Predictive Performance of the Simplified Acute Physiology Score (SAPS) II and the Initial Sequential Organ Failure Assessment (SOFA) Score in Acutely Ill Intensive Care Patients: Post-Hoc Analyses of the SUP-ICU Inception Cohort Study

Anders Granholm*, Morten Hylander Møller, Mette Krag, Anders Perner, Peter Buhl Hjortrup

Department of Intensive Care, Copenhagen University Hospital – Rigshospitalet, Copenhagen, Denmark

* andersgran@gmail.com

Abstract

Purpose

Severity scores including the Simplified Acute Physiology Score (SAPS) II and the Sequential Organ Failure Assessment (SOFA) score are used in intensive care units (ICUs) to assess disease severity, predict mortality and in research. We aimed to assess the predictive performance of SAPS II and the initial SOFA score for in-hospital and 90-day mortality in a contemporary international cohort.

Methods

This was a post-hoc study of the Stress Ulcer Prophylaxis in the Intensive Care Unit (SUP-ICU) inception cohort study, which included acutely ill adults from ICUs across 11 countries (n = 1034). We compared the discrimination of SAPS II and initial SOFA scores, compared the discrimination of SAPS II in our cohort with the original cohort, assessed the calibration of SAPS II customised to our cohort, and compared the discrimination for 90-day mortality vs. in-hospital mortality for both scores. Discrimination was evaluated using areas under the receiver operating characteristics curves (AUROC). Calibration was evaluated using Hosmer-Lemeshow's goodness-of-fit $\hat{C}$-statistic.

Results

AUROC for in-hospital mortality was 0.80 (95% confidence interval (CI) 0.77–0.83) for SAPS II and 0.73 (95% CI 0.69–0.76) for initial SOFA score (P<0.001 for the comparison). Calibration of the customised SAPS II for predicting in-hospital mortality was adequate (P = 0.60). Discrimination of SAPS II was reduced compared with the original SAPS II validation sample (AUROC 0.80 vs. 0.86; P = 0.001). AUROC for 90-day mortality was 0.79 (95% CI

0.76–0.82; P = 0.74 for comparison with in-hospital mortality) for SAPS II and 0.71 (95% CI 0.68–0.75; P = 0.66 for comparison with in-hospital mortality) for the initial SOFA score.

Conclusions

The predictive performance of SAPS II was similar for in-hospital and 90-day mortality and superior to that of the initial SOFA score, but SAPS II's performance has decreased over time. Use of a contemporary severity score with improved predictive performance may be of value.

Introduction

Severity scoring systems are frequently used in intensive care units (ICUs) to assess disease severity, predict mortality, compare ICU performances, and in research [1–3]. The scores generally belong to one of two groups: 1) scores that aim to predict mortality based on parameters obtained upon ICU admission or during the first 24 hours of ICU stay, or 2) scores that aim to quantify the level of organ dysfunction daily during ICU stay [1–3].

The Simplified Acute Physiology Score (SAPS) II was developed and validated in a European and North American cohort (n = 12,997), and published in 1993 [4]. The score includes 17 variables collected during the first 24 hours of ICU stay. Based on the sum of the score the in-hospital mortality risk can be estimated [4].

The Sequential Organ-Failure Assessment (SOFA) score was developed by an expert panel in 1996 [5]. The worst values recorded for every 24-hour period in the ICU is used to assign a score of 0–4 for six organ systems. Although the score was developed to describe changes in organ dysfunction throughout ICU stay, and not to predict mortality, an association between increasing initial organ-specific SOFA scores and mortality has been suggested [5].

The predictive performance of SAPS II has been evaluated in multiple studies, generally showing acceptable discrimination but poor calibration in other populations than the original one [6–15]. In a systematic review, Minne et al. found that the initial SOFA score showed good to excellent discrimination between subsequent hospital survivors and non-survivors, with an area under the receiver operating characteristic curve (AUROC) similar to that of SAPS II [16]. The authors concluded that the SOFA score performed similarly to the SAPS II, however their systematic review included only one small single-centre study directly comparing the two scores, and this needs to be further assessed in larger multi-centre studies.

The value of the mortality prediction scores appear to deteriorate over time [13, 17]. SAPS II has been re-assessed in several national studies [6–15], but its performance is affected by case-mix [2, 18] and national differences. Performance in multinational, contemporary cohorts has, however, been less studied. Additionally, SAPS II was developed to predict in-hospital mortality. In-hospital mortality is considered an inadequate outcome measure today, as it is affected by hospital discharge practice [19–21]. Use of a long-term, fixed-time mortality endpoint might affect the predictive performance of these scores.

The aim of this study was to assess the predictive performances of SAPS II and the initial SOFA score for in-hospital mortality vs. 90-day mortality in a contemporary international general ICU cohort of acutely ill adults. We hypothesised that the predictive performance of SAPS II and the initial SOFA score for in-hospital mortality in ICUs today would be low, and that the predictive performance for 90-day mortality would be superior to that for in-hospital mortality.

Methods

This study was a post-hoc study of the Stress Ulcer Prophylaxis in the Intensive Care Unit (SUP-ICU) 7-day international inception cohort study [22], in which all acutely admitted patients aged 18 year or above were included in 97 ICUs in 11 countries during a single 7-day period between 1st December 2013 and 30th April 2014. Exclusion criteria were gastrointestinal (GI) bleeding upon ICU admission and previous ICU admission during the index hospital admission. The SUP-ICU cohort study was approved by the Danish Data Protection Agency (No. 30–1115) and the Danish Health and Medicines Authorities (No. 3-3013-463/1/). Ethical approval was not applicable owing to the non-interventional (observational) design. Patient information in the SUP-ICU cohort study database was anonymised and de-identified.

For the present study, an internal protocol and statistical analysis plan was written prior to the analyses. We included all the 1034 patients in the SUP-ICU cohort study database and extracted the following data: Age, gender, SAPS II, initial SOFA score, comorbidities, interventions given at ICU admission, in-hospital mortality and 90-day mortality. The present manuscript was prepared according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement [23].

Registration of SAPS II and initial SOFA scores

Data were registered prospectively using a secure web-based electronic case report form, as previously reported [22]. Data for the SAPS II and initial SOFA scores were recorded during the first 24 hours of ICU admission, as defined in the original scores [4, 5]. The Glasgow Coma Scale (GCS) score was registered in accordance with the SAPS II definition, i.e. if patients were sedated, the last known value before sedation was recorded and used for both the SAPS II and the SOFA score.

Measures of predictive performance

For the evaluation of the predictive performances of the scores, we primarily focused on the discrimination of the scores. Discrimination is the ability of the score to separate patients that die from patients that live, and discrimination was assessed using AUROCs as described in the next section.

We compared the discrimination of the SAPS II and initial SOFA scores, and compared the discrimination of SAPS II in the present cohort with the original cohort. Additionally, we customised SAPS II by re-calibrating it to our cohort and assessed calibration (the prognostic accuracy of the model at different risk intervals, described in the next section) of the customised SAPS II. We did this to assess the association between SAPS II and mortality in the original cohort and in the present cohort.

Finally, we compared the discrimination when using 90-day vs. in-hospital mortality as the outcome measure for both the SAPS II and the initial SOFA score

Statistical analysis

We did all statistical analyses according to the analysis plan using SAS version 9.3 (SAS Institute Inc., Cary, NC, USA). As this was a post-hoc analysis of the SUP-ICU cohort study, a convenience sample was used and thus no sample size calculation was made.

We stratified patient characteristics on inclusion in the SUP-ICU cohort study by 90-day mortality and presented these as medians with interquartile ranges (IQR) for continuous data, and numbers (%) for categorical data [24]. We assessed differences by Mann–Whitney U test

and Chi^2-test. All statistical tests were two-tailed, and P-values < 0.05 were considered statistically significant.

We knew variables were missing in the SAPS II and initial SOFA scores [22], and as reported in the main SUP-ICU cohort publication, data were not *missing completely at random* (Little's test, P <0.001) [22]; thus multiple imputations for the missing SAPS II and SOFA scores were performed [25, 26]. Multiple imputations were performed using the fully conditional specification method and 10 imputed datasets. For patients with missing data, the complete scores were imputed, not their individual components, based on the following baseline variables: age; gender; SAPS II/initial SOFA score; chronic obstructive pulmonary disease, asthma or chronic lung disease; previous myocardial infarction; severe chronic heart failure (New York Heart Association Functional Classification class 3–4); chronic renal failure; liver cirrhosis or increased bilirubin; metastatic cancer; active haematological cancer; AIDS; mechanical ventilation at ICU admission; circulatory support at ICU admission; and renal replacement therapy at ICU admission. The discrimination of the two scores (the ability of the model to separate patients who dies from patients who lives) was assessed by their AUROCs. Comparison of differences in the AUROCs between SAPS II and the initial SOFA score according to in-hospital mortality was performed as described by DeLong et al [27]. The method by DeLong et al. is a non-parametric method for comparing different AUROCs, based on the theory of generalised U-statistics that generates a covariance matrix and produces a test-statistic with a Chi^2-distribution [27].

The standard error of the AUROC in the original SAPS II validation sample [4] was calculated from the 95% confidence interval [28] and used for the comparison. We used Chi^2-test to compare the differences in AUROCs between SAPS II for in-hospital mortality in the original SAPS II validation sample and in our cohort and to compare the differences in AUROC between in-hospital vs. 90-day mortality for SAPS II and the initial SOFA score in our cohort [29].

We performed a first-level customisation of SAPS II for in-hospital mortality in the SUP-ICU cohort, as the original SAPS II equation has already been reported to have deteriorated since its development [6–15]. In a first-level customisation, logistic regression analysis is used to derive a new equation while the variables included in the score and their weights are left unchanged.

The calibration (the prognostic accuracy of the model at different risk intervals) of the customised SAPS II was evaluated using the Hosmer-Lemeshow goodness-of-fit $\hat{C}$-statistic.

A post-hoc sensitivity analysis of complete cases only (instead of the multiple imputated dataset) was conducted during the peer-review process (S1 Appendix).

Results

All 1034 patients were included; in-hospital and 90-day mortality rates were 22.5% (95% confidence interval (CI) 20.0–25.2; 233/1027, 7 missing values) and 26.2% (95% CI 23.6–29.0; 271/1034), respectively.

One or more of the variables in the SAPS II and initial SOFA scores were missing for 17.4% (n = 180) and 23.4% (n = 245) of the included patients, respectively [22]. Thus, for 17.4% of the patients, multiple imputations were made for the full SAPS II scores, and for 23.4% of the patients, multiple imputations were made for the full initial SOFA scores.

Patients who died within 90 days were older, had more comorbidities, and higher SAPS II and initial SOFA scores than those who survived (Table 1).

Table 1. Patient characteristics at ICU admission.

Characteristic	All (n = 1034)	Alive 90 days after ICU admission (n = 763)	Dead 90 days after ICU admission (n = 271)	P-value*	Patients with missing values, n (%)
Age—years—median (IQR)	63 (48–74)	60 (45–71)	71 (60–79)	< 0.001	0 (0.0)
Male gender—no. (%)	576 (55.7)	421 (55.2)	155 (57.2)	0.57	0 (0.0)
Initial SOFA score—median (IQR)	6 (4–8)	6 (3–7)	8 (5–11)	< 0.001	245 (23.4)
SAPS II—median (IQR)	42 (31–54)	37 (28–48)	57 (47–68)	< 0.001	180 (17.4)
Chronic obstructive pulmonary disease, asthma or other chronic lung disease—no. (%)	205 (19.8)	138 (18.1)	67 (24.7)	0.02	0 (0.0)
Previous myocardial infarction—no. (%)	103 (10.0)	63 (8.3)	40 (14.8)	0.002	0 (0.0)
Severe chronic heart failure (NYHA 3–4)—no. (%)	56 (5.4)	22 (2.9)	34 (12.6)	< 0.001	0 (0.0)
Chronic renal failure—no. (%)	74 (7.2)	45 (5.9)	29 (10.7)	0.008	0 (0.0)
Liver cirrhosis or increased bilirubin (> 33 µmol/l)—no. (%)	124 (12.5)	79 (10.8)	45 (17.2)	0.007	38 (3.7)
Metastatic cancer—no. (%)	46 (4.5)	23 (3.0)	23 (8.5)	< 0.001	0 (0.0)
Active hematologic cancer—no. (%)	36 (3.5)	17 (2.2)	19 (7.0)	< 0.001	0 (0.0)
AIDS—no. (%)	3 (0.3)	2 (0.3)	1 (0.4)	1.0	0 (0.0)
Immunosuppression**—no. (%)	50 (4.8)	33 (4.3)	17 (6.3)	0.20	0 (0.0)
Coagulopathy on ICU admission***—no. (%)	128 (12.4)	68 (8.9)	60 (22.1)	< 0.001	0 (0.0)
Comorbidities—no. (%)					
0	501 (48.5)	414 (54.3)	87 (32.1)	< 0.001	0 (0.0)
1	318 (30.8)	232 (30.4)	86 (31.7)	0.68	0 (0.0)
2	153 (14.8)	96 (12.6)	57 (21.0)	< 0.001	0 (0.0)
3	46 (4.5)	16 (2.1)	30 (11.1)	< 0.001	0 (0.0)
> 3	16 (1.6)	5 (0.7)	11 (4.1)	< 0.001	0 (0.0)
Mechanical ventilation on ICU admission—no. (%)	544 (52.6)	377 (49.4)	167 (61.6)	< 0.001	0 (0.0)
Circulatory support on ICU admission—no. (%)	469 (45.7)	293 (38.6)	176 (65.7)	< 0.001	7 (0.7)
Renal replacement therapy on ICU admission—no. (%)	70 (6.8)	36 (4.7)	34 (12.6)	< 0.001	0 (0.0)

* For the comparison of patients stratified by 90-day mortality.

** Defined as treatment with at least 0.3 mg/kg/day of prednisolone equivalent for one month or longer in the 6 months prior to ICU admission.

*** Platelets < 50 * 10^9/l (50,000 mm^3) and/or International Normalised Ratio (INR) > 1.5 during current hospital admission.

IQR: Interquartile range; NYHA: New York Heart Association Functional Classification; AIDS: Acquired Immune Deficiency Syndrome.

doi:10.1371/journal.pone.0168948.t001

SAPS II vs. initial SOFA score

The ROC curves for SAPS II and the initial SOFA score for in-hospital mortality are shown in Fig 1. The AUROCs for SAPS II was higher than that for initial SOFA scores for in-hospital mortality; 0.80 (95% CI 0.77–0.83) and 0.73 (95% CI 0.69–0.76), respectively (P < 0.001).

SAPS II in the present cohort vs. SAPS II in the original cohort

Discrimination of SAPS II for in-hospital mortality in the present cohort was lower than that of the original SAPS II validation sample (AUROC 0.80 vs. AUROC 0.86; P = 0.001).

Calibration of the customised SAPS II was good (Fig 2). The correlation between SAPS II and in-hospital mortality risk in our cohort and the original SAPS II validation sample is shown in Fig 3 and Table 2.
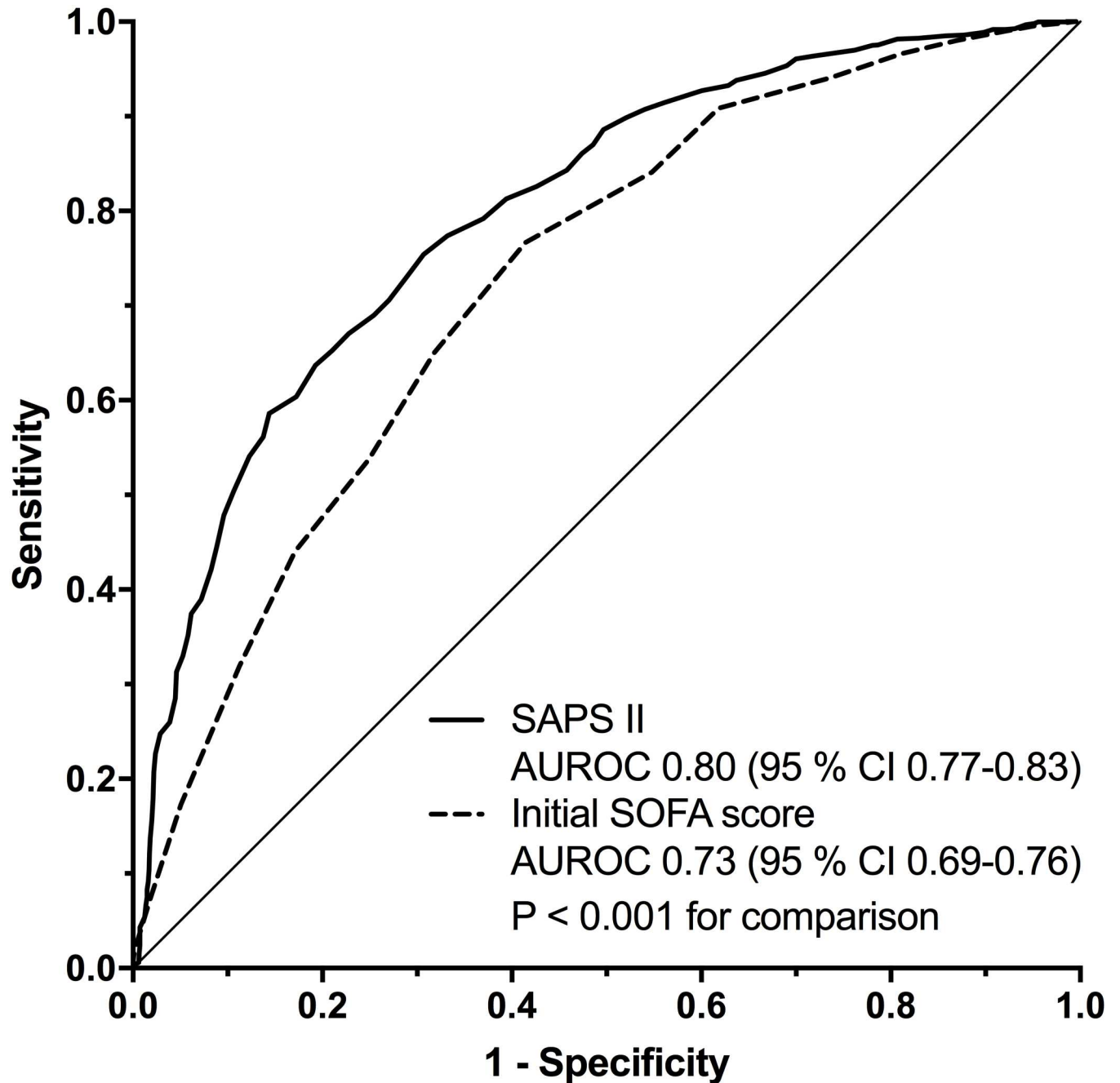
Fig 1. Discrimination of SAPS II and the initial SOFA score for in-hospital mortality. Receiver operating characteristics curves for SAPS II and the initial SOFA score for in-hospital mortality. AUROC: Area under the receiver operating characteristics curve.

In-hospital vs. 90-day mortality

The ROC curves for SAPS II and initial SOFA scores for in-hospital and 90-day mortality are shown in Fig 4. For SAPS II the AUROCs were 0.80 (95% CI 0.77–0.83) and 0.79 (95% CI 0.76–0.82) for in-hospital mortality and 90-day mortality, respectively (P = 0.74 for the comparison). For the initial SOFA score the AUROCs were 0.73 (95% CI 0.69–0.76) and 0.71 (95%
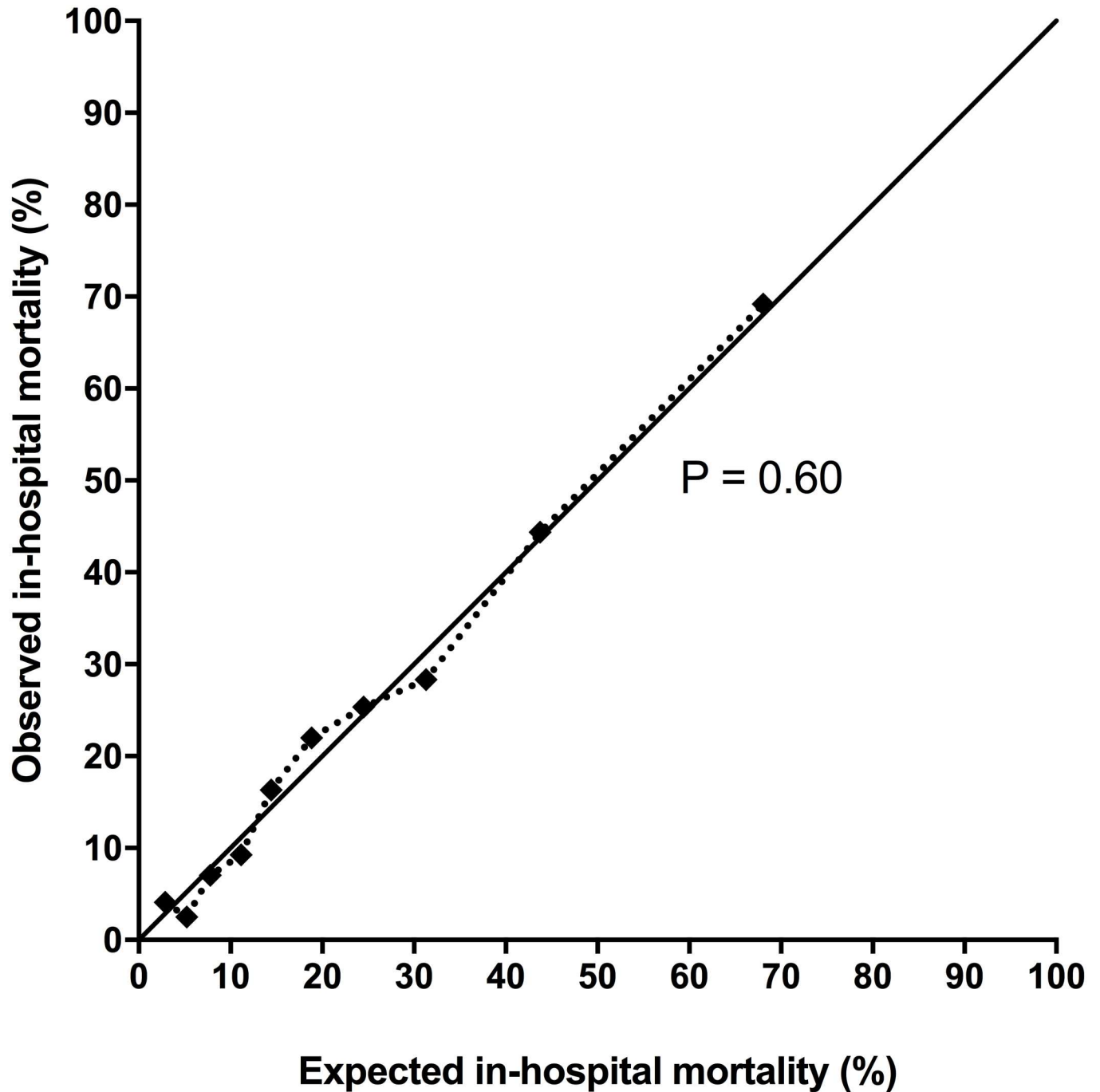
Fig 2. Calibration of the customised SAPS II for in-hospital mortality. Calibration curve of SAPS II customised by logistic regression analysis for in-hospital mortality and assessed by the Hosmer-Lemeshow goodness-of fit $\hat{C}$-statistic. The full line included on the figure is the line of equality (expected = observed mortality).

doi:10.1371/journal.pone.0168948.g002

CI 0.68–0.75) for in-hospital mortality and 90-day mortality, respectively (P = 0.66 for the comparison).

The results of the post-hoc sensitivity analysis of the complete cases were consistent with the primary analysis of the imputed dataset (S1 Appendix).
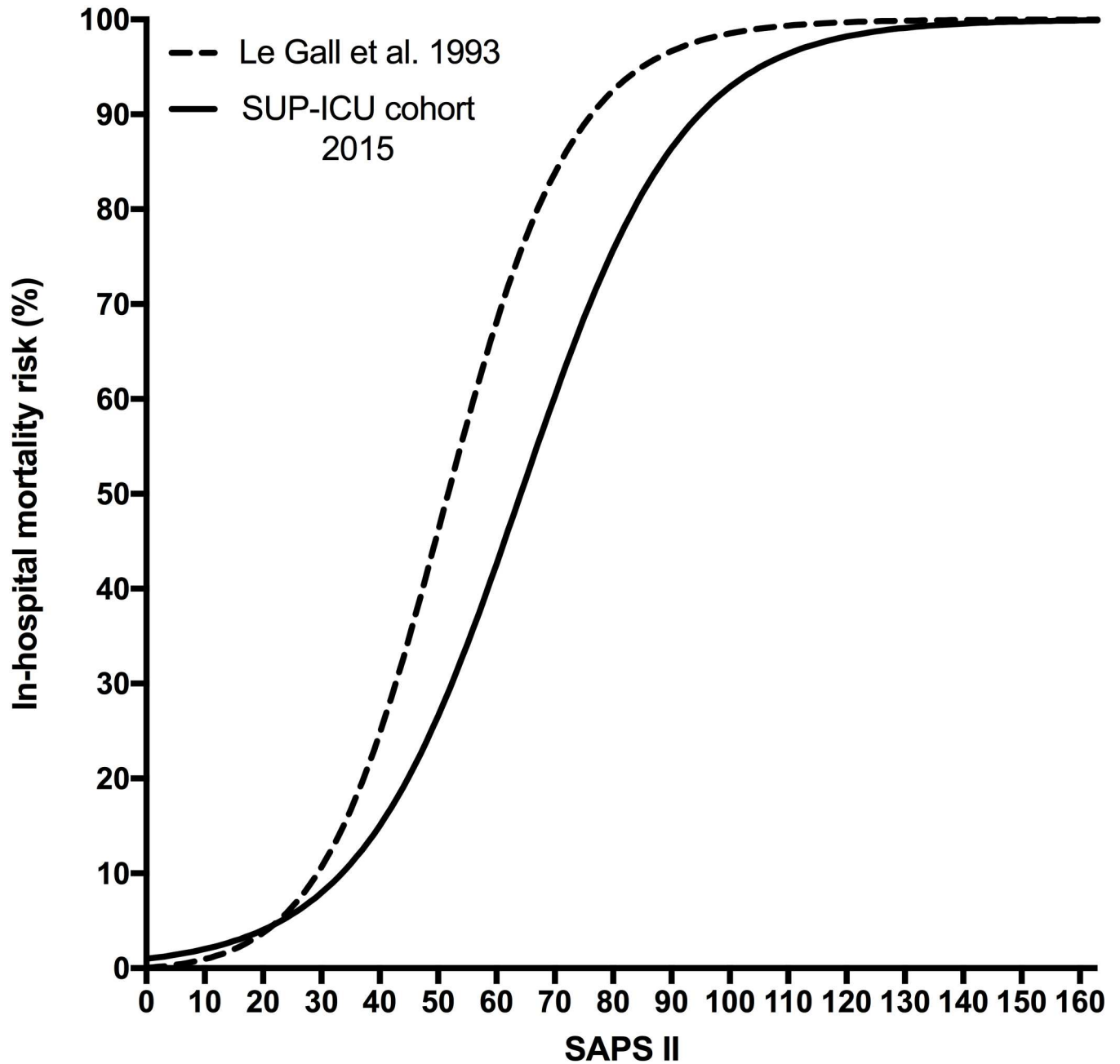
Fig 3. SAPS II and in-hospital mortality risk in the original and the present cohort. Correlation between SAPS II scores and in-hospital mortality risk in the original and the present cohort. In-hospital mortality risks are according to the original SAPS II equation (Le Gall et al. 1993) and the equation of SAPS II re-calibrated in the present cohort.

doi:10.1371/journal.pone.0168948.g003

Discussion

In this study of the performances of SAPS II and the initial SOFA score in a contemporary cohort of general ICU patients, we found that 1) the discrimination of SAPS II was superior to that of the initial SOFA score for in-hospital mortality, 2) the predictive performance of SAPS II was lower in our cohort than that in the original cohort, and 3) the discriminative powers of the two scores were comparable for in-hospital mortality and 90-day mortality.

Table 2. SAPS II and in-hospital mortality risk in the original and the present cohort.

SAPS II score	Le Gall et al. 1993	SUP-ICU cohort 2015
10	1.0%	2.0%
20	3.7%	4.0%
30	10.6%	7.9%
40	24.7%	15.0%
50	46.1%	26.6%
60	68.1%	42.6%
70	83.8%	60.3%
80	92.5%	75.7%
90	96.7%	86.4%
100	98.5%	92.9%

Correlation between SAPS II scores and in-hospital mortality risk in the original and the present cohort for selected SAPS II scores.

doi:10.1371/journal.pone.0168948.t002

SAPS II vs. the initial SOFA score

The predictive performance of the initial SOFA score has been assessed in a systematic review from 2008 by Minne et al. They summarised AUROCs of the initial SOFA score in 12 studies and concluded that the initial SOFA score was not inferior to SAPS II [16], although their review only included a single study that directly compared the two scores and this was in a small, selected patient population from a single ICU [30]. The discrimination of the initial SOFA score in our study was within the range reported by Minne et al., but inferior when compared with that of SAPS II. Importantly, the SOFA score was not developed to predict mortality, but to assess organ failure over time [5].

SAPS II in the present cohort vs. SAPS II in the original cohort

Studies revisiting the performance of SAPS II have generally found good discrimination but poor calibration of the original equation [6–15]. Due to this, expected mortality ratios calculated using the original SAPS II equation in contemporary ICU patients are imprecise and in general too high, which is supported by the different correlations between SAPS II and in-hospital mortality in the original study vs. the present study. The imprecise observed/expected mortality ratios are likely caused by advances in intensive care, as the in-hospital mortality rate for ICU patients has declined over time [31–34], despite increasing age and severity of illness [31, 34]. Case-mix is known to affect the predictive powers of severity scores [18], however, even after adjustment for risk factors and case-mix, in-hospital mortality for patients admitted to the ICU has declined over time [32–34]. Accordingly, SAPS II has been re-calibrated in multiple studies improving the predictive performance [6, 8, 10, 12, 13, 15].

Our results show that not only the calibration of SAPS II is affected. The discrimination of the score has decreased slightly when used in contemporary ICU patients compared to the original study, and this decrease was statistically significant. While first-level customisations are simple to conduct and generally lead to adequate calibration of SAPS II—as was seen in the present study—first-level customisations do not increase discrimination, as the individual parameters and weights of the score are unchanged.

The somewhat inadequate predictive performance of SAPS II in contemporary general ICU patients has intrigued refinement of the score by including additional variables [12], and in 2005 the SAPS 3 score was developed [35, 36]. When compared with SAPS II, SAPS 3 has not
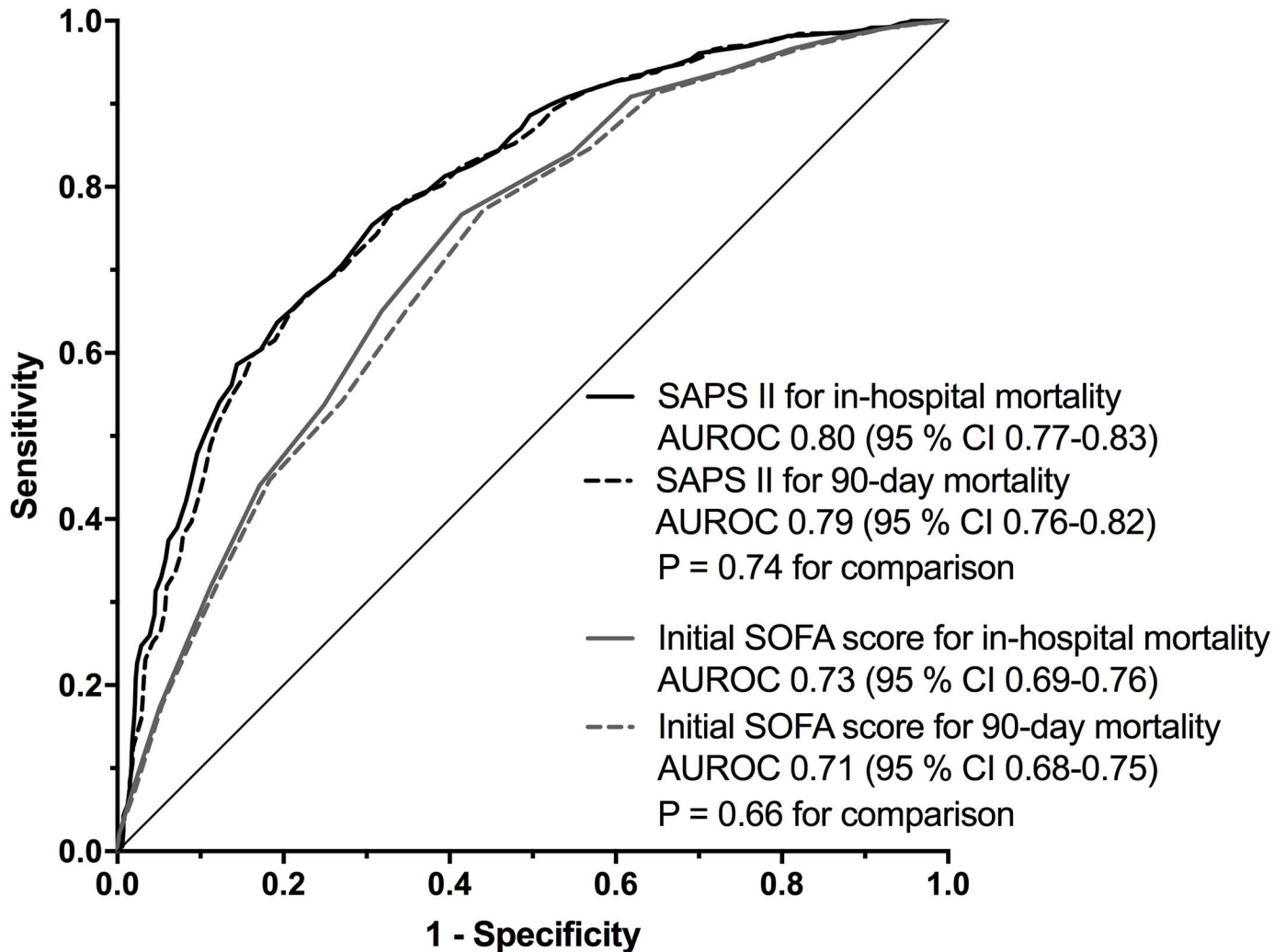
Fig 4. Discrimination of SAPS II and the initial SOFA score for in-hospital and 90-day mortality. Receiver operating characteristics curves for SAPS II and the initial SOFA score for in-hospital and 90-day mortality. AUROC: Area under the receiver operating characteristics curve.

doi:10.1371/journal.pone.0168948.g004

been shown to perform markedly better [37–41], which may be why SAPS II is still widely used, despite its limitations.

In-hospital vs. 90-day mortality

The AUROCs for SAPS II and the initial SOFA score for in-hospital vs. 90-day mortality were not different. Several studies have demonstrated that the performance of scores that rely solely on data collected during the first 24 hours of ICU stay (SAPS, SAPS II and the initial SOFA score) decrease with increasing lengths of stay [15, 42–44]. It has also been argued that the scores predict the obvious [43, 45]; patients presenting with the worst conditions will have the most deranged physiological parameters within the first 24 hours, causing high scores and high predicted as well as observed mortality rates. This may explain why 90-day mortality, though generally considered a more adequate outcome measure than in-hospital mortality [19–21], did not lead to improved predictive performance of SAPS II and initial SOFA scores in our study.

Limitations with the use of SAPS II and initial SOFA scores for mortality prediction today

As highlighted above and in previous studies, the ability of SAPS II and initial SOFA scores to predict mortality has limitations, as the SOFA score was not developed to predict mortality and the predictive performance of the SAPS II has deteriorated. This may limit their clinical use in contemporary ICU cohorts in research and in benchmarking using standardised mortality ratios, unless the scores are regularly re-calibrated. Other potential challenges with the use of SAPS II and the initial SOFA score exist. First, missing values for some variables often occur (e.g. bilirubin and urea levels) [46, 47], which may affect the sum of the scores, and hence result in inadequate prediction. Second, the quality of treatment in the ICU can affect the scores and mask the true baseline differences, due to the use of the worst recorded parameters over a relatively long period of 24 hours [48]. Third, if the scores are used as originally proposed (worst values recorded during first 24 hours in the ICU), there is a risk of intervention effect on the scores in randomised clinical trials, which also may affect the predictive performance. This is often avoided by using values recorded within the 24 hours prior to, instead of after, inclusion in the trial [46, 47]. Finally, both scores include the GCS score, which poses a challenge when assessing sedated and mechanically ventilated patients. Of note, with SAPS II the GCS before sedation should be used [4], while the actual or assumed GCS should be used in the SOFA score [5]; this may result in discrepancies between the two scores.

When severity scores are used to describe populations in ICU studies, we believe that using both SAPS II and SOFA scores and additionally reporting age, important comorbidities and vital signs is appropriate. Use of severity scores that include variables readily available upon ICU admission may also prove to be of value.

Strengths and limitations

The strengths of this study includes high generalisability (external validity), as contemporary ICU patients from 97 ICUs in 11 countries were included [22]. Second, data were prospectively registered in an electronic case report form, limiting the risk of information and selection bias, and third, the protocol and statistical analysis plan was prepared prior to analysis, reducing the risk of selection bias and data-driven analyses. Our study comes with limitations as well. First, patients suffering from GI bleeding upon admission to the ICU were excluded from the SUP-ICU cohort study, and thus the results from our study cannot be generalised to patients with GI bleeding admitted to the ICU. Second, SAPS II and initial SOFA scores were missing for 17.4% and 23.4% of the patients in the SUP-ICU database, respectively. We addressed this issue as recently recommended by using multiple imputations, as complete case analysis can bias the results [25]. The results of our sensitivity analysis of the complete case dataset were similar to the analysis of the imputed dataset. Third, this was a post-hoc study, which generally increases the risk for selection bias and spurious findings. Fourth, the sample used in this study was relatively small compared to the original development samples and several retrospective studies evaluating the scores. Finally, the participating ICUs may not be representative of ICUs in other countries.

Conclusion

In this post-hoc study we assessed the performances of SAPS II and the initial SOFA score in contemporary ICU patients. We found that discrimination of SAPS II was superior to that of the initial SOFA score for in-hospital mortality, the predictive performance of SAPS II has

decreased over time, and the discriminative powers of the two scores were not affected by extending the observation period from in-hospital mortality to 90-day mortality.

Consequently, the use of SAPS II and the initial SOFA score for mortality prediction purposes in contemporary ICU patients is not without limitations, and the development of a contemporary mortality prediction score may be of value.

Supporting Information

S1 Appendix. Sensitivity analysis using complete case analyses.
(DOCX)

S1 Dataset. Dataset SAS7BDAT.
(SAS7BDAT)

Acknowledgments

We thank the SUP-ICU co-authors for their work in the original SUP-ICU cohort study.

Author Contributions

Conceptualization: AG MHM MK AP PBH.

Data curation: AG MK PBH.

Formal analysis: AG MK PBH.

Investigation: MHM MK AP.

Methodology: AG MHM MK AP PBH.

Project administration: AG MHM PBH.

Software: AG MK PBH.

Supervision: AG MHM MK AP PBH.

Validation: MHM PBH.

Visualization: AG MHM MK AP PBH.

Writing – original draft: AG MHM MK PBH.

Writing – review & editing: AG MHM MK AP PBH.

References

1. Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: A compendium. Indian J Crit Care Med 2014; 18: 220–228. doi: 10.4103/0972-5229.130573 PMID: 24872651

2. Strand K, Flaatten H. Severity scoring in the ICU: a review. Acta Anaesthesiol Scand 2008; 52: 467–478. doi: 10.1111/j.1399-6576.2008.01586.x PMID: 18339152

3. Vincent J-L, Moreno R. Clinical review: scoring systems in the critically ill. Crit Care 2010; 14: 207. doi: 10.1186/cc8204 PMID: 20392287

4. Le Gall J-R, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993; 270: 2957–2963. [Erratum, JAMA 1994;271: 1321]. PMID: 8254858

5. Vincent J-L, 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. Intensive Care Med 1996; 22: 707–710. PMID: 8844239

6. Apolone G, Bertolini G, D'Amico R, Iapichino G, Cattaneo A, De Salvo G, et al. The performance of SAPS II in a cohort of patients admitted to 99 Italian ICUs: results from GiViTI. Intensive Care Med 1996; 22: 1368–1378. PMID: 8986488

7. Moreno R, Morais P. Outcome prediction in intensive care: results of a prospective, multicentre, Portuguese study. Intensive Care Med 1997; 23: 177–186. PMID: 9069003

8. Beck DH, Smith GB, Pappachan JV. The effects of two methods for customising the original SAPS II model for intensive care patients from South England. Anaesthesia 2002; 57: 785–793. PMID: 12133092

9. Beck DH, Smith GB, Pappachan J V, Millar B. External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. Intensive Care Med 2003; 29: 249–256. doi: 10.1007/s00134-002-1607-9 PMID: 12536271

10. Metnitz PG, Valentin A, Vesely H, Alberti C, Lang T, Lenz K, et al. Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Intensive Care Med 1999; 25: 192–197. PMID: 10193547

11. Glance LG, Osler TM, Dick A. Rating the quality of intensive care units: is it a function of the intensive care unit scoring system? Crit Care Med 2002; 30: 1976–1982. PMID: 12352029

12. Le Gall JR, Neumann A, Hemery F, Bleriot JP, Fulgencio JP, Garrigues B, et al. Mortality prediction using SAPS II: an update for French intensive care units. Crit Care 2005; 9: R645–652. doi: 10.1186/cc3821 PMID: 16280063

13. Harrison DA, Brady AR, Parry GJ, Carpenter JR, Rowan K. Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. Crit Care Med 2006; 34: 1378–1388. doi: 10.1097/01.CCM.0000216702.94014.75 PMID: 16557153

14. Desa K, Peric M, Husedzinovic I, Sustic A, Korusic A, Karadza V, et al. Prognostic performance of the Simplified Acute Physiology Score II in major Croatian hospitals: a prospective multicenter study. Croat Med J 2012; 53: 442–449. doi: 10.3325/cmj.2012.53.442 PMID: 23100206

15. Haaland ØA, Lindemark F, Flaatten H, Kvåle R, Johansson KA. A calibration study of SAPS II with Norwegian intensive care registry data. Acta Anaesthesiol Scand 2014; 58: 701–708. doi: 10.1111/aas.12327 PMID: 24819749

16. Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. Crit Care 2008; 12: R161. doi: 10.1186/cc7160 PMID: 19091120

17. Minne L, Eslami S, de Keizer N, de Jonge E, de Rooij SE, Abu-Hanna A. Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment. Intensive Care Med 2012; 38: 40–46. doi: 10.1007/s00134-011-2390-2 PMID: 22042520

18. Metnitz PGH, Lang T, Vesely H, Valentin A, Le Gall JR. Ratios of observed to expected mortality are affected by differences in case mix and quality of care. Intensive Care Med 2000; 26: 1466–1472. PMID: 11126258

19. Rydenfelt K, Engerström L, Walther S, Sjöberg F, Strömberg U, Samuelsson C. In-hospital vs. 30-day mortality in the critically ill—a 2-year Swedish intensive care cohort analysis. Acta Anaesthesiol Scand 2015; 59: 846–858. doi: 10.1111/aas.12554 PMID: 26041018

20. Brinkman S, Abu-Hanna A, de Jonge E, de Keizer NF. Prediction of long-term mortality in ICU patients: model validation and assessing the effect of using in-hospital versus long-term mortality on benchmarking. Intensive Care Med 2013; 39: 1925–1931. doi: 10.1007/s00134-013-3042-5 PMID: 23921978

21. Jammer I, Wickboldt N, Sander M, Smith A, Schultz MJ, Pelosi P, et al. Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European Perioperative Clinical Outcome (EPCO) definitions: A statement from the ESA-ESICM joint taskforce on perioperative outcome measur. Eur J Anaesthesiol 2015; 32: 88–105. doi: 10.1097/EJA.0000000000000118 PMID: 25058504

22. Krag M, Perner A, Wetterslev J, Wise MP, Borthwick M, Bendel S, et al. Prevalence and outcome of gastrointestinal bleeding and use of acid suppressants in acutely ill adult intensive care patients. Intensive Care Med 2015; 41: 833–845. doi: 10.1007/s00134-015-3725-1 PMID: 25860444

23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. J Clin Epidemiol 2008; 61: 344–349. doi: 10.1016/j.jclinepi.2007.11.008 PMID: 18313558

24. Alberti C, Boulkedid R. Describing ICU data with tables. Intensive Care Med 2014; 40: 667–673. doi: 10.1007/s00134-014-3248-1 PMID: 24615560

25. Vesin A, Azoulay E, Ruckly S, Vignoud L, Rusinovà K, Benoit D, et al. Reporting and handling missing values in clinical studies in intensive care units. Intensive Care Med 2013; 39: 1396–1404. doi: 10.1007/s00134-013-2949-1 PMID: 23685609

26. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol 2009; 9: 57. doi: 10.1186/1471-2288-9-57 PMID: 19638200

27. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics 1988; 44: 837–845. PMID: 3203132

28. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. http://www.cochrane-handbook.org.

29. Gönen M. Analyzing Receiver Operating Characteristic Curves with SAS®. 1st ed. Cary, NC, USA: SAS Institute, Inc.; 2007.

30. Janssens U, Graf C, Graf J, Radke PW, Königs B, Koch KCh, et al. Evaluation of the SOFA score: a single-center experience of a medical intensive care unit in 303 consecutive patients with predominantly cardiovascular disorders. Intensive Care Med 2000; 26: 1037–1045. PMID: 11030159

31. Zimmerman JE, Kramer AA, Knaus WA. Changes in hospital mortality for United States intensive care unit admissions from 1988 to 2012. Crit Care 2013; 17: R81. doi: 10.1186/cc12695 PMID: 23622086

32. Moran JL, Bristow P, Solomon PJ, George C, Hart GK. Mortality and length-of-stay outcomes, 1993–2003, in the binational Australian and New Zealand intensive care adult patient database. Crit Care Med 2008; 36: 46–61. doi: 10.1097/01.CCM.0000295313.08084.58 PMID: 18090383

33. Hutchings A, Durand MA, Grieve R, Harrison D, Rowan K, Green J, et al. Evaluation of modernisation of adult critical care services in England: time series and cost effectiveness analysis. BMJ 2009; 339: b4353. doi: 10.1136/bmj.b4353 PMID: 19906740

34. Williams TA, Ho KM, Dobb GJ, Finn JC, Knuiman MW, Webb SAR. Changes in case-mix and outcomes of critically ill patients in an Australian tertiary intensive care unit. Anaesth Intensive Care 2010; 38: 703–709. PMID: 20715735

35. Metnitz PGH, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. Intensive Care Med 2005; 31: 1336–1344. doi: 10.1007/s00134-005-2762-6 PMID: 16132893

36. Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Campos RA, 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. doi: 10.1007/s00134-005-2763-5 PMID: 16132892

37. Poole D, Rossi C, Latronico N, Rossi G, Finazzi S, Bertolini G. Comparison between SAPS II and SAPS 3 in predicting hospital mortality in a cohort of 103 Italian ICUs. Is new always better? Intensive Care Med 2012; 38: 1280–1288. doi: 10.1007/s00134-012-2578-0 PMID: 22584793

38. Sakr Y, Krauss C, Amaral ACKB, Réa-Neto A, Specht M, Reinhart K, et al. Comparison of the performance of SAPS II, SAPS 3, APACHE II, and their customized prognostic models in a surgical intensive care unit. Br J Anaesth 2008; 101: 798–803. doi: 10.1093/bja/aen291 PMID: 18845649

39. Ledoux D, Canivet J-L, Preiser J-C, Lefrancq J, Damas P. SAPS 3 admission score: an external validation in a general intensive care population. Intensive Care Med 2008; 34: 1873–1877. doi: 10.1007/s00134-008-1187-4 PMID: 18592214

40. Strand K, Søreide E, Aardal S, Flaatten H. A comparison of SAPS II and SAPS 3 in a Norwegian intensive care unit population. Acta Anaesthesiol Scand 2009; 53: 595–600. doi: 10.1111/j.1399-6576.2009.01948.x PMID: 19419352

41. Capuzzo M, Scaramuzza A, Vaccarini B, Gilli G, Zannoli S, Farabegoli L, et al. Validation of SAPS 3 Admission Score and comparison with SAPS II. Acta Anaesthesiol Scand 2009; 53: 589–594. doi: 10.1111/j.1399-6576.2009.01929.x PMID: 19419351

42. Sicignano A, Carozzi C, Giudici D, Merli G, Arlati S, Pulici M. The Influence of length of stay in the ICU on power of discrimination of a multipurpose severity score (SAPS). Intensive Care Med 1996; 22: 1048–1051. PMID: 8923068

43. Suistomaa M, Niskanen M, Kari A, Hynynen M, Takala J. Customised prediction models based on APACHE II and SAPS II scores in patients with prolonged length of stay in the ICU. Intensive Care Med 2002; 28: 479–485. doi: 10.1007/s00134-002-1214-9 PMID: 11967604

44. Nfor TK, Walsh TS, Prescott RJ. The impact of organ failures and their relationship with outcome in intensive care: analysis of a prospective multicentre database of adult admissions. Anaesthesia 2006; 61: 731–738. doi: 10.1111/j.1365-2044.2006.04707.x PMID: 16867083

45. Vincent JL, de Mendonça A, Cantraine F, Moreno R, Takala J, Suter PM, et al. Use of the SOFA score to asses the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. Crit Care Med 1998; 26: 1793–1800. PMID: 9824069

46.   Perner A, Haase N, Guttormsen AB, Tenhunen J, Klemenzson G, Åneman A, et al. Hydroxyethyl Starch 130/0.42 versus Ringer's Acetate in Severe Sepsis. N Engl J Med 2012; 367: 124–134. doi: 10.1056/NEJMoa1204242 PMID: 22738085

47.   Holst LB, Haase N, Wetterslev J, Wernerman J, Guttormsen A, Karlsson S, et al. Lower versus Higher Hemoglobin Threshold for Transfusion in Septic Shock. N Engl J Med 2014; 371: 1381–1391. doi: 10.1056/NEJMoa1406617 PMID: 25270275

48.   Vincent J-L, Opal SM, Marshall JC. Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions. Crit Care Med 2010; 38: 283–287. doi: 10.1097/CCM.0b013e3181b785a2 PMID: 19730252