diagnosis, and interventions. (3)

Many severity scores have been published but only few of them are used. Most scores are calculated from data collected during the first day in the ICU; these include the Acute Physiology and Chronic Health Evaluation (APACHE) score, Simplified Acute Physiology Score (SAPS), and Mortality Prediction Model (MPM). Others are repetitive and collect data every day throughout the ICU stay or for the first 3 days; these include the Organ System Failure (OSF), Organ Dysfunction and Infection System (ODIN), Sequential Organ Failure Assessment (SOFA), Multiple Organs Dysfunction Score (MODS), Logistic Organ Dysfunction (LOD) model, and Three-Day Recalibrating ICU Outcomes (TRIOS). (7)

The scoring system used regularly in our ICU is the APACHE II score. This system was first introduced by Knaus et al. in 1984 which is a revision of the first APACHE system (1981). A revision was made after analyzing a large physiological database of critically ill patients. This system utilized scoring points according to heuristic groupings of 12 physiologic variables, Glasgow Coma Score (GCS), age, and chronic health evaluation status The score can be determined directly upon admission or by determining the worst score obtained within the first 24 hours after admission. To predict the outcome, the risk of hospital death is computed by combining APACHE II score with Knaus' weighted coefficient for different types of disease entities. Later on, these are assigned according to the need for emergent surgery and the diagnostic category that requires ICU admission. (2,3,8,9) The APACHE II scoring system has been applied to patients with pancreatitis with or without sepsis, (10-13) acute myocardial infarction, (14) acute renal failure, (15) liver failure, (16) post liver transplant, (17) chronic obstructive pulmonary disease, (18) oropharyngeal carcinoma, (19) trauma, (20) and abdominal infection (21) with varying results. However, several other studies revealed a few weaknesses in its ability to predict outcomes in patients with sepsis and peritonitis. (22-26) APACHE II score has also performed consistently less accurate in predicting the

outcomes of trauma patients, (27-29) patients with unstable hemodynamic which require invasive cardiovascular monitoring, (30) and post cardiopulmonary bypass surgery. (9) Tanaka et al. (31) discovered no satisfying explanation for the exclusion of burned patients in the original study by Knaus et al. (9)

The APACHE II score is quite cumbersome to perform and with less satisfactory prediction accuracy in many surgical cases as described above. Therefore, we initiate this study to compare several simpler alternatives; the sequential organ failure assessment (SOFA) and modified sequential organ failure assessment (MSOFA) scores. The SOFA score was proposed and published by Vincent et al. in 1996 (32) through a heuristic consensus process and afterwards validated in a larger population of 1449 critically ill patients. The system is comprised of calculation scores from six organ systems (respiratory, cardiovascular, hepatic, coagulation, renal, and neurological) graded from 0-4 according to the consensus-derived degree of dysfunction or failure. It has been validated in a prospective multi-institutional study with an 84% correct classification. (33)

Grissom et al. (34) proposed and published a simplified version of the SOFA score known as the MSOFA score. The MSOFA score was originally developed for a fast screeing method during the Avian Influenza outbreak. It eliminates the necessity of laboratory examinations such as platelet count and substitutes measurements of P/F ratio and serum bilirubin level with S/F ratio and clinical examination for sign of jaundice. S/F ratio is obtained by dividing pulse oxymeter saturation with fraction of inspired oxygen.
These scoring systems can be very useful, simpler, and less expensive alternatives to APACHE II score to predict outcomes of surgical critical patients in the ICU of Dr. Hasan Sadikin General Hospital if they could perform equal to or even better than APACHE II score.

Materials and Methods

Study design and subject population

This prospective observational cohort study was performed in a single medical-surgical ICU of Dr. Hasan Sadikin General Hospital, a referral teaching hospital of Padjadjaran University Medical Faculty in Bandung, West Java, Indonesia.

Data collection

Data collection took place from January 2008 to December 2008. We performed consecutive sampling for all surgical patients fulfilling the study protocols, which were admitted to the ICU. Patients who were excluded from the present study were those who were younger than 14 year-old, suffered burn injuries, went home against medical advice, withdrawn from treatment, resigned from the study, died or those who stayed in the ICU less than 24 hours. Approval for the project was obtained from the hospital Ethics Committee. Among the data collected were basic demographic characteristics, which included sex and age, category of cases, which included emergency or elective, and working diagnosis leading to ICU admission. The ICU length of stay (LOS) was also determined. Patients were followed until they died or discharged from the ICU for a maximum of 10 days.

We determined the initial scores of APACHE II, SOFA, and MSOFA during the first 24 hours of ICU admission. We also performed serial score measurements every 48-72 hours until patient discharged for a maximum of 10 ICU-day for SOFA and MSOFA. These measurements were done to determine the predictive value of mean and maximum SOFA and MSOFA scores.

Statistical analysis

Subjects' characteristics were presented by descriptive method. Quantitative normally distributed variables were presented as means±standard deviation (SD) and non normally distributed variables as medians. Univariate comparison was performed to compare variables between groups. The organ dysfunction scores were compared using the Mann-Whitney test. Categorical variables were expressed as actual numbers and percentages. In all comparisons, $p<0.05$ was considered statistically significant. Chi-square test was used for the statistical significance of categorical variables. The ability of the models for predicting ICU mortality was determined by examining their discrimination power, which was tested by examining the graph of the area under the Receiver Operating Characteristics curve

(AuROC). AuROC summarizes the relationship between sensitivity (number of true positives) and 1-specificity (number of false positives) for all the possible values of the organ dysfunction scores. AuROC estimates the ability of the model to assess a higher risk of death to patients who died.

Multivariate logistic regression analysis was applied to determine the independent contribution of organ dysfunction scores to the prediction of the mortality as a dependent variable. The maximum scores for each of the six organ systems were used as the independent variables. The regression equation represents how much of the dependent variable will change with any given change of the independent variables, represented as a regression line on a scatter diagram. The constant and coefficients (β) for all variables are used in the regression equation. Constant signifies the distance above the baseline at which the regression line cuts the vertical axis. The odds ratio with 95% confidence intervals (CI) were used to estimate the association between the independent variables and the dependent variable. The Hosmer-Lemeshow statistic was used to evaluate the calibration, and correspondence between the observed and predicted mortality. Lower Hosmer-Lemeshow value and a higher p value indicate a better fit. Good fit was defined as $p>0.05$. Statistical analysis was performed using the SPSS 16.0 statistical package (SPSS Inc, Chicago, IL, USA).

Scores

The worst values of each organ failure score in the first 24 hours following ICU admission were used for the calculations. For patients who were sedated, a Glasgow Coma Score (GCS) was determined either from their medical records before sedation or through interviewing the physician who ordered the sedation.

Results

During the study period, a total of 144 eligible surgical patients were included in the study. Fifty-seven subjects died (39.58%). The age of the non-survivors were significantly older than the survivors (51.54±15.06 and 44.56±17.97 respectively). The subjects were mostly males (64.6%). All cases in this study were post operative, comprised of 92 cases of emergency surgery and 52 of elective surgery. Emergency abdominal surgery ranked first with 71 cases (49.3%) followed with emergency thoracic surgery with 14 cases (9.7%). We found that the length of hospital stay of non-survivors (11.14±10.14 days) was slightly longer than survivors (10.91±10.78 days). This was not statistically significant, probably because some patients with non progressive organ failure may stay for a longer time in the ICU and still survive, as stated by Ferreira et al. (4)

The patients' demographic characteristics, case categories, and hospital length of stay are shown in **Table 1**. APACHE II score and all subvariables of SOFA and MSOFA scores were significantly higher in non-survivors compared with the survivors. The assessment of patients' severity of illness with APACHE II, SOFA, and MSOFA scores and their respective outcomes can be seen in **Table 2**. For SOFA and MSOFA scores, we determined their initial, mean, and maximum scores during ICU stay. **Table 3** shows the discriminatory power of each score in predicting mortality. It determines their cut off points according to receiver operating characteristics curve (ROC), and then determines their sensitivity, specificity, accuracy, and AuROCs as shown in **Table 4** and **Figures 1** to **7**. The relationship between each subvariable of SOFA and MSOFA scores with their probability of death can be seen in **Tables 5** and **6**, while the relationship of APACHE II score with its probability of death is shown in **Table 7**. Analysis of each component of SOFA and MSOFA scores to determine the most significantly influencing component affecting mortality of the subjects are shown in **Tables 8** and **9**.

Discussion

The results of the present study showed that the initial score and progression of organ dysfunction were closely related to the outcome of the surgical patients admitted to ICU. The mean and maximum values of SOFA and MSOFA scores were more useful as a tool for outcome prediction compared with their initial score. High mean and maximum scores reflected the worsening organ systems dysfunction during the course in ICU which mostly occurs in non-surviving patients.

A model's discrimination (the ability of the model to distinguish patients who died from those who survive)

was assessed by numerically examining the AuROC. An AuROC of 1 means a perfect dicrimination while 0.5 is a random chance. A model is considered acceptable if the AuROC is ≥0.7 and is considered excellent if the AuROC is ≥0.9. (3,6,7) Discriminatory power in outcome prediction as estimated by the AuROC was acceptable for initial SOFA and MSOFA scores (0.732 and 0.751 respectively), but less satisfactory for APACHE II score (0.694). Serial measurements of SOFA and MSOFA scores as shown in mean and maximum subvariables of SOFA (AuROC=0.917 and AuROC=0.909 respectively) and MSOFA (AuROC=0.903 and AuROC=0.901 respectively) revealed an excellent discriminatory power.

Calibration evaluates the degree of correspondence between the estimated probabilities of mortality produced by a model and the actual mortality experienced by patients. It can be statistically evaluated using the formal Hosmer-Lemeshow goodness-of-fit tests. To test calibration formally patients are rank-ordered according to their probability of mortality and grouped into range-defined strata. Typically ten such strata are formed, each containing approximately the same number of patients (called "risk deciles"). To obtain the predicted number of deaths in a stratum, the probability of mortality for all patients in that group is summed. Formal goodness-of-fit testing compares the observed with the predicted number of deaths and the observed with the predicted number of survivors in each stratum of patients. The resulting value can be used to determine whether the combined discrepancy between observed and predicted outcome across all strata is within sampling variability. If differences are large, the model does not correctly reflect the outcome in that cohort of patients. The iSOFA score has a p value of 0.128 with χ^2 of 5.684, maxSOFA has a p value of 0.316 with χ^2 of 5.901. The iMSOFA score has a p value of 0.303 with χ^2 of 6.034, maxMSOFA has a p value of 0.068 with χ^2 of 13.189, while APACHE II score has a p value of 0.014 with χ^2 of 10.627, which means APACHE II score has a poor goodness-of-fit result.

Subscore analysis shows that all components of SOFA and MSOFA scores significantly affect outcome except for liver failure. Two most important components of the SOFA scores are the cardiovascular component ($p \leq 0.05$, OR [95% CI]=2.634) and respiratory component ($p \leq 0.05$, OR [95% CI]=2.273), which means that the increase of one point in cardiovascular and respiratory component will increase the mortality risk as high as 2.634% and 2.273% respectively. The two most important components of the MSOFA scores are also the cardiovascular ($p \leq 0.05$, OR [95% CI]=2.784) and respiratory components ($p \leq 0.05$, OR [95% CI]=4.863). The present study does have some limitations. First, it was done in a single centre, therefore case-mix and quality of ICU care may have a large influence on the results. Second, serial measurements of SOFA and MSOFA scores were taken every 48-72 hours due to limited funds instead of exactly every 48 or 72 hours, so it may affect the results.

Conclusion

SOFA and MSOFA scores with their subvariables have better discriminatory power and calibration than APACHE II score in predicting mortality of surgical patients in ICU of Dr. Hasan Sadikin General Hospital.

Acknowledgement

The authors thank Drs. Jeffrey and YP Sari for software preparation and data collection, and also Mr. H. Sukandar and Mrs. N. Fauziyah for data processing and statistical analysis.

Table 1. Characteristics of the Subjects

Variable	Outcome		Total (n=144)	p
	Survivors (n=87)	Non-survivors (n=57)		
Age				Z_{M-W}=2.25
mean (SD)	44.56 (17.97)	51.54 (15.06)	47.33 (17.17)	p=0.024
range (years)	15-85	20-81	15-85	
<30	23 (26.4%)	4 (7.0%)	27 (18.8%)	
30-39	12 (13.8%)	5 (15.8%)	17 (11.9%)	
40-49	14 (16.1%)	13 (17.5%)	27 (18.8%)	
50-59	17 (19.5%)	19 (33.3%)	36 (25.2%)	
60-69	16 (18.4%)	7 (12.3%)	23 (16.1%)	
≥70	4 (5.7%)	9 (14.0%)	13 (9.2%)	
Gender				
Male	59 (67.8%)	34 (59.6%)	93 (64.6%)	χ^2=1.004
Female	28 (32.2%)	23 (40.4%)	51 (35.4%)	p=0.316
Case category				
Emergency Thoracic	13 (14.9%)	1 (1.8%)	14 (9.7%)	χ^2=14.834
Emergency GI	37 (42.5%)	34 (59.6%)	71 (49.3%)	p=0.062
Emergency Neurosurg.	4 (4.6%)	3 (5.3%)	7 (4.9%)	
Elective Orthopedic	3 (3.4%)	2 (3.5%)	5 (3.5%)	
Elective Thoracic	9 (10.3%)	3 (5.3%)	12 (8.3%)	
Elective GI	4 (4.6%)	3 (5.3%)	7 (4.9%)	
Elective Neurosurg.	11 (12.6%)	3 (5.3%)	14 (9.7%)	
Elective Urology	2 (2.3%)	3 (5.3%)	5 (3.5%)	
Others	4 (4.6%)	5 (8.8%)	8 (6.3%)	
Hospital LOS				
mean (SD)	10.91 (11.79)	11.14 (10.15)		Z_{M-W}=0.008
range	2-75	2-46		p=0.993
Total	87 (100.0%)	57(100.0%)	144 (100%)	

Legend:
Z_{M-W}=Mann-Whitney test
χ^2=Chi square test
LOS=length of stay

Table 2. Comparison of Scores and Their Outcome

Variable	Outcome		p
	Survivors (n=87)	Non-survivors (n=57)	
iSOFA			<0.001
mean (SD)	3.70 (2.24)	5.86 (2.88)	
range	0-10	0-14	
meanSOFA			<0.001
mean (SD)	3.18 (1.77)	7.35 (2.56)	
range	0-9	2-14	
maxSOFA			<0.001
mean (SD)	4.06 (2.15)	9.32 (3.33)	
range	0-10	2-18	
iMSOFA			<0.001
mean (SD)	3.98 (1.95)	5.79 (1.99)	
range	1-9	2-11	
meanMSOFA			<0.001
mean (SD)	3.39 (1.64)	6.68 (2.07)	
range	1-9	3-13	
maxMSOFA			<0.001
mean (SD)	4.11 (1.88)	8.18 (2.64)	
range	1-9	3-16	
APACHE II			<0.001
mean (SD)	11.63 (5.56)	14.95 (4.28)	
range	2-27	4-28	

Legend: utilizing Mann-Whitney test

Table 3. Cut-off Points for Each Variable

Variable	Survivors (n=87)	Non-survivors (n=57)	χ^2	p	RR	95% CI
iSOFA			19.94	<0.001	1.88	1.38-2.56
≤4.5	59 (77.6%)	17 (22.4%)				
>4.5	28 (41.2%)	40 (58.8%)				
meanSOFA			72.04	<0.001	5.75	3.02-10.9
≤5.5	79 (86.8%)	12 (13.2%)				
>5.5	8 (15.1%)	45 (84.9%)				
maxSOFA			65.28	<0.001	3.53	2.36-5.53
≤5.5	69 (92.0%)	6 (8.0%)				
>5.5	18 (26.1%)	51 (73.9%)				
iMSOFA			25.78	<0.001	2.05	1.50-2.79
≤4.5	59 (80.8%)	14 (19.2%)				
>4.5	28 (39.4%)	43 (60.6%)				
meanMSOFA			66.16	<0.001	5.05	2.76-9.21
≤5.5	78 (85.7%)	13 (14.3%)				
>5.5	9 (17.0%)	44 (83.0%)				
maxMSOFA			69.94	<0.001	3.86	2.50-5.95
≤5.5	71 (92.2%)	6 (7.8%)				
>5.5	16 (23.9%)	51 (76.1%)				
APACHE II			20.27	<0.001	1.83	1.41-2.37
≤11.5	48 (82.8%)	10 (17.2%)				
>11.5	39 (45.3%)	47 (54.7%)				

Legend:
χ^2=Chi-square test
CI=confidence interval
RR=relative risk

Table 4. Discriminatory Power of Each Score

Variable	Se	Sp	PPV	NPV	Accuracy
iSOFA	70.2%	67.8%	58.8%	77.6%	68.8%
meanSOFA	78.9%	90.8%	84.9%	86.8%	86.1%
maxSOFA	89.5%	79.3%	73.9%	92.0%	83.0%
iMSOFA	75.4%	67.8%	60.6%	80.8%	70.8%
meanMSOFA	77.2%	89.7%	83.0%	85.7%	84.7%
maxMSOFA	89.5%	81.6%	76.1%	92.2%	85.0%
APACHE II	82.5%	55.2%	54.7%	82.8%	66.0%

Legend:
Se=sensitivity
Sp=specificity
PPV=positive predictive value
NPV=negative predictive value

Table 5. Calibration of iSOFA and maxSOFA

Score	POD (%)	Survivors (n=87)		Non-survivors (n=57)		
		Observed	Expected	Observed	Expected	
iSOFA						χ^2=5.684
0-1	0	14	14.044	3	2.956	p=0.128
2-3	7	26	24.922	7	8.078	
4-5	20	32	28.448	17	20.552	
6-7	22	10	15.408	18	12.592	
≥8	≥33	5	4.178	12	12.822	
maxSOFA						χ^2=5.901
0-1	0	10	9.440	0	0.560	p=0.316
2-3	2	22	21.336	1	1.664	
4-5	7	37	36.376	5	5.624	
6-7	18	13	14.733	12	10.267	
8-9	26	3	4.540	11	9.460	
10-11	46	2	0.572	17	18.428	
>11	86	0	0.001	11	10.999	

Table 6. Calibration of iMSOFA and maxMSOFA

Score	Survivors (n=87)		Non-survivors (n=57)		
	Observed	Expected	Observed	Expected	
iMSOFA					χ^2=6.034
0-2	14	13.262	1	1.738	p=0.303
3-4	45	42.858	13	15.142	
5-6	19	24.544	27	21.456	
7-8	7	5.074	10	11.926	
≥9	2	1.262	6	6.738	
maxMSOFA					χ^2=13.189
0-2	11	10.839	0	0.161	p=0.068
3-4	46	46.653	5	4.347	
5-6	21	19.762	8	9.238	
7-8	7	8.021	20	18.979	
9	2	1.388	11	11.612	
≥10	0	0.337	13	12.663	

Table 7. Calibration of APACHE II

Score	POD (%)	Survivors (n=87)		Non-survivors (n=57)		
		Observed	Expected	Observed	Expected	
						χ^2=10.627
0-4	1	7	5.56	1	2.43	p=0.014
5-9	3	25	19.56	4	9.44	
10-14	6	30	33.34	22	18.66	
15-19	11	18	23.31	22	16.69	
≥20	≥29	7	5.23	8	9.77	

Table 8. Analysis of Each Component of SOFA Score as a Predictor of Mortality

Component	Coef β	SE(β)	p	OR (95% CI)
Respiratory	0.821	0.281	0.003	2.273 (1.310-3.942)
Coagulation	0.682	0.260	0.009	1.978 (1.118-3.295)
Liver	0.034	0.362	0.925	1.035 (0.508-2.105)
Cardiovascular	0.969	0.223	<0.001	2.634 (1.703-4.075)
Neurology	0.594	0.212	0.005	1.811 (1.196-2.743)
Renal	0.597	0.255	<0.001	1.816 (1.102-2.993)

Table 9. Analysis of Each Component of MSOFA Score as a Predictor of Mortality

Component	Coef β	SE(β)	p	OR (95% CI)
Respiratory	1.582	0.476	0.001	4.863 (1.912-12.367)
Liver	0.172	0.442	0.697	1.188 (0.499-2.828)
Cardiovascular	1.024	0.224	<0.001	2.784 (1.796-4.315)
Neurology	0.693	0.209	0.001	1.999 (1.326-3.013)
Renal	0.619	0.246	0.012	1.857 (1.146-3.010)

Figure 1. AuROC of iSOFA Score

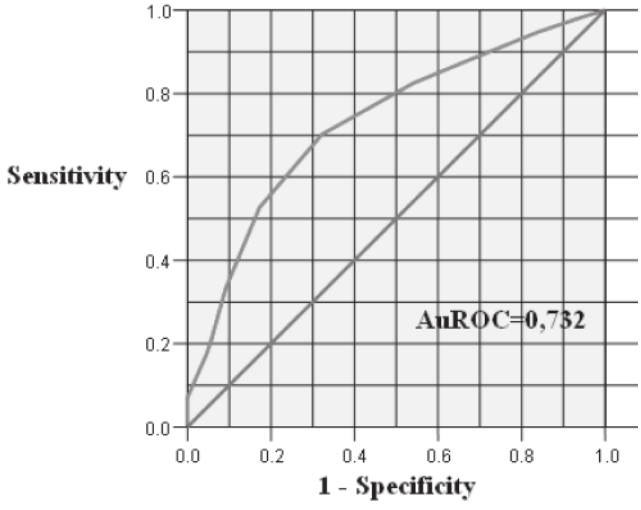


Figure 2. AuROC of meanSOFA Score

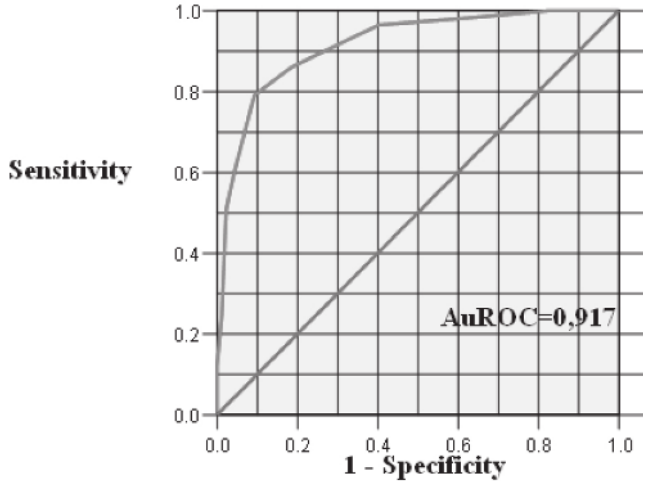


Figure 3. AuROC of maxSOFA Score

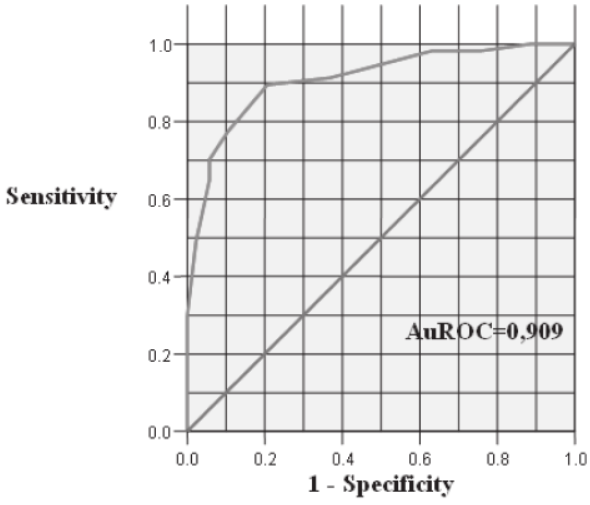


Figure 4. AuROC of iMSOFA Score

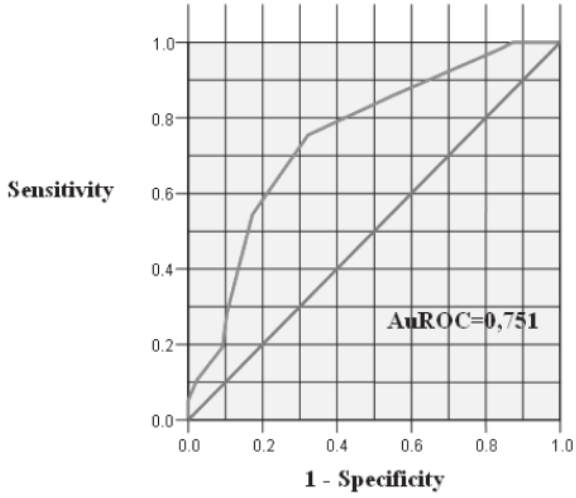


Figure 5. AuROC of meanMSOFA Score

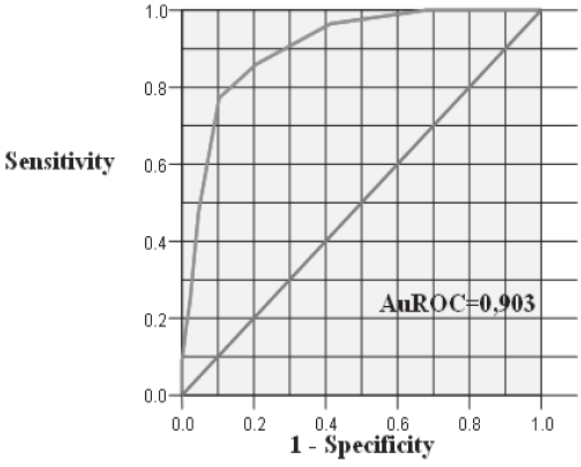


Figure 6. AuROC of maxMSOFA Score

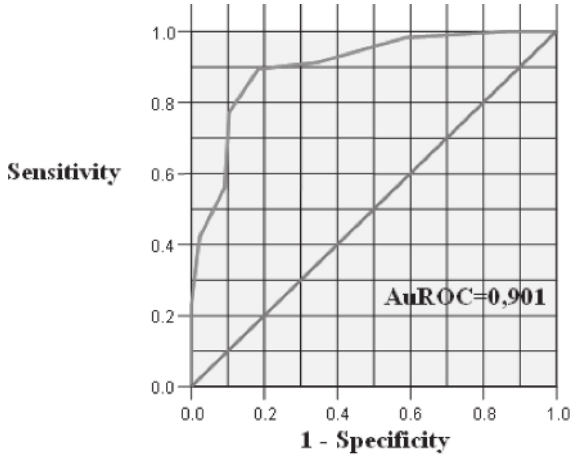
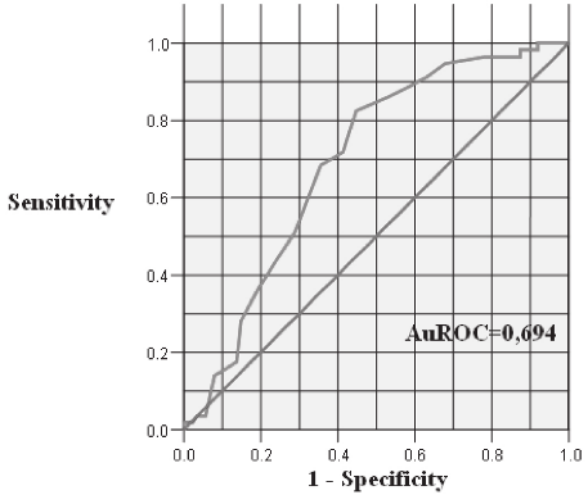


Figure 7. AuROC of APACHE II Score



References

1. Shortell SM, Zimmerman JE, Rousseau DM, Gilles RR, Wagner DP, Draper EA, et al. The performance of intensive care units: does good management make a difference? Med Care 1994;32:508-25.
2. Siegel JH, Rixen D. Clinical and physiologic scoring systems for sepsis and organ dysfunction.In: Deitch EA, Vincent JL, Windsor A, editors. Sepsis and multiple organ dysfunction, a multidisciplinary approach. London: W.B. Saunders; 2002. p. 161-78.
3. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. Br J Anaesth 2008;8:181-5.
4. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 2001;286:1754-8.
5. Cabré L, Mancebo J, Solsona JF, Saura P, Gich I, Blanch L, et al. Multicenter study of the multiple organ dysfunction syndrome in intensive care units: the usefulness of Sequential Organ Failure Assessment scores in decision making. Intensive Care Med 2005;31:927-33.
6. Lemeshow S, Le Gall JR. Modelling the severity of illness of ICU patients. A systems update. JAMA 1994;272:1049-55.
7. Le Gall JR. The use of severity scores in the intensive care unit. Intensive Care Med. 2005;31:1618-23.
8. 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.
9. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985;13:818-29.
10. Arregui LM, Moyes DG, Lipman J, Fatti LP. Comparison of disease severity scoring systems in septic shock. Crit Care Med 1991;19:1165-71.
11. Roumen RM, Schers TJ, de Boer HH, Goris RJ. Scoring systems for predicting outcome in acute hemorrhagic necrotizing pancreatitis. Eur J Surg 1992;158:167-71.
12. Wilson C, Health DI, Imrie CW. Prediction of outcome in acute pancreatitis: a comparative study of APACHE II, clinical assessment and multiple factor scoring systems. Br J Surg 1990;77:1260-4.
13. Fan ST, Lai EC, Mok FP, Lo CM, Zheng SS, Wong J. Prediction of the severity of acute pancreatitis. Am J Surg 1993;166:262-8.
14. Schuster HP, Schuster FP, Ritschel P, Wilts S, Bodmann KF. The ability of the Simplified Acute Physiology Score (SAPS II) to predict outcome in coronary care patients. Intensive Care Med 1997;23:1056-61.
15. Halstenberg WK, Goormastic M, Paganini EP. Validity of four models for predicting outcome in critically ill acute renal failure patients. Clin Nephrol 1997;47:81-6.
16. Zauner CA, Apsner RC, Kranz A, Kramer L, Madl C, Schneider B, et al. Outcome prediction for patients with cirrhosis of the liver in a medical ICU: a comparison of the APACHE scores and liver-specific scoring systems. Intensive Care Med 1996;22:559-63.
17. Bein T, Fröhlich D, Pömsl J, Forst H, Pratschke E. The predictive value of four scoring systems in liver transplant recipients. Intensive Care Med 1995;21:32-7.
18. Goel A, Pinckney RG, Littenberg B. APACHE II predicts long-term survival in COPD patients admitted to a general medical ward. J Gen Intern Med 2003;18:824-30.
19. de Cássia Braga Ribeiro K, Kowalski LP. APACHE II, POSSUM, and ASA scores and the risk of perioperative complications in patients with oral or oropharyngeal cancer. Arch Otolaryngol Head Neck Surg 2003;129:739-45.
20. Rutledge R, Fakhry S, Rutherford E, Muakkassa F, Meyer A. Comparison of APACHE II, Trauma Score, and Injury Severity Score as predictors of outcome in critically injured trauma patients. Am J Surg 1993;166:244-7.
21. Bohnen JM, Mustard RA, Schouten BD: Steroids, APACHE II score, and the outcome of abdominal infection. Arch Surg 1994;129:33-7.
22. Berger MM, Marazzi A, Freeman J, Chioléro R. Evaluation of the consistency of Acute Physiologic and Chronic Health Evaluation (APACHE II) scoring in surgical intensive care unit. Crit Care Med 1992;20:1681-7.
23. Bosscha K, Reijnders K, Hulstaert PF, Algra A, van der Werken C. Prognostic scoring systems to predict outcome in peritonitis and intra-abdominal sepsis. Br J Surg 1997;84:1532-4.
24. Cerra FB, Negro F, Abrams J. APACHE II score does not predict multiple organ failure or mortality in post operative surgical patients. Arch Surg 1990;125:519-22.
25. McLauchlan GJ, Anderson ID, Grant IS, Fearon KC. Outcome of patients with abdominal sepsis treated in an intensive care unit. Br J Surg 1995;85:524-9.
26. Ohmann C, Yang Q, Hau T, Wacha H. Prognostic modelling in peritonitis. Peritonitis Study Group of the Surgical Infection Society Europe. Eur J Surg 1997;163:53-60.
27. McAnena OJ, Moore FA, Moore EE, Mattox KL, Marx JA, Pepe P: Invalidation of the APACHE II scoring system for patients with acute trauma. J Trauma 1992;33:504-7.
28. Rixen D, Siegel JH, Friedman HP. "Sepsis/SIRS," physiologic classification, severity stratification, relation to cytokine elaboration and outcome prediction in posttrauma critical illness. J Trauma 1996;41:581-98.
29. Vassar MJ, Wilkerson CL, Duran PJ, Perry CA, Holcroft JW. Comparison of APACHE II, TRISS, and a proposed 24-hour ICU point system for prediction of outcome in ICU trauma patients. J Trauma 1992;32:490-500.
30. Civetta JM, Hudson-Civetta JA, Kirton O, Aragon C, Salas C. Further appraisal of APACHE II limitations and potential. Surg Gynecol Obstet 1992;175:195-203.
31. Tanaka Y, Shimizu M, Hirabayashi H. Acute physiology, age, chronic health evaluation (APACHE) III score is an alternative efficient predictor of mortality in burn patients. Burns 2007;33:316-20.
32. 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.
33. Vincent JL, de Mendonça A, Cantraine F, Moreno R, Takala J, Suter PM, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working Group on "sepsis-related problems" of the European Society of Intensive Care Medicine. Crit Care Med 1998;26:1793-800.
34. Grissom CK, Orme JF, Jensen RL, Jephson AR. A modified sequential organ failure assessment (SOFA) score to predict mortality in critically-ill patients (abstract). Crit Care Med 2007;35:A9.