Validation and customization of the SAPS 3 score in Brazilian ICUs

On Behalf of "UTIs Brasileiras", The Brazilian ICU Registry

July 2025

Steering Committee:

Marcio Soares (study coordinator)

Vice-President of Development and Research, Epimed Solutions, Rio de Janeiro

Senior Researcher, D'Or Institute for Researcher and Education, Rio de Janeiro

Ederlon Rezende

National Coordinator, UTIs Brasileiras

Past-President (2010-2011), Brazilian Association of Intensive Care (AMIB)

ICU Director, Hospital do Servidor Público Estadual, São Paulo

Suzana M. Lobo

Past-President (2020-2021), Brazilian Association of Intensive Care (AMIB)

ICU Director, Hospital de Base, São José do Rio Preto

Associate Professor, Faculdade de Medicina de São José do Rio Preto - FAMERP, São José do Rio de Preto

Marcelo O. Maia

Past-President (2022-2023), Brazilian Association of Intensive Care (AMIB)

ICU Director, Hospital Anchieta, Distrito Federal

Patrícia M. V. C. Melo

President (2024-2025), Brazilian Association of Intensive Care (AMIB)

Associate Professor, Universidade Federal do Piauí, Teresina

Jorge I. F. Salluh

Scientific Director, Epimed Solutions, Rio de Janeiro

Senior Researcher, D'Or Institute for Researcher and Education, Rio de Janeiro

Editor-in-Chief, Critical Care Science

Lucas R. Garcia de Mello

Customer Services Director, Epimed Solutions, Rio de Janeiro

Associate Professor, Departamento de Enfermagem Médico Cirúrgico (DEMC) da Faculdade de Enfermagem da Universidade Estadual do Rio de Janeiro – UERJ, Rio de Janeiro

Statistical Analysis

Lunna P. Borges

Data Science Manager, Epimed Solutions, Rio de Janeiro

Thaís A. Machado

Statistician, Epimed Solutions, Rio de Janeiro

External Methods Review

Leonardo S. L. Bastos

Associate Professor, Department of Industrial Engineering, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro

Otávio T. Ranzani

Senior Investigator, Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain

Conflicts of interest

M.S. and J.I.F.S. are founders and equity shareholders of Epimed Solutions®, which commercializes the Epimed Monitor System®, a cloud-based software for ICU management and benchmarking. L.R.G.M., L.P.B. and T.A.M. are employees of Epimed Solutions®. M.S. and J.I.F.S. are supported by individual research grants from the National Council for Scientific and Technological Development (CNPq), Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ). The other authors declare that they have no conflicts of interest.

Take-home Messages

What is already known about this topic:

The Simplified Acute Physiology (SAPS) 3 was published in 2005, and it is the most commonly used severity of illness score in Brazil. The performance of a score is expected to deteriorate over time, particularly in terms of calibration. Therefore, it should be reassessed periodically to evaluate whether it remains appropriate. In addition, the parameters and metrics used to estimate ICU resources were reported in 2007 and, to the best of our knowledge, have not been revalidated. The present study evaluated SAPS 3 performance in a contemporary cohort of 1,306,811 patients admitted to 1,239 ICUs participating in the Brazilian ICU Registry (UTIs Brasileiras) and investigated the need for its customization.

What this study adds:

We found that the estimated hospital mortality using the standard equation of SAPS 3 (SAPS 3-SE) and ICU resources using the original metrics to estimate the length of stay (LOS) per survivor were significantly overestimated, which resulted in low standardized mortality and resource use rates, respectively. Therefore, we performed a first-level customization (recalibration) of SAPS 3 (SAPS 3-Custom) and derived a new set of metrics to estimate the expected number of ICU days per survivor. The customization procedures corrected these overestimations and resulted in more accurate predictions of the two outcomes of interest.

How this study may affect practice or policy:

The study results support the use of SAPS 3-Custom over SAPS 3-SE to evaluate ICU performance and efficiency and for benchmarking in Brazilian ICUs. However, reevaluations of the performance of the SAPS 3-Custom and its customized LOS per survivor should be regularly performed to assess whether they remain accurate in the near future.

Contents

Tables

Figures

Abbreviations

AMIB: Associação de Medicina Intensiva Brasileira

AUROC: area under the receiver operating characteristic curve

CI: confidence interval

ICU: intensive care unit

IQR: 25%-75% interquartile range

LOS: length of stay

SAPS 3: Simplified Acute Physiology Score 3

SAPS 3-SE: SAPS 3 standard equation

SAPS 3-Custom: SAPS 3 customized equation

SMR: standardized mortality rate

SRU: standardized resource use rate

Introduction

The severity of illness scores is routinely used to evaluate the performance and efficiency of intensive care units (ICUs) for the benchmarking and assessment of temporal severity-adjusted trending of mortality.(1–3) These scores are also used in clinical trials and observational studies to characterize and stratify subgroups of patients in terms of illness severity. Therefore, these instruments provide valuable clinical and administrative information.

However, model performance inherently varies across different settings because of differences in case-mix, clinical management, admission and discharge policies, among other factors.(3, 4) Moreover, the performance of these scores is also expected to deteriorate over time, particularly in terms of calibration.(5, 6) Therefore, the severity of illness scores must be validated prior to their use in a specific setting or geographic region and reassessed periodically to evaluate whether their performance remains appropriate.

The Simplified Acute Physiology (SAPS) 3 score was published in 2005, and it was developed from a database with 16,784 patients who were admitted to 303 ICUs from 35 countries, including Brazil.(4) Since 2009, the Brazilian Association of Intensive Care (Associação de Medicina Intensiva Brasileira, AMIB) has defined the SAPS 3 score as the recommended severity of illness score for assessing ICU performance and benchmarking Brazilian ICUs.(7) This decision has been supported by different multicenter studies.(8–10) The last large and multicenter validation study was published in 2017 and used data from 48,818 patients who were admitted to 70 ICUs in 50 hospitals during 2013.(8) In this study, the SAPS 3 standard equation (SAPS 3-SE) had good discrimination and calibration, but the customized equation for Central and South American countries overestimated mortality.(8) Recently, the critical care setting was seriously challenged by the worldwide COVID-19 pandemic, with significant changes in patient clinical management and ICU organization and management.(11) Therefore, validation of the SAPS 3 in a post-pandemic critical care population is needed to assess whether this model still performs well. In addition, the SAPS 3 provides information for evaluating ICU efficiency using standardized resource use (SRU). Nevertheless, the parameters and metrics used to estimate the SRU were reported in 2007 by Rothen et al. using the SAPS 3 original dataset, and to the best of our knowledge, they have not been revalidated.(12) The present study aimed to evaluate SAPS 3 performance in a contemporary cohort of patients who were admitted to ICUs participating in the Brazilian ICU Registry (UTIs Brasileiras)(13) and investigate whether it needs to be customized.

Methods

Study design and setting

We performed a retrospective analysis of prospectively collected data from adult patients (≥16 years old) who were admitted to ICUs participating in the Brazilian ICU Registry between January 1st, 2023 and September 30th, 2024.

Selection of Participants, Data Collection and Definitions

We included adult ICUs with ≥5 beds that participated in the Brazilian ICU Registry. The registry is an initiative led by AMIB in partnership with Epimed Solutions® (Rio de Janeiro, Brazil) to characterize the epidemiological profile of Brazilian ICUs and share useful information to guide health policies and strategies to improve the care of critically ill patients in the country.(13) In addition, the project aims to encourage the use of quality indicators and performance in the management of ICUs and improve the quality of intensive care and increase patient safety in Brazil.

We included all patients aged ≥16 years who were admitted to the participating ICUs during the study period. We excluded patients who were potential organ donors and brain dead at ICU admission or readmitted to the ICU during the same hospitalization, had an ICU length of stay (LOS) <1 hour and without hospital discharge..

Among included patients, for the purpose of ICU eligibility, valid patients were those with a recorded main diagnosis. ICUs with <60% of valid patients among the total patients or >20% of patients transferred to another hospital/institution, home-care, or hospice at hospital discharge were also excluded.

The participating ICUs routinely collected patients by trained nurses and medical personnel using the Epimed Monitor System (Epimed Solutions®, Rio de Janeiro, Brazil), which is a cloud-based registry for ICU quality improvement and benchmarking purposes.(14) The data included demographics, admission source, hospital LOS before ICU admission, primary ICU admission diagnosis, Sequential Organ Failure Assessment (SOFA) score at admission,(15) comorbidities based on the Charlson Comorbidity Index,(16) frailty assessed by the Modified Frailty Index,(17) use of organ support during the ICU stay, ICU and hospital LOS, vital status at hospital discharge (dead or alive) and destination after hospital discharge. We calculated the SAPS 3-SE score as recommended using the logit [-32.6659 + ln(SAPS 3 score + 20.5958) × 7.3068] to estimate the probability of death: $e^{logit}/(1+e^{logit})$.(4)

Outcomes

The primary outcome was all-cause in-hospital mortality at the patient level. The ICU LOS was the secondary outcome.

Missing data

Following the recommendations for the SAPS 3 calculation, we input normal values for laboratory and physiological variables.(4) The **a-Table** in the Appendix presents the frequencies of missing data for each SAPS 3 variable or component.

Statistical analysis

We described the ICU and patient characteristics using standard descriptive statistics and report continuous variables as the means ± standard deviation or medians (25%-75% interquartile range, IQR), as appropriate. We reported categorical variables as absolute numbers (frequency percentages).

We evaluated model discrimination (i.e., the ability of each model to discriminate between patients who lived and patients who died) by estimating the area under the receiver operating characteristic curve (AUROC). We plotted calibration curves with 95% confidence intervals (CIs) to investigate the relationships between the observed and expected outcomes within each risk decile. We used Brier's score as an additional parameter for comparing the overall agreement between the predicted and observed outcomes.(18) Standardized mortality (SMR) and resource use (SRU) ratios with 95% CIs were estimated to evaluate clinical performance and resource use efficiency, respectively. The SMR is the ratio between the observed and predicted hospital mortality. The SRU estimates the average observed-to-expected ratio of resources (based on the ICU LOS) used per surviving patient in a specific ICU adjusted for the SAPS 3, as proposed by Rothen et al.(12) We fitted funnel plot graphs of SMRs and SRUs considering only those ICUs with greater than 150 admissions to evaluate potential biases in the estimation of these indicators.(19)

We temporally stratified the samples into training (01/01/2023 – 03/31/2024) and validation (04/01/2024 – 09/30/2024) datasets. After confirming the poor calibration of the original SAPS 3-SE, we performed a first-level customization (recalibration) by computing a new logistic coefficient while maintaining the same variables with the same weights as the original model. A logistic regression was fitted with the SAPS 3 score as the independent variable and in-hospital mortality as the dependent variable in the training dataset (SAPS 3 customized equation, SAPS 3-Custom). To update the average number of expected ICU days to produce a survivor in our dataset for the purpose of SRU estimation, we stratified patients into the nine SAPS 3 strata (< 24; 25-34; 35-44; 45-54; 55-64; 65-74; 75-84; 85-94; ≥ 95 points) originally proposed by Rothen et al.(12) First,

the average number of resources expected to produce a survivor in each stratum was estimated by dividing the sum of the ICU LOS of all patients in that stratum by the number of surviving patients in that stratum. To calculate the SRU for a given ICU, the sum of the ICU LOS of all patients was divided by the sum of the total number of expected days to produce survivors according to the SAPS 3 strata to which patients were assigned. “Zero” expected days was assigned to non-survivors. To avoid noise and the wide variability of the SRUs introduced by patients with disproportionately high ICU LOSs, we truncated it to 30 days. We performed a pre-specified sensitivity analysis for medical and surgical patients.

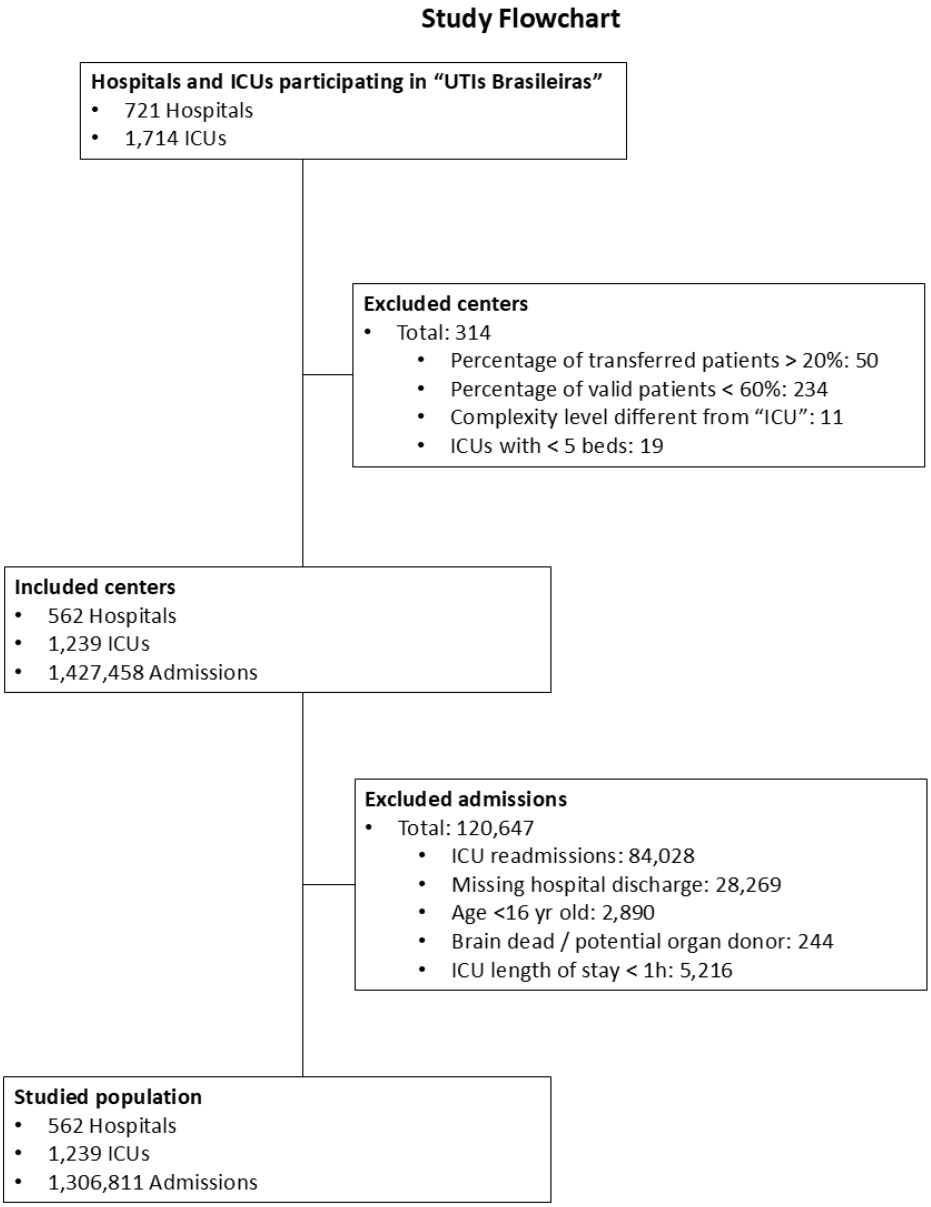
A two-tailed p value < 0.05 was considered statistically significant. We performed the statistical analyses using R version 3.5.2 (http://www.r-project.org).

Results

Characterization of the studied population and participating centers

A total of 1,306,811 patients admitted to 1,239 ICUs in 562 hospitals during the study period were considered eligible and constituted the study population (**Figure 1**).

Figure 1. Eligibility of ICUs and patients.



ICU: intensive care unit.

Table 1 presents the main hospital and ICU characteristics. Most ICUs were medical-surgical (n=966, 78.0%) located at private hospitals (n=811, 65.5%). The median number of patients per ICU was 880 (561 – 1,286).

Table 1. ICU characteristics (n=1,239)

Variable	n (%)
Hospital Type	
Public	428 (34.5)
Private	811 (65.5)
Region	
Southeast	654 (52.8)
Northeast	286 (23.1)
Central-West	166 (13.4)
South	73 (5.9)
North	60 (4.8)
ICU beds (n)	
10 - 20	1,042 (84.1)
20 - 30	150 (12.1)
≥ 30	47 (3.8)
Unit Type	
General/Mixed	966 (78.0)
Cardiac/Coronary care	175 (14.1)
Surgical	37 (3.0)
Neurological	25 (2.0)
Oncological	26 (2.1)
Other	10 (0.8)

In general, data completeness was good for all SAPS 3 variables, except for bilirubin and blood gas analysis results (**a-Table 1** of Appendix). For assessing the model's performance, the samples were divided into training (n=912,164) and validation (n=394,647) datasets. **Table 2** shows that the main patient characteristics and outcomes were comparable between the two datasets. The main reasons for ICU admission were medical (70.4% and 69.1%), followed by postoperative care after scheduled (20.8% and 21.3%) and emergency (7.0% and 7.3%) surgeries. Cardiovascular complications, infection/sepsis and neurological complications represented almost two-thirds of the medical admissions. At ICU admission, invasive mechanical ventilation was used in 12.8% and 13.1%, vasopressors were used in 14.0% and 14.8%, and renal replacement therapy was used in 2.0% and 1.9% of the training and validation datasets, respectively. The median SAPS 3 score was 44 points, with an interquartile range of (34–54) and (35–55) in the training and validation datasets, respectively. The outcomes were also comparable. The median ICU and hospital LOS were 3 (1–6) and 7 (4–15) days, respectively. The ICU and hospital mortality rates were 10.5% and 10.9%, and 15.7% and 16.0%, respectively.

Table 2. Characteristics of the study population in the training and validation datasets. *

Variable	Training (n=912,164)	Validation (n=394,647)	*P value*
Age, years	66 (49-77)	66 (50-78)	<0.001
Sex			0.353
Female	456,690 (50.1)	197,440 (50.0)	
Male	455,384 (49.9)	197,177 (50.0)	
Unknown	90 (0.0),	29 (0.0)	
SAPS 3, points	44 (34-54)	44 (35-55)	<0.001
SOFA, points	1 (0-3)	1 (0-4)	<0.001
Hospital LOS prior to ICU Admission, days	0 (0-1)	0 (0-1)	<0.001
Charlson Comorbidity Index, points	1 (0-2)	1 (0-2)	<0.001
Modified Frailty Index, points	1 (0-2)	1 (0-2)	<0.001
Admission Source			<0.001
Emergency room	421,327 (46.2)	171,562(43.5)	
Operating room	184,963 (20.2)	77,032 (19.5)	
Transfer from other hospital	96,360 (10.6)	53,241 (13.5)	
Ward/floor	77,302 (8.5)	33,062 (8.4)	
Cardiovascular intervention room	46,934 (5.1)	20,421 (5.2)	
Other ICU of your hospital	42,702 (4.7)	23,329 (5.9)	
Other	42,576 (4.7)	16,000 (4.0)	
Admission type			<0.001
Medical	641,819 (70.4)	272,763 (69.1)	
Cardiovascular	178,757 (19.6)	73,836 (18.7)	
Infection/sepsis	176,869 (19.4)	80,783 (20.5)	
Neurological	87,439 (9.6)	36,210 (9.2)	
Renal/Metabolic	39,964 (4.4)	16,543 (4.2)	
Respiratory	34,113 (3.7)	15,324 (3.9)	
Gastrointestinal	11,226 (1.2)	4,579 (1.2)	
Other	113,451 (12.4)	45,488 (11.5)	
Scheduled surgery	189,542 (20.8)	84,060 (21.3)	
Emergency surgery	63,783 (7.0)	29,002 (7.3)	
Invasive Support use at ICU admission (±1 h)			
Mechanical ventilation	116,742 (12.8)	51,527 (13.1)	<0.001
Vasopressors	127,806 (14.0)	58,236 (14.8)	<0.001
Dialysis	18,568 (2.0)	7,586 (1.9)	<0.001
Outcomes			
Unit LOS, days	3 (1-6)	3 (1-6)	0.927
ICU mortality	95,872 (10.5)	42,911 (10.9)	<0.001
Hospital LOS, days	7 (4-15)	7 (4-15)	<0.001
Hospital mortality	143,164 (15.7)	63,016 (16.0)	<0.001

SAPS: Simplified Acute Physiology Score; SOFA: Sequential Organ Failure Assessment; ICU: intensive care unit; LOS: length of stay.

*All continuous variables are reported as median (25%-75% interquartile range), and categorical variables are reported as number (%).

Performance of SAPS 3-SE in the training dataset

Table 3 reports the performance analyses of SAPS 3-SE in the training sample (n=912,164). In summary, the discrimination was very good [AUROC = 0.833 (95% CI, 0.832–0.835)]. However, the calibration was poor because SAPS 3-SE uniformly overestimated mortality in all risk ranges (**Figures 2 and 3**). The overall SMR was 0.874 (95% CI, 0.869–0.878), and the SRU was 0.928 (95% CI, 0.925–0.931). In the ICU-level analysis, the distributions of SMR and SRU individual values considering only the ICUs with at least 150 patients (n=1,194) are reported in **Table 4** and the funnel plot graphics (**Figures 4 and 5**). Most ICUs had SMRs and SRUs below 1.00.

Customization of the SAPS 3 score (SAPS 3-Custom) in the training dataset

Considering the poor calibration of SAPS 3-SE, we performed a first-level customization of SAPS 3 using the same function of the standard equation and vital status at hospital discharge as the dependent variable. The following equation was derived:

[-20.9447434 + ln(SAPS 3 score + 1) × 4.894223].

The SAPS 3-Custom model had good discrimination (AUROC: 0.833 (0.832–0.835) (**Table 3**)) and calibration in all risk ranges (**Figures 2 and 3**), which corrected the overestimation of SAPS 3-SE.

To correct for the observed overestimation of the expected number of days per survivor according to the severity of illness originally proposed by Rothen et al.(12), we derived a new set using the training dataset. To interpret our results compared to Rothen et al., we stratified patients into the same original SAPS 3 strata (**Table 5**). Overall, we found that the expected LOS per survivor was comparable to the original value (2.39 days vs. 2.3 days) in the lowest risk patients [stratum 1 (SAPS 3 = ≤24 points)], followed uniformly by lower values in the low–middle risk patients [strata 2–6 (SAPS 3 =25–74 points)]. The expected LOS per survivor was also comparable (26.1 days vs. 22.2 days) in the risk strata 7 group (SAPS 3 = 75–84 points) and, finally, by higher values in the highest risk patients [strata 8 and 9 (SAPS 3 ≥ 85 points)].

In the ICU-level analysis that considered only the ICUs with at least 150 patients (n=1,194), the SAPS 3-Custom yielded more even distributions of estimated SMR (median=0.95) and SRU (median=0.99) around the unit (**Table 4** and **Figures 4 and 5**).

Table 3. Performance of SAPS-SE and SAPS-Custom for all patients in the training dataset (n=912,164).

Score	Observed mortality	Predicted mortality	AUROC (95% CI)	SMR (95% CI)	SRU (95% CI)	Brier
SAPS 3-SE	15.69%	17.96%	0.833 (0.832-0.835)	0.874 (0.869-0.878)	0.928 (0.925 – 0.931)	0.099
SAPS 3-Custom		15.70%		1.000 (0.995-1.005)	1.003 (1.000 – 1.006)	0.098

Table 4. Distribution of SMR and SRU individual values in the ICU-level analysis considering only ICUs with at least 150 patients in the training dataset.

	Training dataset (ICUs=1,194; Patients=910,954)			
	SMR		SRU	
Percentile	SAPS 3-SE	SAPS 3-Custom	SAPS 3-SE	SAPS 3-Custom
5%	0.27	0.31	0.53	0.58
10%	0.38	0.43	0.60	0.65
25%	0.54	0.61	0.73	0.79
33%	0.63	0.72	0.78	0.85
50%	0.83	0.95	0.92	0.99
66%	1.05	1.20	1.09	1.17
75%	1.20	1.39	1.30	1.38
90%	1.72	1.98	1.92	2.07
95%	2.02	2.30	2.41	2.62

SAPS 3-SE: Simplified Acute Physiology Score 3, original standard equation; SAPS 3-CE: SAPS 3, customized equation; AUROC: area under the receiver operating curve; CI: confidence interval; SMR: standardized mortality rate; SRU: standardized resource use rate.

Table 5. Length of stay (LOS) per surviving patient, stratified by SAPS 3 in patients (n=910,954) admitted to ICUs with ≥ 150 admissions (n=1,194) for the purposes of SRU estimation.

SAPS 3 Stratum	SAPS 3 (points)	Patients (n)	Survivors (n)	Total ICU LOS (days)	Customized LOS per survivor (days)	Original LOS per survivor (days) (12)
1	0–24	47,468	46,732	111,844	2.39	2.3
2	25–34	187,972	183,240	547,511	2.99	3.2
3	35–44	236,529	223,486	926,043	4.14	4.3
4	45–54	210,140	182,556	1,096,248	6.00	7.2
5	55–64	118,909	85,600	808,587	9.45	11.0
6	65–74	56,196	28,952	466,770	16.12	16.6
7	75–84	27,978	9,654	251,954	26.10	22.2
8	85–94	13,405	3,154	117,785	37.34	29.4
9	≥ 95	6,991	910	52,001	57.14	39.0

LOS: length of stay; SAPS 3: Simplified Acute Physiology Score 3; SRU: standardized resource use rate. ICU: intensive care unit

To update the average number of expected LOS per survivor in our dataset for the purpose of SRU estimation, we stratified patients into the nine SAPS 3 classes (< 24; 25-34; 35-44; 45-54; 55-64; 65-74; 75-84; 85-94; ≥ 95 points) originally proposed by Rothen et al. (12). First, the average number of resources expected to produce a survivor in each stratum was estimated by dividing the sum of the ICU LOS of all patients in each stratum by the number of surviving patients in that stratum.

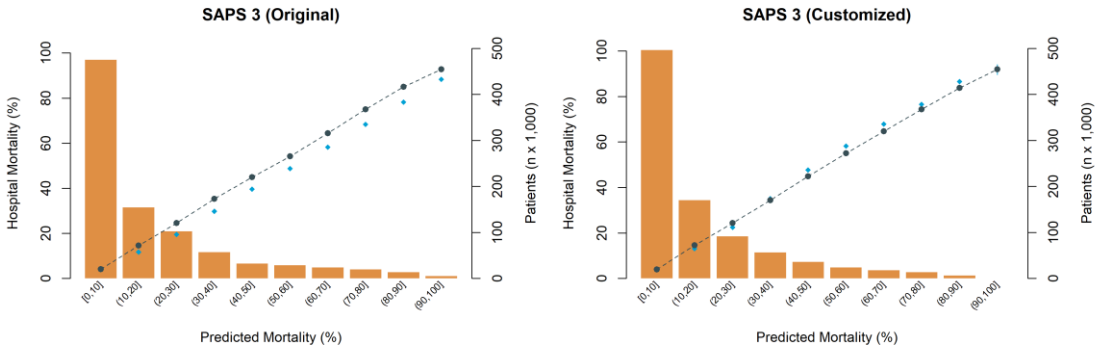


Figure 2. Calibration curves for the original (SAPS 3-SE) and customized (SAPS 3-Custom) equations in the training dataset (n=912,164). Patients were stratified into equal 10%-risk deciles. The columns represent the number of patients per decile. The thin dotted line represents a perfect fit. The observed mortality in each decile is represented by blue diamonds with 95% confidence intervals.

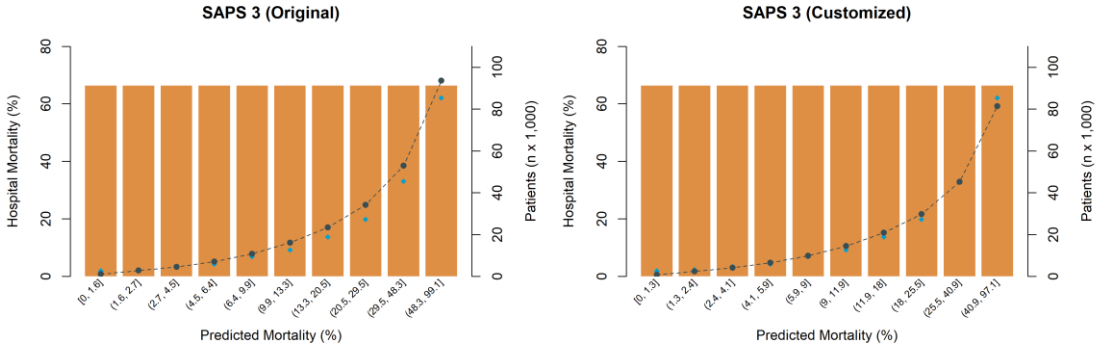


Figure 3. Calibration curves for the original (SAPS 3-SE) and customized (SAPS 3-Custom) equations in the training dataset (n=912,164). Patients were equally stratified into deciles. The columns represent the number of patients per decile. The thin dotted line represents a perfect fit. The observed mortality in each decile is represented by blue diamonds with 95% confidence intervals.

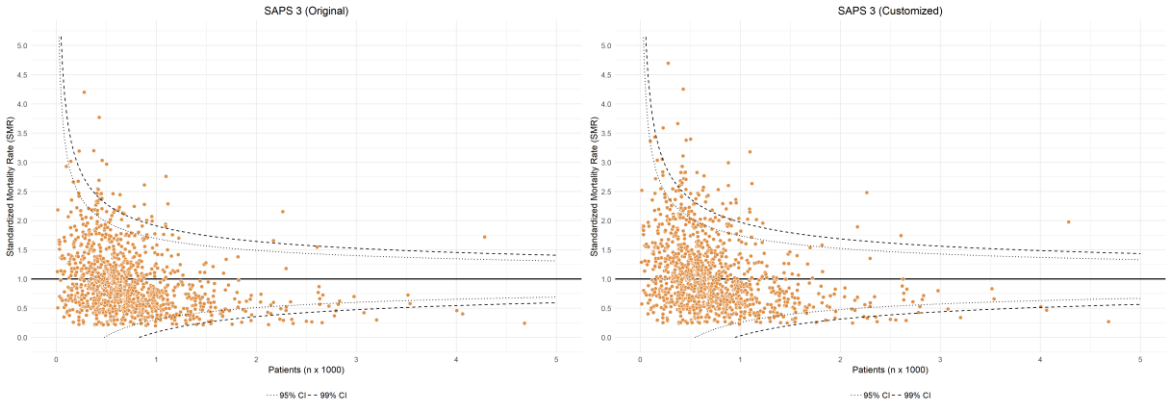


Figure 4. Funnel plot graphs of individual SMR values using the original (SAPS 3-SE) and SAPS 3-Custom equations in ICUs with more than 150 admissions (n=1,194) in the training dataset. The dotted thin and dashed lines represent the 95% and 99% confidence intervals, respectively.

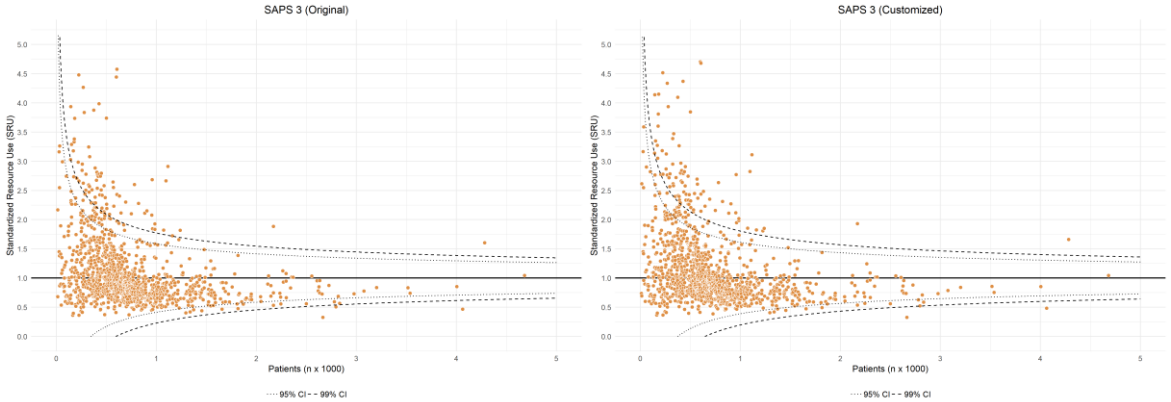


Figure 5. Funnel plot graphs of individual SRU values using the original (SAPS 3-SE) and SAPS 3-Custom expected LOS per surviving patient according to the severity of illness in ICUs with greater than 150 admissions (n=1,194) in the training dataset. The thin dotted and dashed lines represent the 95% and 99% confidence intervals, respectively.

Performance of SAPS 3-SE and SAPS 3-Custom in the validation dataset

We report the performance of SAPS 3-SE and SAPS 3-Custom for all patients in the validation dataset in **Table 6** and **Figures 6 and 7,** with comparable results to those observed in the training dataset. Despite good accuracy, SAPS 3-SE significantly overestimated the observed mortality in all risk ranges, whereas SAPS 3-Custom was well calibrated. The subgroup analysis of medical and surgical patients revealed similar results **(Table 6** and **Figures 8-11).** The ICU-level analysis (n= **(Table 7 and Figures 12 and 13)** with distribution patterns for the SMR and the SRU confirmed that the SAPS 3-Custom was more appropriate for evaluating ICU performance than the SAPS 3-SE.

Table 6. Performance of SAPS-SE and SAPS-Custom for all (n=394,647), medical (n=272,763) and surgical (n=113,062) patients in the validation dataset.

Patients	Observed mortality	Score	Predicted mortality	AUROC (95% CI)	SMR (95% CI)	SRU (95% CI)	Brier
All	15.97%	SAPS 3-SE	18.87%	0.835 (0.833 - 0.836)	0.846 (0.840 - 0.853)	0.901 (0.897 - 0.905)	0.101
		SAPS 3-Custom	16.49%		0.969 (0.961 - 0.976)	0.970 (0.966 - 0.975)	0.099
Medical	18.93%	SAPS 3-SE	23.01%	0.834 (0.833 - 0.836)	0.823 (0.816 - 0.830)	0.920 (0.915 – 0.924)	0.114
		SAPS 3-Custom	20.09%		0.942 (0.934 - 0.951)	0.994 (0.989 – 1.000)	0.113
Surgical	8.44%	SAPS 3-SE	9.96%	0.847 (0.843 - 0.851)	0.848 (0.831 – 0.865)	0.799 (0.792 – 0.806)	0.062
		SAPS 3-Custom	8.74%		0.966 (0.947 - 0.986)	0.849 (0.842 – 0.857)	0.061

SAPS 3-SE: Simplified Acute Physiology Score 3, original standard equation; SAPS 3-CE: SAPS 3, customized equation; AUROC: area under the receiver operating curve; CI: confidence interval; SMR: standardized mortality rate; SRU: standardized resource use rate.

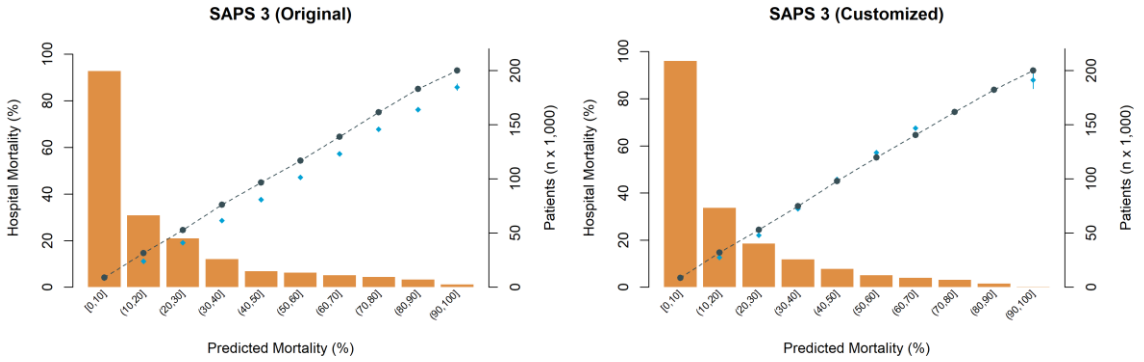


Figure 6. Calibration curves for the original (SAPS 3-SE) and customized (SAPS 3-Custom) equations in the validation dataset (n=394,647). Patients were stratified into equal 10%-risk deciles. The columns represent the number of patients per decile. The thin dotted line represents a perfect fit. The observed mortality in each decile is represented by blue diamonds with 95% confidence intervals.

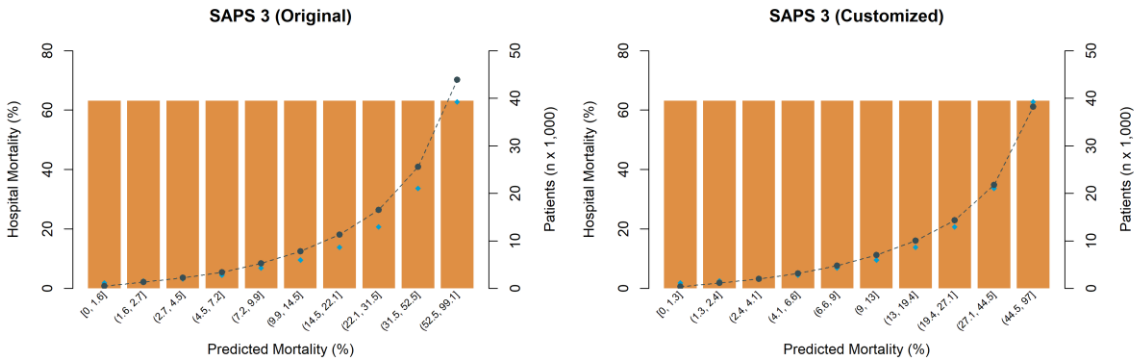


Figure 7. Calibration curves for the original (SAPS 3-SE) and customized (SAPS 3-Custom) equations in the validation dataset (n=394,647). Patients were equally stratified into deciles. The columns represent the number of patients per decile. The thin dotted line represents a perfect fit. The observed mortality in each decile is represented by blue diamonds with 95% confidence intervals.

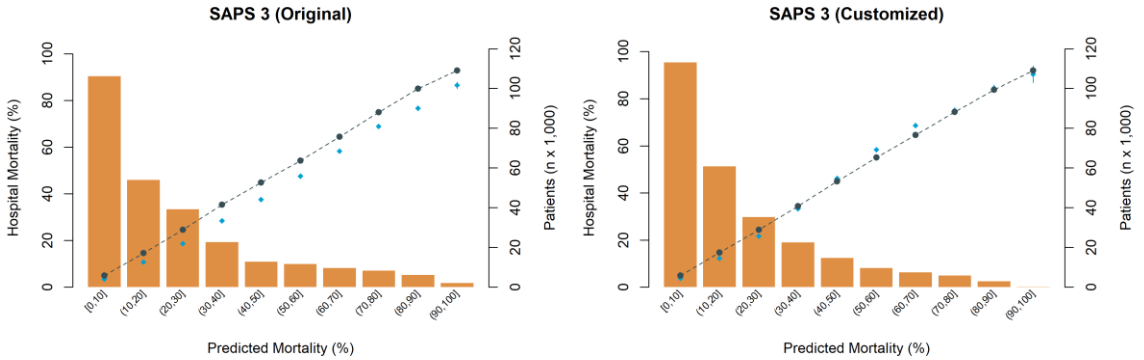


Figure 8. Calibration curves for the original (SAPS 3-SE) and customized (SAPS 3-Custom) equations for medical patients in the validation dataset (n=272,763). Patients were stratified into equal 10%-risk deciles. The columns represent the number of patients per decile. The thin dotted line represents a perfect fit. The observed mortality in each decile is represented by blue diamonds with 95% confidence intervals.

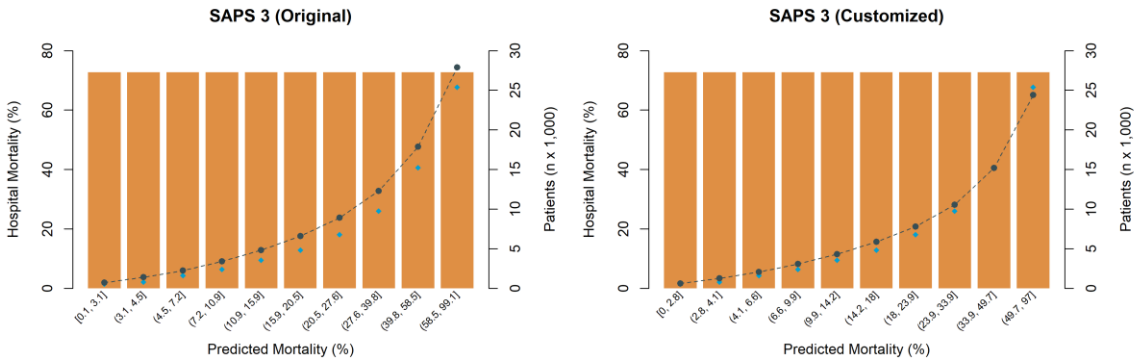


Figure 9. Calibration curves for the original (SAPS 3-SE) and customized (SAPS 3-Custom) equations for medical patients in the validation dataset (n=272,763). Patients were equally stratified into deciles. The columns represent the number of patients per decile. The thin dotted line represents a perfect fit. The observed mortality in each decile is represented by blue diamonds with 95% confidence intervals.

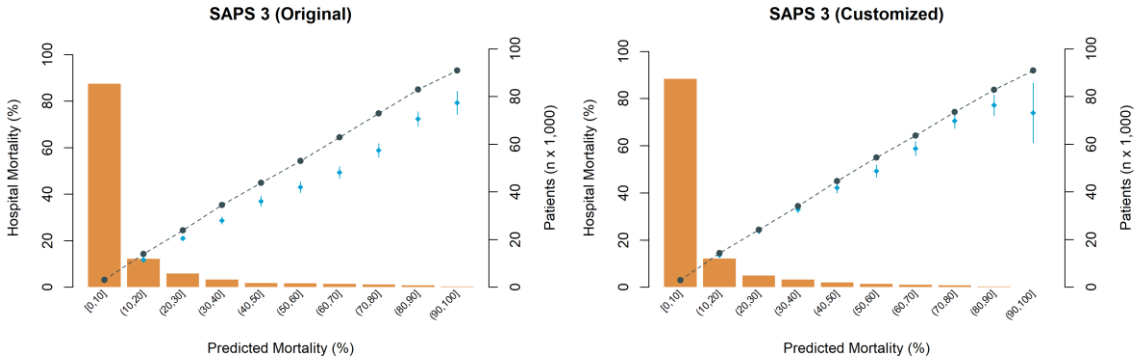


Figure 10. Calibration curves for the original (SAPS 3-SE) and customized (SAPS 3-Custom) equations for surgical patients in the validation dataset (n=113,062). Patients were stratified into equal 10%-risk deciles. The columns represent the number of patients per decile. The thin dotted line represents a perfect fit. The observed mortality in each decile is represented by blue diamonds with 95% confidence intervals.

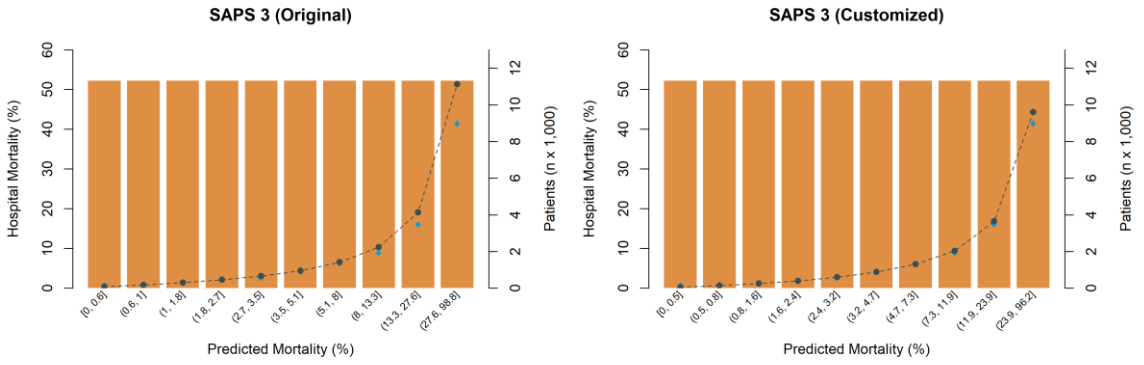


Figure 11. Calibration curves for the original (SAPS 3-SE) and customized (SAPS 3-Custom) equations for surgical patients in the validation dataset (n=113,062). Patients were equally stratified into deciles. The columns represent the number of patients per decile. The thin dotted line represents a perfect fit. The observed mortality in each decile is represented by blue diamonds with 95% confidence intervals.

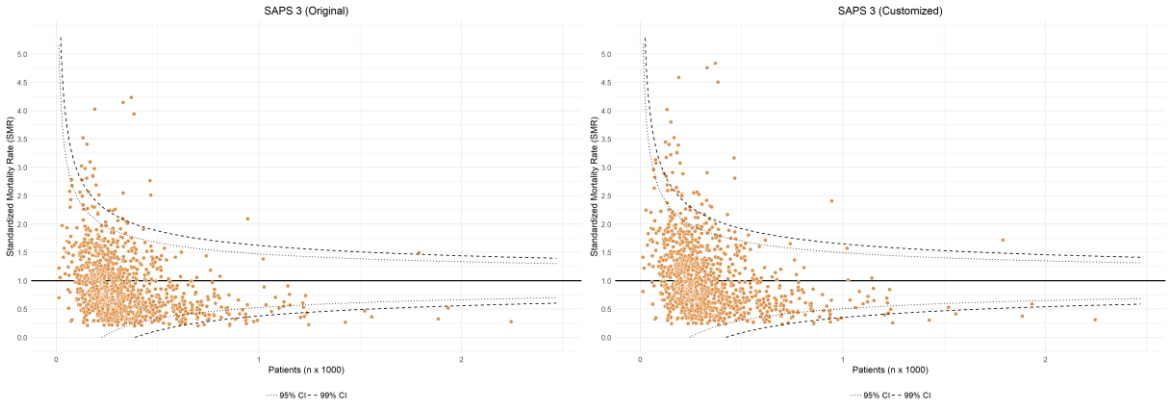


Figure 12. Funnel plot graphs of individual SMR values using the original (SAPS 3-SE) and SAPS 3-Custom equations in ICUs with more than 150 admissions (n=1,149) in the validation dataset. The thin dotted and dashed lines represent the 95% and 99% confidence intervals, respectively.

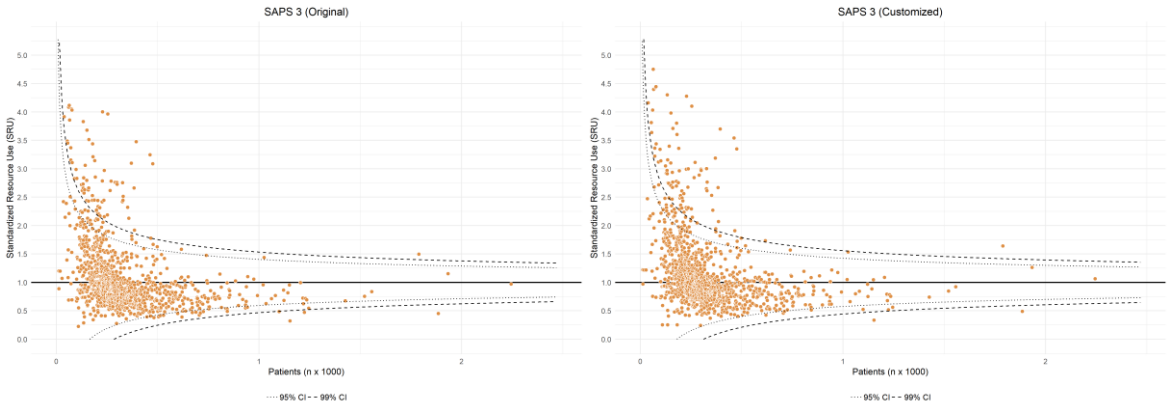


Figure 13. Funnel plot graphs of individual SRU values using the original (SAPS 3-SE) and SAPS 3-Custom expected LOS per surviving patient according to the severity of illness in ICUs with more than 150 admissions (n=1,149) in the validation dataset. The thin dotted and dashed lines represent the 95% and 99% confidence intervals, respectively.

Table 7. Distribution of SMR and SRU individual values in the ICU-level analysis considering only ICUs with at least 150 patients in the validation dataset.

	Validation dataset (ICUs=1,149; Patients=393,347)			
	SMR		SRU	
Percentile	SAPS 3-SE	SAPS 3-Custom	SAPS 3-SE	SAPS 3-Custom
5%	0.27	0.31	0.50	0.53
10%	0.33	0.38	0.57	0.61
25%	0.52	0.59	0.71	0.77
33%	0.60	0.69	0.76	0.83
50%	0.77	0.89	0.91	0.98
66%	1.05	1.20	1.08	1.17
75%	1.20	1.38	1.28	1.37
90%	1.64	1.88	1.95	2.07
95%	1.95	2.24	2.61	2.76

SAPS 3-SE: Simplified Acute Physiology Score 3, original standard equation; SAPS 3-CE: SAPS 3, customized equation; AUROC: area under the receiver operating curve; CI: confidence interval; SMR: standardized mortality rate; SRU: standardized resource use rate.

Discussion and Interpretation of the Study Results

The performance of any prognostic score should be reassessed regularly for several reasons, including changes in case mix, ICU admission and discharge policies and patient clinical management, among others. Although the score's discrimination tends to remain relatively stable over time, calibration can deteriorate, and recalibration is required to maintain its performance.(20–23) In addition, recalibration may be required to adjust for local specificities in a given country or region.(24, 25) According to the Brazilian ICU Registry data, there was a trend toward progressively lower SMRs and SRUs in the years preceding the COVID-19 pandemic.(13) Although these results may suggest improvements in the efficiency of the ICUs participating in the registry, a reassessment of SAPS 3 performance was already opportune. However, the COVID-19 pandemic imposed serious restrictions on any evaluation of model performance during that period. Moreover, some studies demonstrated the poor performance of prognostic scores in COVID-19 patients.(26–28) The present study evaluated the performance of SAPS 3 to predict outcomes in a large contemporary cohort of patients admitted to the ICUs participating in the Brazilian ICU Registry, which is comprised of approximately 50% of all adult ICU beds in Brazil, with a balanced mix of private and public ICUs from the five national regions. To the best of our knowledge, this database is the largest ICU registry worldwide.

We demonstrated that the SAPS 3-SE was mis-calibrated to predict outcomes in the training and validation datasets. In general, the SAPS 3-SE significantly overestimated hospital mortality in all risk ranges (including the subgroup analysis of medical and surgical patients). We also demonstrated that the expected LOS per surviving patient according to the severity of illness proposed by Rothen et al.(12) were also not appropriate in our database, with lower values in the low–middle-risk classes and higher values in the highest-risk patients. Therefore, lower than expected (<1.0) SMR and SRU values were observed in most ICUs.

The abovementioned results indicated that recalibration was needed. Because the discrimination was still very good [AUROC=0.833 (0.832–0.835)], we performed a first-level recalibration of the score. The derived equation of SAPS 3-Custom had a very good calibration in all risk ranges in the training and validation datasets, including the medical and surgical patient subgroups, which corrected the overestimation of the original standard equation. We also derived "customized" LOS per survivor by stratifying patients as originally proposed. Comparing our results with Rothen et al.(12), we found that resource use was overestimated in the low–middle-risk classes and underestimated in the highest-risk patients. The new customized values substantially improved the accuracy of the SRU estimation.

The present study has potential practical implications because we demonstrated that the SAPS 3-SE overestimated mortality and resource use in more recent years. The recalibration procedures corrected these overestimations. Therefore, the study results support the use of SAPS 3-Custom over SAPS 3-SE to evaluate ICU performance and efficiency and for benchmarking in Brazilian ICUs. However, reevaluations of the performance of the SAPS 3-Custom and its customized LOS per survivor should be regularly performed to assess whether they remain accurate in the near future.

References

1. Salluh JIF, Soares M: ICU severity of illness scores: APACHE, SAPS and MPM. *Curr Opin Crit Care* 2014; 20:557–565

2. Keegan MT, Gajic O, Afessa B: Severity of illness scoring systems in the intensive care unit. *Crit Care Med* 2011; 39:163–169

3. Keegan MT, Soares M: What every intensivist should know about prognostic scoring systems and risk-adjusted mortality. *Rev Bras Ter Intensiva* 2016; 28:264–269

4. Moreno RP, Metnitz PGH, Almeida E, et al.: SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med* 2005; 31:1345–1355

5. Ferrando-Vivas P, Jones A, Rowan KM, et al.: Development and validation of the new ICNARC model for prediction of acute hospital mortality in adult critical care. *J Crit Care* 2016;

6. Soares M, Dongelmans DA: Why should we not use APACHE II for performance measurement and benchmarking? *Rev Bras Ter Intensiva* 2017; 29:268–270

7. Recomendacoes [Internet]. *AMIB* [cited 2021 Oct 15] Available from: https://www.amib.org.br/informacao/recomendacoes/

8. Moralez GM, Rabello LSCF, Lisboa TC, et al.: External validation of SAPS 3 and MPM0-III scores in 48,816 patients from 72 Brazilian ICUs. *Ann Intensive Care* 2017; 7:53

9. Maccariello E, Valente C, Nogueira L, et al.: SAPS 3 scores at the start of renal replacement therapy predict mortality in critically ill patients with acute kidney injury. *Kidney Int* 2010; 77:51–56

10. Soares M, Lobo SMA, Torelly AP, et al.: Outcomes of cancer patients admitted to Brazilian intensive care units with severe acute kidney injury. *Rev Bras Ter Intensiva* 2010; 22:236–244

11. Zampieri FG, Soares M, Salluh JIF: How to evaluate intensive care unit performance during the COVID-19 pandemic. *Rev Bras Ter Intensiva* 2020; 32:203–206

12. Rothen HU, Stricker K, Einfalt J, et al.: Variability in outcome and resource use in intensive care units. *Intensive Care Med* 2007; 33:1329–1336

13. UTIs Brasileiras [Internet]. [cited 2018 May 27] Available from: http://www.utisbrasileiras.com.br

14. Soares M, Borges LP, Bastos LDSL, et al.: Update on the Epimed Monitor Adult ICU Database: 15 years of its use in national registries, quality improvement initiatives and clinical research. *Crit Care Sci* 2024; 36:e20240150en

15. Vincent JL, Moreno R, Takala J, 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; 22:707–710

16. Charlson ME, Pompei P, Ales KL, et al.: A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40:373–383

17. Zampieri FG, Iwashyna TJ, Viglianti EM, et al.: Association of frailty with short-term outcomes, organ support and resource use in critically ill patients. *Intensive Care Med* 2018; 44:1512–1520

18. Brier GW: Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 1950; 75:1–3

19. Spiegelhalter DJ: Funnel plots for comparing institutional performance. *Stat Med* 2005; 24:1185–1202

20. Soares M, Salluh JIF: Validation of the SAPS 3 admission prognostic model in patients with cancer in need of intensive care. *Intensive Care Med* 2006; 32:1839–1844

21. Zimmerman JE, Kramer AA, McNair DS, et al.: Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297–1310

22. Genu DHS, Lima-Setta F, Colleti J, et al.: Multicenter validation of PIM3 and PIM2 in Brazilian pediatric intensive care units. *Front Pediatr* 2022; 10:1036007

23. Davis SE, Lasko TA, Chen G, et al.: Calibration Drift Among Regression and Machine Learning Models for Hospital Mortality. *AMIA Annu Symp Proc AMIA Symp* 2017; 2017:625–634

24. B M, E S, R M, et al.: Austrian validation and customization of the SAPS 3 Admission Score [Internet]. *Intensive Care Med* 2009; 35[cited 2024 Aug 6] Available from: https://pubmed.ncbi.nlm.nih.gov/18846365/

25. Sy L, So K, K J, et al.: Validation of SAPS3 admission score and its customization for use in Korean intensive care unit patients: a prospective multicentre study [Internet]. *Respirol Carlton Vic* 2013; 18[cited 2024 Aug 6] Available from: https://pubmed.ncbi.nlm.nih.gov/23663287/

26. Higgins TL, Stark MM, Henson KN, et al.: Coronavirus Disease 2019 ICU Patients Have Higher-Than-Expected Acute Physiology and Chronic Health Evaluation-Adjusted Mortality and Length of Stay Than Viral Pneumonia ICU Patients. *Crit Care Med* 2021; 49:e701–e706

27. Metnitz PGH, Moreno RP, Fellinger T, et al.: Evaluation and calibration of SAPS 3 in patients with COVID-19 admitted to intensive care units. *Intensive Care Med* 2021; 47:910–912

28. Kurtz P, Bastos LSL, Salluh JIF, et al.: SAPS-3 performance for hospital mortality prediction in 30,571 patients with COVID-19 admitted to ICUs in Brazil. *Intensive Care Med* 2021; 47:1047–1049

Appendix

a-Table 1. SAPS 3 missing variables (patients: 1,306,811)

Missing variables	n (%)
Age	0
Comorbidities	218,977 (16.8)
Admission source	15,108 (1.2)
Length of hospital stay before ICU admission	0
Vasoactive drugs before ICU admission	27,896 (2.1)
SAPS 3 admission diagnoses	25,842 (2.0)
Admission type	25,842 (2.0)
Surgical status	25,503 (2.0)
Acute infection	25,503 (2.0)
Glasgow coma scale	188,636 (14.4)
Total bilirubin	713,219 (54.6)
Body temperature	104,323 (8.0)
Creatinine	173,222 (13.3)
Heart rate	51,962 (4.0)
Leukocyte	164,512 (12.6)
pH	766,302 (58.6)
Platelets	167,679 (12.8)
Systolic blood pressure	53,735 (4.1)
PaO_2/FiO_2	904,493 (69.2)
PaO_2	773,686 (59.2)
PaO_2/FiO_2in ventilated patients*	57,705 (34.3)
PaO_2 in ventilated patients*	36,485 (21.7)
Ventilatory support	27,896 (2.1)

SAPS: Simplified Acute Physiology Acute Score; ICU: intensive care unit

Ventilated patients at ICU admission (n=168,269)