EVALUATION OF APACHE- IV & SAPS- II SCORING SYSTEMS AND CALCULATION OF STANDARDISED MORTALITY RATE IN SEVERE SEPSIS AND SEPTIC SHOCK PATIENTS- A PROSPECTIVE OBSERVATIONAL STUDY

Siddhartha Chakraborty1, Sarbari Swaika2, Rajat Choudhuri3, Suchismita Mallick4

1Senior Resident, Department of Anaesthesiology and Critical Care, Institute of Post Graduate Medical Education and Research, Kolkata, West Bengal, India.
2Associate Professor, Department of Anaesthesiology and Critical Care, Institute of Post Graduate Medical Education and Research, Kolkata, West Bengal, India.
3Associate Professor, Department of Anaesthesiology and Critical Care, Institute of Post Graduate Medical Education and Research, Kolkata, West Bengal, India.
4Assistant Professor, Department of Anaesthesiology and Critical Care, Institute of Post Graduate Medical Education and Research, Kolkata, West Bengal, India.

ABSTRACT

BACKGROUND
Sepsis and septic shock are major causes of mortality in the intensive care units worldwide. The scoring systems are very useful to predict risk of mortality and evaluating outcome in critically ill patients. In this study, we aimed to research the effectiveness of SAPS II and APACHE IV scoring systems in the evaluation of prognosis in severe sepsis and septic shock patients hospitalized in ICU.

MATERIALS AND METHODS
A prospective observational study was conducted on 50 consecutive patients of severe sepsis and septic shock admitted to ICU between April 2016 to April 2017. Predicted mortality was calculated using online calculator. Standardised mortality rate (SMR) was calculated with 95% confidence intervals. Calibration was assessed using Hosmer-Lemeshow test, statistic and Cohen’s kappa statistic. Discrimination was assessed using receiver operating characteristic curves.

RESULTS
The actual mortality rate in this study was 52%. Predicted mortality rate of APACHE IV and SAPS II were 39.21% (SMR 1.32) and 45.85% (SMR 1.13) respectively. The Cohen’s kappa for APACHE IV and SAPS II were 0.369 and 0.426 respectively. Hosmer-Lemeshow goodness of fit statistic indicates good logistic regression model fit for both APACHE IV and SAPS II scoring system (p value > 0.05). AUROC of APACHE IV and SAPS II were 0.748 and 0.760 respectively.

CONCLUSION
SAPS II had a closer prediction and better discriminative ability than APACHE IV.

KEY WORDS
APACHE, SAPS II, Severe Sepsis, Septic Shock, ICU


BACKGROUND
Sepsis is one of the leading causes of death in ICU’s worldwide (Mortality rate 30%; and 50% when shock-associated).[1] Because of its aggressive, multi factorial nature, sepsis is a rapid killer affecting up to 75% of ICU patients, accounting for as much as 50% of ICU bed days and carrying mortality rate of 20 - 80%.[2, 3] Intensivists are frequently faced with questions regarding prognosis of critically ill sepsis patients in ICU. Only subjective evaluation of patients cannot give clear idea regarding severity of illness.

'Financial or Other Competing Interest': None.
Corresponding Author:
Dr. Sarbari Swaika,
Associate Professor,
Department of Anaesthesiology and Critical Care,
Institute of Post Graduate Medical Education and Research,
Kolkata, West Bengal, India.
E-mail: dr.s.swaika@gmail.com
DOI: 10.14260/jemds/2019/188

Over the past 20 years, numerous efforts have been made to design a model that can objectively quantify prognosis of such patients which is immensely helpful in clinical decision making. These scores also have been used as a surrogate measure of ICU performance and helpful in resource management.[4,5,6]

Most of these scoring systems were developed for general ICU patients, using large population databases in European and American ICUs. The predictive accuracy of these scores in Indian ICUs may not fit well because of differences in case mix.[7] So when applied to a particular group of patients, such as those with sepsis, their accuracy further declines. Till now data from the Indian subcontinent is not adequate for validation of these scoring systems in sepsis and septic shock patients.

Our aim was to assess the performance and utility of APACHE IV & SAPS II scoring system in predicting ICU mortality in severe sepsis and septic shock patients in a single tertiary multidisciplinary ICU. The other objective is to assess accuracy of prediction of ICU length of stay by APACHE IV in same study group.
MATERIALS AND METHODS
After approval by the Institutional Ethics Committee this prospective observational study was carried out in a mixed ICU of a tertiary care teaching hospital in one-year duration (April 2016 to April 2017) which was pre decided. This was a nine bedded ICU and all patients were admitted under one consultant (Closed ICU). The consent for participation was obtained at time of admission from the patients themselves or from the relatives who were most acquainted with the patient; a waiver of informed consent was granted by the Institutional ethics committee due to the minimal risk of the observational study. Patients <18 yrs and >70 yrs. of age and patients whose duration of stay in ICU was less than 4 hrs, were excluded from the study. The study included 50 consecutive patients who fulfilled the criteria for severe sepsis and septic shock within first 24 hrs. of ICU admission irrespective of reason of ICU admission (Medical or Surgical reason) within one-year study duration. After that they were followed up till death or discharge or transfer out from ICU. None of the study subject left ICU against medical advice.

At the time of the study, definitions of severe sepsis and septic shock were based on Surviving Sepsis Campaign 2012.[8] Severe sepsis was defined as harmful host response to infection (Systemic Inflammatory Response Syndrome) associated with some degree of organ hypofunction and Septic shock was defined as sepsis-induced hypotension and perfusion abnormalities despite fluid resuscitation, necessitating vasopressor support.[1]

Basic demographic characteristics, clinical findings and laboratory investigation reports were noted within first 24hrs of admission. Collected data was converted to Severity score and Predicted Mortality Rate for both scoring systems using online calculator.[9,10]

Primary objective of the study was to compare the effectiveness of the two scoring system in predicting mortality in sepsis and septic shock cases in terms of standardized mortality rate. Being an observational study, formal sample size calculation was not done. With the time and logistics at a disposal, we proposed at least 50 subjects to be recruited consecutively, subject to fulfillment of selection criteria and informed consent.

The performance of prognostic models encompasses two objective measures - calibration and discrimination.[11]

Calibration refers to how closely the predicted mortality correlates with the actual mortality over the entire range of probabilities. Comparison of both scoring systems was done in terms of standardised mortality rate (SMR). The SMR value expresses two things: first, the performance of ICU, and, second, how well a score is calibrated. An SMR of 1.00 means that actual and estimated death rates are equal and imply that the ICU has an average performance. A ratio greater than 1.00 would suggest a lower than average performance, whereas ICUs with a low SMR might be categorised as ‘high-performance’ units. Calibration of the prognostic models was assessed using Hosmer-Lemeshow goodness of fit Statistic. A higher p-value (>0.05) would indicate a good fit for the model.[12,13]

Discrimination refers to how well the model discriminates between individuals who will live and those who will die. Methods used to evaluate the ability of each model to discriminate include calculation of the area under the receiver operating characteristic (ROC) curve with its 95% CI. The area under the curve (AUC) represents the number of patients who died. Typically, model developers require an AUC of the ROC curve to be 0.70.[12] In mortality prediction models, a huge grey area exists between those who die and those who survive. Therefore, a number of pair of sensitivity–specificity values produces the ROC curve across the range of mortality prediction scores.[13,14]

Statistical Analysis
Statistical analysis of the data was done by Statistica version 6 [Tulsa, Oklahoma: StatSoft Inc, 2001] and MedCalc version 11.6 [Mariakerke, Belgium: MedCalc Software 2011]. The distribution of data was first evaluated using the Kolmogorov-Smirnov test. The Student’s t-test was used to compare continuous variables as indicated. Categorical variables were analysed using Chi square test and Fisher’s exact test. A value of p < 0.05 was considered as significant. Calibration of the prognostic models was assessed using Hosmer-Lemeshow goodness of fit Statistic. A higher p-value (>0.05) would indicate a good fit for the model. Calibration of both models was also assessed by Cohen’s kappa statistic. Predictive ability of ICU length of stay of APACHE IV was assessed by Intraclass Correlation Coefficients (ICC). ICC values less than 0.5 are indicative of poor reliability.

RESULTS
Fifty patients fulfilling the inclusion criteria were recruited consecutively as study subjects. Mean age of the subjects were 52.10 + 16.48 years with equal male: female ratio. Reasons of admission were mostly medical (70%), few were emergency surgery (26%) and scheduled surgery (4%). Forty six percent patients were referred from other hospitals and few patients were shifted to ICU from OT/Recovery (34%) and floor (20%) of our hospital. In this study population, it was observed that most common organ system involved was Respiratory system (32%) followed by Gastrointestinal system (24%), Neurological system (22%) and Genitourinary system (20%). Few patients (12%) also had some chronic disease like chronic kidney disease, chronic liver failure, diabetic mellitus and metastatic carcinoma. Presence of co morbidities did not affect survival of the patient significantly (p value 0.101).

The temperature and respiratory rate in the first 24 hours of ICU stay were found to be significantly higher in non survivors as compared to the survivors. Non survivors had significantly lower P02/FIO2 ratio when compared to survivors. Among non survivors 80.77% and among survivors 58.33% patients required mechanical ventilation in first 24 hrs of ICU admission (p value 0.124). Despite adequate fluid resuscitation among non survivors 26.9% and among survivors 20.83% required vasopressor support to maintain MAP more than 90 mmHg (p value 0.614) (Table 1). In this study population three patients needed haemodialysis support but two of them died.

Scoring systems consist of two parts: a severity score, which is a number (The higher the number the more severe is the condition) and a calculated probability of mortality. The mean score of APACHE-IV and SAPS-II of the subjects in this study was 79± 25.26 and 49.42 ± 16.18 respectively. For both scoring systems mean score for non survivors was significantly higher than for survivors (p value 0.001) (Table 1).
The Actual Mortality Rate (AMR) in this study was 52% (95% CI: 38.15 to 65.85%) since 26 out of 50 study patients had expired. Mean of Predicted Mortality Rate (PMR) for APACHE IV was 39.21% (SMR: 1.32) and for SAPS II was 45.85% (SMR: 1.13) (Table 2).

The discriminatory capability, as measured by the AUROC, was generally good for both models. Area under ROC for APACHE IV and SAPS II were 0.748 and 0.760 respectively. So, SAPS II showed better discriminative ability than APACHE IV (Figure 1). But there was no statistically significant difference between two scoring system (p value: 0.028). In Cohen's kappa statistic, agreement between actual and predicted mortality for APACHE IV is 0.369 (mortality implies score > 88) indicating fair agreement and for SAPS II scoring is 0.426 (mortality implies score > 31) indicating moderate agreement. So, both the systems predicted well, and SAPS II shows better calibration than APACHE IV.

Mean actual ICU length of stay (LOS) (9.60 ± 5.14) was significantly greater than APACHE IV predicted LOS (7.36 ± 2.40) (P value 0.003). With increase of APACHE IV score ICU-LOS initially increased and then gradually decreased. The Pearson's correlation coefficient (r) values for linear association between actual ICU-LOS and APACHE IV score was 0.06.

Table 5 shows on a case-by-case comparison basis, overall, in 68% cases (range: -0.40 to -22.40), there was a negative difference between Predicted LOS and Actual LOS meaning under prediction of ICU-LOS and in 32% cases (range: 0.4 to 7.3), the difference was positive indicating over prediction by APACHE-IV than actual ICU-LOS. Overall, APACHE-IV predicted ICU-LOS for severe sepsis patients very poorly and inconsistently (Intraclass correlation coefficient value is 0.181 for absolute agreement) (Fig. 2).

Contrasting result was obtained in surgical category patients where discriminative ability of APACHE IV (AUROC 0.929) was better than SAPS II (AUROC 0.848) (p value 0.2545) (Table 3).

To assess calibration for the two models, Hosmer-Lemeshow goodness of fit Statistic indicates good logistic regression model fit for both APACHE IV and SAPS II scoring system (p value > 0.05). In Cohen's kappa statistic, agreement between actual and predicted mortality for APACHE IV is 0.369 (mortality implies score > 88) indicating fair agreement and for SAPS II scoring is 0.426 (mortality implies score > 31) indicating moderate agreement. So, both the systems predicted well, and SAPS II shows better calibration than APACHE IV.

### Table 1. Comparison of Different Parameters Among Survivors and Non-Survivors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total (Mean ± SD)</th>
<th>Non-Survivors (Mean ± SD)</th>
<th>Survivors (Mean ± SD)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical Profile and Laboratory Profile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>37.73 ± 1.01</td>
<td>38.32 ± 0.83</td>
<td>37.08 ± 0.77</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean Arterial Pressure (mmHg)</td>
<td>67.28 ± 6.16</td>
<td>66.65 ± 5.98</td>
<td>67.96 ± 6.41</td>
<td>0.460</td>
</tr>
<tr>
<td>Mean Heart Rate (/min)</td>
<td>131 ± 10.94</td>
<td>131.8 ± 12.21</td>
<td>130.1 ± 9.56</td>
<td>0.584</td>
</tr>
<tr>
<td>Respiratory Rate (/min)</td>
<td>27.48 ± 2.67</td>
<td>29.7 ± 2.50</td>
<td>25.79 ± 1.38</td>
<td>0.000</td>
</tr>
<tr>
<td>Glasgow Coma Scale</td>
<td>7.5 ± 3.50</td>
<td>6.577 ± 3.55</td>
<td>8.917 ± 3.51</td>
<td>0.051</td>
</tr>
<tr>
<td>PO2/Fio2 †</td>
<td>243.5 ± 161.8</td>
<td>194.9 ± 24.88</td>
<td>296.2 ± 36.93</td>
<td>0.025</td>
</tr>
<tr>
<td>Arterial pH</td>
<td>7.36 ± 0.13</td>
<td>7.35 ± 0.02</td>
<td>7.39 ± 0.02</td>
<td>0.243</td>
</tr>
<tr>
<td>WBC Count</td>
<td>17476 ± 5019</td>
<td>18060 ± 948.2</td>
<td>15660 ± 499.30</td>
<td>0.396</td>
</tr>
<tr>
<td>Haematocrit</td>
<td>30.83 ± 7.61</td>
<td>31.92 ± 1.65</td>
<td>29.65 ± 1.35</td>
<td>0.295</td>
</tr>
<tr>
<td>Serum Creatinine (mg/dL)</td>
<td>1.64 ± 12.22</td>
<td>1.84 ± 0.29</td>
<td>1.43 ± 0.17</td>
<td>0.239</td>
</tr>
<tr>
<td>Serum Sodium (mEq/L)</td>
<td>136.5 ± 9.42</td>
<td>136.2 ± 1.82</td>
<td>136.7 ± 1.99</td>
<td>0.861</td>
</tr>
<tr>
<td>Serum Potassium (mEq/L)</td>
<td>3.94 ± 0.90</td>
<td>3.96 ± 0.21</td>
<td>3.91 ± 0.14</td>
<td>0.833</td>
</tr>
<tr>
<td><strong>Requirement of Mechanical Ventilation within First 24 Hours of ICU Admission</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required</td>
<td>35</td>
<td>21</td>
<td>14</td>
<td>0.124</td>
</tr>
<tr>
<td>Not Required</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>(Fisher’s exact test 2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Requirement of Vasopressors within First 24 Hours of ICU Admission</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required</td>
<td>28</td>
<td>7</td>
<td>5</td>
<td>0.614</td>
</tr>
<tr>
<td>Not Required</td>
<td>22</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>(Fisher’s exact test 2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Presence of Comorbidities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0.101</td>
</tr>
<tr>
<td>Absent</td>
<td>44</td>
<td>21</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>(Fisher’s exact test 2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Comparison of Reason of Admission among Survivors and Non-Survivors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical</td>
<td>35</td>
<td>18</td>
<td>17</td>
<td>0.987</td>
</tr>
<tr>
<td>Emergency Surgery</td>
<td>13</td>
<td>7</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Scheduled Surgery</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Severity Score of Both Scoring Systems</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APACHE IV</td>
<td>79 ± 25.26</td>
<td>89.96 ± 22.91</td>
<td>67.13 ± 22.50</td>
<td>0.001</td>
</tr>
<tr>
<td>SAPS II</td>
<td>49.42 ± 16.18</td>
<td>56.5 ± 14.84</td>
<td>41.75 ± 14.15</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 1. Comparison of Reason of Admission among Survivors and Non-Survivors

* PO2: Partial pressure of oxygen, FiO2: Fraction of oxygen in inspired air † WBC: White blood cell

---

**DISCUSSION**

The ideal scoring system should be simple, reliable, based on easily/routinely recordable variables, well calibrated with high level of discrimination (High sensitivity/specificity), applicable to all types of patient populations (All age groups, different diagnosis). No scoring system currently incorporates all these features.[11,15]

In present study, we used ICU mortality as the primary outcome, while the development of original SAPS II and APACHE IV scoring systems were established to record the hospital death as primary outcome. ICU death is better than the hospital death to measure outcome in our ICU for many reasons. The first reason is that no intermediate care or step-down unit is available in our hospital. Secondly, number of critically ill patients who need ICU care, are more than the beds available in ICU. As a result, some patients who are improving but still at risk and need intermediate support have to move directly from the ICU to the medical ward. These reasons probably lead to a higher mortality rate in hospital that is not related to the performance of ICU.[16]

Comparing our study's actual death rate of 52% in sepsis and septic shock patient we found that it is nearly similar to those of the other studies in India (60.71%)[17] and abroad (Saudi Arabia 46%[5], Italy 46.7%[18], France 42%[19], UK 50%[20]). In the present study, APACHE IV and SAPS II showed good calibration in Hosmer-Lemeshow goodness of fit Statistic (p value >0.05).
In terms of Standardised Mortality Rate (SMR), both APACHE IV and SAPS II models under predicted overall mortality and SAPS II was more accurate than APACHE IV (SAPS II 1.13 Vs APACHE IV 1.32). SAPS II had also a better discriminative ability than APACHE IV (APACHE IV AUROC 0.748 Vs SAPS II AUROC 0.760). These differences between observed and expected mortality might have been caused by poor management before ICU transfer which can partly correct physiological derangements without arresting underlying pathology. This phenomenon is known as 'lead-time bias' and can partly explain the reason behind the high mortality rates in patients with relatively lower calculated predicted mortality.

Similar results had been described by Abed N et al, where they reported that, APACHE IV (SMR 1.69) and SAPS II (SMR 1.47) both under predicted mortality but SAPS II was better in prognostication. Discrimination ability of SAPS II (AUROC 0.836) was also better than APACHE IV (AUROC 0.833). Our results on the performance of the SAPS II scoring system are in agreement with other reports published by Khan M et al,[3] Arbi Y et al,[5] Salir Y et al.[13] However Ayazoglu T found promising result in favour of APACHE IV.[21]

In present study, higher mortality rate was observed among surgical category patients than medical category patients. Most probable reason behind this was the high severity of illness score of surgical patients as 13 out of 15 surgical category patients admitted to our ICU after emergency surgery. SAPS II predicted better for post-surgical patients (SMR 1.02) than medical category patients (SMR 1.19). As a whole, APACHE IV under predicted for both medical and surgical category patients. Similar performance with APACHE IV system was demonstrated in a study among cancer patients in Korea by Xing X.[22] Customization or adding new variables may improve the ability of calibration.

The sensitivity and specificity of both scoring systems in our study were not very satisfactory (SAPS II: 100% sensitivity and 41.67% specificity and APACHE IV: 50% sensitivity and 87.5% specificity) when compared with others. Sharma S et al.[23] found SAPS II predicting with 88.23% sensitivity and 100% specificity in sepsis patients. In a Turkish intensive care unit, Ayazoglu T,[21] observed excellent result by APACHE IV on stroke patients (at score >84.5, 94.7% sensitivity and 94.4% specificity). Small sample size of our study may be the most probable reason for this discrepancy in sensitivity and specificity.

APACHE-IV is used internationally widely as a prediction tool for ICU-LOS. But in our ICU there was significant difference between APACHE IV predicted LOS and actual LOS (p value 0.003). In another study conducted in Kolkata, Chattopadhyay A et al[24] also found very poor and inconsistent ICU LOS prediction in sepsis and septic shock patients. APACHE IV was developed in a cohort of mixed ICU patients but not specifically for any subgroup of patients like sepsis/severe sepsis patients which may be the possible reason for poor prediction.

Sepsis, a syndrome of physiologic, pathologic, and biochemical abnormalities induced by infection. For better understanding of sepsis outcome, it is important to focus on derangement of those physiological measurements. Dhib et al[17] found that non-survivors had significantly high leucocyte count and required high FiO2 when compared with survivors; while survivors had significantly higher serum bicarbonate, albumin and pH as compared tonon-survivors. In present study comparing clinical and laboratory profile, we found significant difference among survivors and non-survivors in terms of temperature, respiratory rate and P02/FiO2 ratio. So different studies revealed different factors significantly affecting sepsis outcome though there are some common factors also. So, it may be necessary to further customize the models, or adding extra parameters to these scoring systems (Like CRP, serum lactate etc). But customization for sepsis and septic shock patients is more complicated due to subsequent involvement of multiple organ in late stage.

Careful evaluation of different studies revealed the fact that accuracy of the mortality prediction models are limited because they are restricted by the items included, and subjected to interpretation. Accuracy of the scoring systems also decreases as treatments and other factors influencing the mortality rate change. Being a single center study, however, some amount of bias due to differences in case mix, small sample size, lead time bias, and quality of care might have possibly occurred. These limiting factors were relevant in performed stratified analysis of calibration of both models.

CONCLUSION
SAPS II predicted much closer to actual mortality than APACHE IV. SAPS II had a better discriminative ability than APACHE IV. Prediction of ICU length of stay by APACHE IV was very poor and inconsistent.

REFERENCES


Online calculator for SAPS II. http://clincalc.com/IcuMortality/SAPSII.aspx [last accessed on 08/08/2018].

Online calculator for APACHE IV. http://intensivecarenetwork.com/Calculators/Files/Apache4.html [last accessed on 08/08/2018].


