Feasibility of transitioning from APACHE II to SAPS III as prognostic model in a Brazilian general intensive care unit. A retrospective study

Viabilidade da transição do APACHE II para o SAPS III como modelo prognóstico em uma unidade de terapia intensiva geral no Brasil. Um estudo retrospectivo

Ary Serpa Neto, Murillo Santucci Cesar de Assunção, Andréia Pardini, Eliézer Silva

Department of Critical Care Medicine, Hospital Israelita Albert Einstein (HIAE), São Paulo, Brazil

ABSTRACT

CONTEXT AND OBJECTIVE: Prognostic models reflect the population characteristics of the countries from which they originate. Predictive models should be customized to fit the general population where they will be used. The aim here was to perform external validation on two predictive models and compare their performance in a mixed population of critically ill patients in Brazil.

DESIGN AND SETTING: Retrospective study in a Brazilian general intensive care unit (ICU).

METHODS: This was a retrospective review of all patients admitted to a 41-bed mixed ICU from August 2011 to September 2012. Calibration (assessed using the Hosmer-Lemeshow goodness-of-fit test) and discrimination (assessed using area under the curve) of APACHE II and SAPS III were compared. The standardized mortality ratio (SMR) was calculated by dividing the number of observed deaths by the number of expected deaths.

RESULTS: A total of 3,333 ICU patients were enrolled. The Hosmer-Lemeshow goodness-of-fit test showed good calibration for all models in relation to hospital mortality. For in-hospital mortality there was a worse fit for APACHE II in clinical patients. Discrimination was better for SAPS III for in-ICU and in-hospital mortality (P = 0.042). The SMRs for the whole population were 0.27 (CI: 0.23 – 0.33) for APACHE II and 0.28 (CI: 0.22 – 0.36) for SAPS III.

CONCLUSIONS: In this group of critically ill patients, SAPS III was a better prognostic score, with higher discrimination and calibration power.

INTRODUCTION

Prognostic models reflect the population characteristics of the countries from which they originate. The development of the Acute Physiology and Chronic Health Evaluation II (APACHE II) system was based on a cohort of patients in the United States, and it has been used in many intensive care units around the world. In contrast, the Simplified Acute Physiology Score III (SAPS III) was validated in a multicenter and multinational cohort study.

Predictive models should be customized to fit the case-mix population where they will be used because the outcomes in the original databases and period from which the models were derived may be different from the databases of intensive care units (ICUs) using the models. It is not clear whether calibration of the established models for local circumstances would enhance their accuracy in stratifying patients.

Although the methods have been reported to adapt well to different periods and case mixes, few studies have formally assessed the models’ predictive accuracy when applied to new populations from other institutions or countries. In South America, SAPS III was calibrated with a level of 1.3 (i.e., the relationship between observed and predicted mortality was 1.3). Recently studies have validated SAPS III in different Brazilian cohorts of patients, obtaining good results.

Comparison between observed and predicted mortality rates could serve as an indicator of ICU performance, and lead to overall improvement in healthcare services. However, ICU profiles vary worldwide, depending on the proportions of medical and surgical patients, admission and discharge policies, availability of intermediate care units and staffing with intensive care specialists.

Any transition from a well-established approach to a new one requires caution and validation. Changing APACHE II for SAPS III has some advantages and the most important is the fact that SAPS III is the only prognostic score that included a cohort of patients from South America in its development.

OBJECTIVE

In the present study, we aimed to perform external validation on two predictive models and directly compare their performance in an independent population of mixed critically ill patients.

METHODS

Data collection

This study was approved by the Ethics Committee of Hospital Israelita Albert Einstein and, because of the retrospective nature of the study, the informed consent requirement was waived. The data were collected from all patients admitted to a mixed 41-bed ICU in the tertiary-level private hospital in Brazil from August 2011 to September 2012.

Data were retrospectively collected using APACHE II only between August 2011 and December 2011, and using SAPS III only between May 2012 and September 2012. From January 2012 to April 2012, during a period of calibration, both scores were calculated for all patients admitted to the ICU and were collected for analysis. The data collection practices were standardized and performed by a trained nurse or physician. All data were checked for implausible and outlying values. The data included age, gender and type of admission (clinical, elective surgery or emergency surgery).

Study population

All ICU admissions were enrolled during the period analyzed. The exclusion criteria were: age < 18 years, missing data and not receiving ICU care. The admissions between January 2012 and April 2012 were used as a validation database to study the performance of APACHE II versus SAPS III for all admissions and in subgroups according to the type of admission.

Scores and predicted mortalities

The calculations of the individual scores for each model were based on the most disordered physiological values recorded during the first 24 hours of ICU admission for APACHE II and were based on the variables measured one hour before and after ICU admission for SAPS III. The mortality probabilities for APACHE II and SAPS III were calculated using the original equations.

Performance of the scores

The calibration of the scores was tested using the Hosmer-Lemeshow goodness-of-fit procedure, which was calculated by dividing the admissions into ten deciles according to the risk of death. The chi-square statistics were determined for each decile and summing the chi-square values for the ten deciles resulted in the test value. A high P value would indicate a good fit for the model. Hosmer-Lemeshow is a test for assessing agreement between the actual and predicted death rates. The discriminative ability of the models was assessed using receiver operating characteristic (ROC) curves and the respective areas under curves (AUC). The AUC is an expression of the model’s ability to discriminate correctly between survivors and non-survivors.

The standardized mortality ratio (SMR) was calculated using the models by dividing the number of observed deaths by the number of expected deaths. Confidence intervals for the SMR were calculated to test the model’s uniformity-of-fit, using the methods that have been put forward. The variables were compared between the three periods using analysis of variance (ANOVA). Calibration curves were constructed by plotting the predicted death rates stratified as 5% intervals of mortality risk (x-axis) versus observed death rates (y-axis). Finally, we constructed a model using Cox regression analysis with APACHE II and SAPS III as independent factors.

Statistical significance was defined as P < 0.05 and the results are presented as mean ± standard deviation unless indicated.
otherwise. All the statistical procedures were performed using the SPSS 20.0 statistical package (SPSS, Chicago, Illinois, USA).

RESULTS

Study population
A total of 3,333 ICU admissions were enrolled until the end of September 2012. The formation of the database is presented in Table 1. The characteristics of the population in the three periods are presented in Table 2. The ICU and hospital mortality and the APACHE II score decreased over time, and the SAPS III score increased during the periods.

Calibration and discrimination
The Hosmer-Lemeshow goodness-of-fit statistics supported model fit for all in-ICU mortality models with the exception of APACHE II for patients in the calibration database undergoing elective surgery. For in-hospital mortality, there was worse fit for APACHE II among clinical patients during the first period and for SAPS III among patients in the calibration database undergoing elective surgery (Table 3). The calibration curves for APACHE II and SAPS III showed overestimation of the risk of death in all ranges of predicted mortality (Figure 1).

Discrimination, as tested by the AUC, among general and clinical patients, was better for SAPS III in relation to in-ICU and in-hospital mortality (P = 0.042) (Table 4). Figure 2 shows the ROC for SAPS III and APACHE II, for in-ICU mortality in the calibration database in different situations.

Standardized mortality ratio
The SMRs for the whole population were 0.27 (CI: 0.23 – 0.33) for APACHE II and 0.28 (CI: 0.22 – 0.36) for SAPS III. In the calibration database, the SMRs for APACHE II and SAPS III were 0.33 (CI: 0.22 – 0.50), and 0.36 (CI: 0.25 – 0.55), respectively.

Table 1. Study database

<table>
<thead>
<tr>
<th>Data gathered (admissions)</th>
<th>3,333</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>65.47 ± 18.02</td>
</tr>
<tr>
<td>Type of admission, n (%)</td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>2,155 (64.65)</td>
</tr>
<tr>
<td>Elective surgery</td>
<td>994 (29.82)</td>
</tr>
<tr>
<td>Emergency surgery</td>
<td>184 (5.53)</td>
</tr>
</tbody>
</table>

Table 2. Characteristics of the study population

<table>
<thead>
<tr>
<th></th>
<th>APACHE II database</th>
<th>Calibration database</th>
<th>SAPS III database</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 1,495 %</td>
<td>n = 425 %</td>
<td>n = 1,413 %</td>
<td></td>
</tr>
<tr>
<td>Type of admission, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>1,061 (70.96)</td>
<td>246 (57.90)</td>
<td>848 (60.00)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Elective surgery</td>
<td>332 (22.20)</td>
<td>38 (8.90)</td>
<td>44 (3.10)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Emergency surgery</td>
<td>102 (6.84)</td>
<td>141 (33.20)</td>
<td>521 (36.90)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Period</td>
<td>2011 (August – December)</td>
<td>2012 (January – April)</td>
<td>2012 (May – September)</td>
<td>0.299</td>
</tr>
<tr>
<td>Age, years</td>
<td>65.76 ± 17.82</td>
<td>64.23 ± 7.99</td>
<td>65.53 ± 18.23</td>
<td>0.299</td>
</tr>
<tr>
<td>Mortality, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICU</td>
<td>127 (8.50)</td>
<td>29 (6.80)</td>
<td>73 (5.20)</td>
<td>0.002</td>
</tr>
<tr>
<td>Hospital</td>
<td>212 (14.20)</td>
<td>50 (11.80)</td>
<td>134 (9.50)</td>
<td>0.001</td>
</tr>
<tr>
<td>Length of stay in ICU, days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>–</td>
<td>3.61 ± 4.91</td>
<td>–</td>
<td>3.59 ± 5.42</td>
</tr>
<tr>
<td>Maximum</td>
<td>–</td>
<td>43.00</td>
<td>–</td>
<td>50.00</td>
</tr>
<tr>
<td>Mean among survivors</td>
<td>–</td>
<td>3.28 ± 4.18</td>
<td>–</td>
<td>3.27 ± 4.79</td>
</tr>
<tr>
<td>Mean among non-survivors</td>
<td>–</td>
<td>8.17 ± 9.79</td>
<td>–</td>
<td>9.38 ± 10.66</td>
</tr>
<tr>
<td>Length of stay in hospital, days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>–</td>
<td>22.04 ± 30.53</td>
<td>–</td>
<td>18.73 ± 41.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>–</td>
<td>217.00</td>
<td>–</td>
<td>792.00</td>
</tr>
<tr>
<td>Mean among survivors</td>
<td>–</td>
<td>19.99 ± 27.88</td>
<td>–</td>
<td>16.98 ± 37.54</td>
</tr>
<tr>
<td>Mean among non-survivors</td>
<td>–</td>
<td>37.28 ± 43.00</td>
<td>–</td>
<td>34.26 ± 62.02</td>
</tr>
<tr>
<td>Scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APACHE II</td>
<td>18.20 ± 7.09</td>
<td>14.97 ± 7.72</td>
<td>–</td>
<td>– 0.127</td>
</tr>
<tr>
<td>SAPS III</td>
<td>–</td>
<td>30.50 ± 17.00</td>
<td>–</td>
<td>45.62 ± 15.67</td>
</tr>
<tr>
<td>Probability of death, %</td>
<td>APACHE II</td>
<td>30.74 ± 19.22</td>
<td>20.56 ± 20.69</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>SAPS III</td>
<td>18.59 ± 22.76</td>
<td>18.40 ± 20.04</td>
<td>–</td>
</tr>
</tbody>
</table>

ICU = intensive care unit; APACHE II = acute physiology and chronic health disease classification system II; SAPS III = simplified acute physiology score III.
For all models, the SMRs showed some variation across the spectrum of patients. The SMRs ranged from 0.24 to 0.46 for APACHE II, and from 0.09 to 0.31 for SAPS III. In the calibration database, the SMRs ranged from 0.13 to 0.38 for APACHE II, and from 0.18 to 0.40 for SAPS III (Table 5).

Cox regression model
The hazard ratios for in-hospital and in-ICU mortality using APACHE II as an independent factor were 1.08 (95% CI: 1.04 – 1.12) and 1.09 (95% CI: 1.04 – 1.14), respectively. For SAPS III, the hazard ratios for in-hospital and in-ICU mortality were 1.03 (95% CI: 1.03 – 1.04) and 1.04 (95% CI: 1.03 – 1.06), respectively.
DISCUSSION

The external validation of these two widely used prognostic models showed good discrimination and good calibration when applied to the same independent population of Brazilian ICU patients. The transition from APACHE II to SAPS III in this Brazilian ICU was feasible and, in some scenarios, SAPS III had even better performance than APACHE II.

The SAPS III score was developed using data from 16,784 patients. However, it was not developed to be representative of all ICU patients, since it was developed in a cohort of general patients. Therefore, external validation is extremely important before applying this score to any type of patient. The cohort used for the conception of the SAPS III model involved patients from Brazilian ICUs, which might explain its superiority to APACHE II in our study. Also, the SAPS III model is based exclusively on data evaluated during the first hour of admission to the ICU. Prognostic systems that include measurements during and/or after the first 24 hours of admission, like APACHE II, often reflect standard care and not the real clinical status of the patient.

Severity scores have the aim of measuring the severity of the disease of ICU patients and are proposed as tools to aid in outcome assessment and resource allocation. Furthermore, they are used for comparisons of outcomes and quality of care between ICUs.

Our findings are supported by other studies published in the literature. Soares et al. demonstrated that SAPS III had excellent discrimination in Brazilian ICUs and that APACHE II was unsatisfactory due to its lower discriminatory power and lack of calibration for some populations. However, they also found that SAPS III overestimated hospital mortality, as also found in our study. In a cohort of Brazilian patients, Silva Júnior et al. concluded that SAPS III is a useful tool for determining which patients will need more care, and for the evolution of high-risk surgical patients.

Alves et al. showed, among elderly patients admitted to an ICU, that SAPS III had excellent discrimination, but that the calibration was inadequate. Soares et al. also showed that SAPS III had good discrimination and inadequate calibration among...

**Table 5. Standardized mortality ratio calculated using the prediction models according to type of admission (95% confidence interval)**

<table>
<thead>
<tr>
<th>Type of admission</th>
<th>APACHE II Admissions (%)</th>
<th>Deaths (%)</th>
<th>SMR</th>
<th>Calibr. database Admissions (%)</th>
<th>Deaths (%)</th>
<th>SMR</th>
<th>SAPS III Admissions (%)</th>
<th>Deaths (%)</th>
<th>SMR</th>
<th>SMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td>1,061 (70.9)</td>
<td>87 (8.19)</td>
<td>0.269 (0.22-0.34)</td>
<td>246 (57.8)</td>
<td>24 (9.80)</td>
<td>0.378 (0.25-0.59)</td>
<td>407 (0.26-0.63)</td>
<td>848 (60.0)</td>
<td>69 (8.13)</td>
<td>0.312 (0.24-0.4)</td>
</tr>
<tr>
<td>Elective surgery</td>
<td>332 (22.2)</td>
<td>25 (7.53)</td>
<td>0.241 (0.16-0.36)</td>
<td>141 (33.1)</td>
<td>2 (1.40)</td>
<td>0.135 (0.03-0.57)</td>
<td>0.183 (0.04-0.81)</td>
<td>521 (36.9)</td>
<td>3 (0.57)</td>
<td>0.093 (0.03-0.3)</td>
</tr>
<tr>
<td>Emergency surgery</td>
<td>102 (6.8)</td>
<td>15 (14.7)</td>
<td>0.460 (0.27-0.81)</td>
<td>38 (9.10)</td>
<td>3 (7.90)</td>
<td>0.327 (0.10-1.14)</td>
<td>0.324 (0.10-1.14)</td>
<td>44 (3.10)</td>
<td>1 (2.27)</td>
<td>0.151 (0.02-1.11)</td>
</tr>
<tr>
<td>Total</td>
<td>1,495 (100)</td>
<td>127 (8.50)</td>
<td>0.276 (0.23-0.33)</td>
<td>425 (100)</td>
<td>29 (6.80)</td>
<td>0.331 (0.22-0.50)</td>
<td>0.366 (0.25-0.55)</td>
<td>1,413 (100)</td>
<td>73 (5.20)</td>
<td>0.281 (0.22-0.36)</td>
</tr>
</tbody>
</table>

ICU = intensive care unit; APACHE II = acute physiology and chronic health disease classification system II; SAPS III = simplified acute physiology score III; SMR: standardized mortality ratio.
cancer patients. Costa e Silva et al.19 showed, in Brazilian critically ill patients with acute kidney injury, that SAPS III presented good discrimination and calibration performances, accurately predicting mortality in this group of patients. Finally, Nassar et al.13 demonstrated that SAPS III had good discrimination and inadequate calibration in a general cohort of Brazilian patients.

Our findings are supported by several studies that compared APACHE II with SAPS in different scenarios.20-22 These scores have also already been used with other aims like estimation of prolonged mechanical ventilation in surgical patients.23 In our hospital, we have a high number of liver transplantation procedures, and the performance of these scores in these population is a matter of debate.24-26 Also, one limitation of this study is that our hospital has a lower mortality rate than other institutions from Brazil and this makes it difficult to differentiate whether we have a score with bad performance or whether we have an ICU with excellent performance.

CONCLUSIONS
We showed, in a Brazilian cohort of critically ill patients, that SAPS III was a better prognostic score, with higher discrimination and calibration power. The transition from APACHE II to SAPS III was feasible in this scenario.

REFERENCES


Sources of funding: None
Conflict of interest: None

Date of first submission: December 16, 2013
Last received: January 3, 2014
Accepted: April 22, 2014

Address for correspondence:
Murillo Santucci César Assunção
Av. Albert Einstein, 700
São Paulo — Brasil
CEP 05651-901
Tel. (+ 55 11) 2151-1500
E-mail: murilloassuncao@gmail.com