# Validation of APACHE II and SAPS II scales at the intensive care unit along with assessment of SOFA scale at the admission as an isolated risk of death predictor

Izabela Kądziołka[1,2], Rafał Świstek[1], Karolina Borowska[3], Paweł Tyszecki[1], Wojciech Serednicki[1,4]

[1]Department of Anaesthesiology and Intensive Therapy, University Hospital in Krakow, Poland
[2]Department of Medical Didactics, Collegium Medicum, Jagiellonian University in Krakow, Poland
[3]University Hospital in Krakow, Poland
[4]Department of Interdisciplinary Intensive Therapy, Collegium Medicum, Jagiellonian University in Krakow, Poland

## Abstract

**Background:** Disease's severity classification systems are applied to measure the risk of death and to choose the best therapy for patients admitted to intensive care unit (ICU). The aim of the study was to verify risk of death calculated with APACHE II (Acute Physiology and Chronic Health Evaluation II), SAPS II (Simplified Acute Physiology Score II), SOFA (Sequential Organ Failure Assessment) and evaluate correlation between these scores. The usefulness of SOFA score as a sole scale also was assessed.

**Methods:** This was a retrospective study conducted in 30-beds ICU in Kraków, Poland. Every male and female patient over 18 years old who was admitted to the ICU between 18.04.2016 and 12.08.2016 was included in the analysis. Patients who were transferred from another ICU were excluded from the research. APACHE II, SAPS II, SOFA were calculated after admission using laboratory results and clinical examination. Discrimination and calibration were used to validate these scoring system.

**Results:** Analysis included 86 patients. The outcomes were compared within survivors and non-survivors groups. The prediction of death was statistically significant only for APACHE II and SAPS II. The best AUROC was for APACHE II 0.737 and SAPS II 0.737; discrimination for SOFA was not statistically significant. There was high correlation only between SAPS II and APACHE II results ($r \geq 0.7$, $P < 0.01$). The calibration was excellent for SAPS II, $P = 0.991$, and slightly worse for APACHE II, $P = 0.685$, and SOFA, $P = 0.540$. Patients who survived spent more days on ICU ($P < 0.01$), mean Length of Stay (LOS) in this group was $38.25 \pm 16.80$ days.

**Conclusions:** APACHE II and SAPS II scales have better discrimination, calibration and power to predict deaths on ICU than SOFA. Among these scales SOFA did not achieve expected results.

**Key words:** APACHE II, SAPS II, SOFA, risk of death on admission, intensive care unit.

**CORRESPONDING AUTHOR:**

Wojciech Serednicki, Department of Anaesthesiology and Intensive Therapy, University Hospital in Krakow, 50 Kopernika St., 30-501 Kraków, Poland, e-mail: w.serednicki@hotmail.com

The first scales evaluating the patient conditions were designed in the 80-ties of the previous century and quickly became popular at intensive care units (ICUs) [1, 2]. Since that time, numerous new scales have been created or the existing ones modified in order to provide most accurate assessment of patients. They are the tools for validation of the therapeutic procedures used as well as for monitoring the quality and costs of treatment. The standardised assessment measures are also extremely useful during prospective and retrospective studies, which, in turn, contribute to implementation of increasingly effective treatment options, both interventional and pharmacological [3].

The best known scales include the Acute Physiology and Chronic Health Evaluation II (APACHE II) and Simplified Acute Physiology Score II (SAPS II), which assess various physiological parameters and consider the data regarding age, chronic diseases and history of surgeries. The scores of the above scales should be verified within the first 24 h following ICU admission, which enables us to assess the patient's condition as well as to estimate the risk of death during a particular hospitalisation [1]. Some parameters are common for both scales yet in some cases they were replaced with other factors determining the performance of the same systems. The largest differences concern concomitant chronic diseases [1, 2].

The next scale that gains in popularity at ICUs is the Sequential Organ Failure Assessment (SOFA), originally termed Sepsis-related Organ Failure Assessment, which constitutes the basis for the current definition of sepsis. SOFA is used for everyday assessment of the patient's condition and is based on the parameters routinely monitored in the ICU settings; it includes parameters regarding the main body systems yet does not consider chronic diseases, age or modes of admissions. This scale is an alternative to the extensive classifications, such as APACHE II or SAPS II, which include not only the basic vital parameters and point-of care testing (POCT) but also laboratory tests performed every several days or only once [4].

Depending on the ICU profile and reference degree, the effectiveness of APACHE II, SAPS II and SOFA in estimating the prognosis markedly differs.

The present study was undertaken to evaluate the reliability of APACHE II and SAPS II in assessing the risk of ICU death. Another objective was to determine whether the introduction of SOFA or replacement of the scales hitherto used with it could improve the effectiveness of death risk evaluation and treatment efficacy.

## METHODS

The study design was approved by the Bioethics Committee of the Jagiellonian University in Krakow. The study encompassed patients treated in the 30-bed ICU of the University Hospital in Krakow. The retrospective analysis included patients > 18 years of age admitted to the ICU between April 18, 2016 and August 12, 2016. The patients transferred from another ICU were excluded.

The study patients were assessed on admission with APACHE II and SAPS II using the calculators configured in the Infomedica system. The SOFA scores were calculated for the worst results within 24 hours after admission (https://www.the-calculator.co/health/Sequential-Organ-Failure-Assessment-(SOFA)-Score-Calculator-862.html). The data entered to the formula were taken from the POCT analyser (concentration of bilirubin, creati-

nine, $HCO_3$, blood pH, and $PaO_2$), laboratory blood tests (PLT and WBC count, haematocrit), and neurological assessment according to the Glasgow Coma Scale (GCS). GCS was used on admission. The study population was divided into two groups: group 1 – patients who died in the ICU (non-survivors) and group 2 – whose final point (outcome) was discharge from ICU or transfer to the conservative department (survivors).

The APACHE II, SAPS II and SOFA results were statistically analysed as for discrimination, prediction of death and calibration of tests. Discrimination was defined as the ability of a scale to differentiate between non-survivors and survivors. The satisfying discrimination result is assumed to be the area under a receiver operator characteristic (AUROC) > 0.70 [5]. Calibration was defined as the degree of agreement between the expected and the observed final point.

Statistical analyses were performed using source data that consisted of qualitative and quantitative variables. Quantitative variables of normal distribution were compared using the Mann-Whitney $U$ test. Inter-scale correlations were checked with the Pearson correlation. Discrimination was checked using ROC and AUROC curves. The Hosmer-Lemenshow test was applied for calibration. $P < 0.01$ was considered statistically significant. The STATISTICA 12 (Statsoft, Tulsa, USA) software was applied for the analysis of data.

## RESULTS

The study included 86 individuals; the mean age was 58.66 ± 20.71 years while the mean ICU length of stay (LOS) – 32.06 ± 18.86 days.

Group 1 consisted of 25 patients (29%). Their ICU treatment was shorter compared to group 2 (U = 253, $P < 0.01$). Demographic characteristics of patients are presented in Table 1.

The means and medians of scores in various scales were calculated for both groups (Table 2). In group 2, the APACHE II and SAPS II scores were lower than in group 1 – 3.90 and 3.96, respectively; $P < 0.01$. There was no statistically significant intergroup dif-

**TABLE 1.** Demographic data of study patients

| Parameter | Total | Group of non-survivors | Group of survivors | P |
|---|---|---|---|---|
| Number of individuals (%) | 86 (100) | 25 (29.1) | 61 (70.9) | |
| Gender | | | | |
| Female (%) | 29 (33.7) | 10 (40.0) | 19 (31.1) | 0.43 |
| Male (%) | 57 (66.3) | 15 (60.0) | 42 (68.9) | |
| Mean age (SD, years) | 58.6 (20.71) | 65.4 (20.46) | 55.9 (20.34) | 0.053 |
| Mean ICU stay (SD, days) | 32.06 (18.86) | 16.96 (14.8) | 38.25 (16.8) | < 0.01 |

ICU — intensive care unit

**TABLE 2.** APACHE II, SAPS II and SOFA scores of study patients presented as means ± SD (medians)

| Parameter | Total | Group of non-survivors | Group of survivors | P |
|---|---|---|---|---|
| APACHE II | 24.15 ± 7.18 (25.0) | 28.52 ± 6.48 (28.0) | 22.36 ± 6.71 (23.0) | < 0.01 |
| SAPS II | 58.61 ± 14.52 (58.0) | 67.52 ± 13.85 (66.0) | 54.9 ± 13.21 (54.5) | < 0.01 |
| SOFA | 14.36 ± 3.73 (14.0) | 15.2 ± 3.83 (14.0) | 14.0 ± 3.66 (14.0) | 0.18 |

APACHE II — Acute Physiology and Chronic Health Evaluation II, SAPS — Simplified Acute Physiology Score II, SOFA — Sequential Organ Failure Assessment

ference in SOFA scores on admission – group 1: 15.2 (± 3.83), group 2: 14.0 (± 3.66), $P = 0.18$.

Based on the Pearson correlation coefficient, a strong correlation between SAPS II and APACHE II scores was found ($r = 0.7$, $P < 0.01$); otherwise, there was no correlation observed between SOFA versus SAPS II ($P = 0.056$) and APACHE II ($P = 0.365$) scores.

The ability to discriminate both study groups as to the risk of death using APACHE II, SAPS II and SOFA scales was presented in Figure 1. AUROC was found to be 0.738 – APACHE II, 0.737 – SAPS II and 0.579 – SOFA; only the first two scales showed statistical significance in predicting death (very similar results).

The best calibration determined using the Hosmer-Lemenshow test was found for SAPS II (1.601; $P = 0.991$). The APACHE II and SOFA results were worse (4.798; $P = 0.685$; 5.033; $P = 0.540$, respectively).

## DISCUSSION

In the present study, the APACHE II and SAPS II scales were found to be effective in reliable evaluation of the risk of death on ICU admission. Their scores were higher in non-survivors, as compared to survivors (discharged or transferred from the ICU). No such a difference was observed in the case of SOFA. Additionally, APACHE II and SAPS II discrimination results were better compared to SOFA.

Numerous papers have been published, compring the effectiveness of the analysed scales in predicting mortality; however, the results of SOFA are ambiguous. The uselessness of this scale for predicting mortality on ICU admission has been reported in one prospective study carried out in Canada [6]. In the above study, the data were collected according to the recommendations described by the authors of SOFA [4]. The results of 1436 patients have demonstrated poor calibration and discrimination (AUROC 0.67) [6]. On the other hand, some reports have confirmed the usefulness of SOFA for predicting the risk of death on ICU admission (AUROC > 0.8) [7–9]. The retrospective study from New Delhi, performed to compare the new generation APACHE, SAPS, and SOFA scales with those used earlier, has demonstrated high discrimination (AUROC ~ 0.89) and poor calibration of SOFA and SAPS II scales [9].

In the above mentioned and some other studies comparing the usefulness of scales, the values for SOFA ranged from 0.67 to 0.84 [6–8, 10–13].
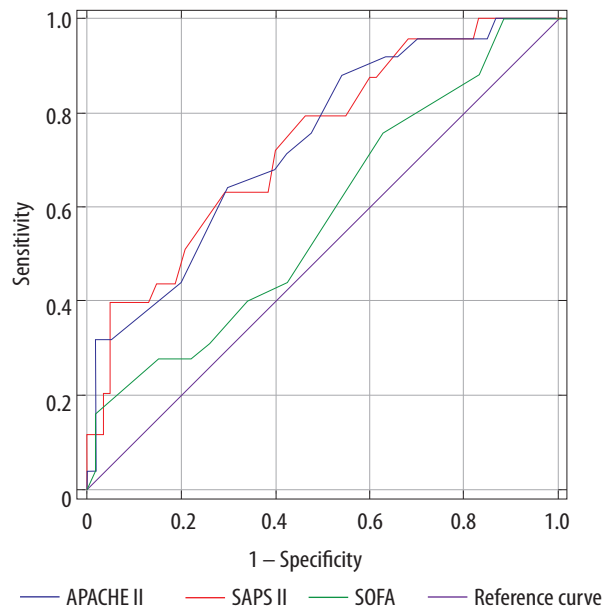


**FIGURE 1.** The ROC curves for APACHE II, SAPS II and SOFA

APACHE II - Acute Physiology And Chronic Health Evaluation II, SAPS II -Simplified Acute Physiology Score II, SOFA — Sequential Organ Failure Assessment Score, AUROC — The Area Under the Receiver Operating Characteristic

The worst AUROC result for APACHE II or SAPS II scales was 0.74 [14]. According to the studies comparing only SOFA with APACHE II or SAPS II, the AUROC result for SOFA was 0.70 while the APACHE II and /or SAPS II discrimination results II were more reliable [10, 11, 13, 15].

Some differences described above are likely to result from different approaches to the Glasgow Coma Scale (GCS) scores in ICU patients [2, 4, 16]. In cases of SAPS II and APACHE II scales, GCS should be evaluated before administering sedatives and implementing mechanical lung ventilation [2, 16]. For the purposes of SOFA, the current score should be taken into account [4]. In our study, the evaluation of GCS was hindered – on admission to ICU, patients

**TABLE 3.** Comparison of discrimination and calibration of APACHE II, SAPS II and SOFA

| Parameter | AUROC | Hosmer-Lemenshow test, P |
|---|---|---|
| APACHE II | 0.738 | 0.685 |
| SAPS II | 0.737 | 0.991 |
| SOFA | 0.579 | 0.540 |

AUROC — the Area under a Receiver Operating Characteristic, APACHE II - Acute Physiology And Chronic Health Evaluation II, SAPS II -Simplified Acute Physiology Score II, SOFA — Sequential Organ Failure Assessment

were usually sedated at the Emergency Department or during anaesthesia. Moreover, the evaluation of GCS in intubated patients remains debatable, which is essential for SOFA, as this parameter plays an important role in it.

As far as calibration is concerned, our findings demonstrated the best results for SAPS II ($P = 0.991$), slightly worse for APACHE II ($P = 0.685$) and the worst ones for SOFA ($P = 0.540$). The above findings are consistent with the results reported by other centres showing better calibrations for the first two scales, i.e. 0.079–0.811 [9, 10, 13, 14] and 0.189–0.86 [9, 11, 13, 14], respectively, as compared to SOFA: 0.301–0.437 [9, 11, 13, 15].

The reports validating the scales in Polish ICUs are scarce. The only data reported in literature come from the Gdansk and Łodz centres where SAPS II, APACHE II and SOFA scales were validated for predicting mortality. Sawicka *et al.* [17] evaluated patients with diseases of the haemopoietic system. The Łodz centre validated the scales for patients with meningococcal cerebrospinal meningitis. According to both studies, all three scales have been found useful. However, only the SAPS II scale was as an independent factor of the risk of death [17, 18]. In our study, the mean results of all three scales on ICU admission were comparable.

The comparison of the mean ICU length of stay with the data of other studies [11, 14, 19] clearly showed a significantly longer stay in the Krakow ICU, which was particularly evident in group 2 – 38.25 days (± 16.8); in group 1, the mean length of stay was 16.96 days (± 14.8), which is also long. In the studies cited, the mean ICU treatment time in patients discharged or transferred from the ICU ranged from 4.0 to 8.42 days and in the group of non-survivors – 5.3–7.51 days [11, 14, 19].

The issue of the length of ICU stay under the conditions of the Polish health care system has been disputed multiple times. In 2014, Adamski *et al.* [20] presented the data comparing the functioning of ICUs in Poland (Olsztyn) and Finland (Pori) and demonstrated radical differences in the mean length of ICU stay (14 and 4.2 days, respectively). According to the authors, the causes of such differences include the lack of legal regulations about the principles regarding maintenance/withdrawal of treatment and procedures ("do not resuscitate"), the lack of planned care of chronically ill patients after discharge from ICUs and lack of intermediate facilities in the Polish health care system dealing with further treatment of patients discharged from ICUs and before admission to the conservative departments. Moreover, in Poland, almost 50% of patients were admitted to ICUs from other hospital departments while in Finland, admissions from the emergency departments predominate [20]. Such results point to the problem of too late response to the gradually deteriorating conditions of patients treated in hospital departments.

In our study, many factors, often dependent on the Polish health care conditions, affected high scores in the scales used, very long ICU hospitalisations, as well as treatment outcomes. They include a wide profile of patients admitted to ICUs, lack of management standards, lack of an appropriate system of transferring patients who do not require further treatment from ICU to other departments or facilities, and limited availability of diagnostic and therapeutic methods. Our analysis demonstrated that despite the lack of efficiency of SOFA in predicting death on ICU admission, its usefulness for continuous assessment of ICU patients was not negated. Everyday assessment of patients according to SOFA and analysis of changes in the results observed in the successive days could be an additional element of making therapeutic decisions [21].

The present study was performed as initial research for further analysis, verifying the usefulness of using scales on admission to the ICU in Krakow. The objective was to improve individual assessment of patients on ICU admission by using one most effective scale and take therapeutic decisions based on objective clinical data. Further research should include larger study populations of patients, divided according to the cause of hospitalisation and use newer assessment scales, e.g. SAPS III, which has showed high discrimination and calibration in numerous studies.

## CONCLUSIONS

In our centre, APACHE II and SAPS II scales hitherto used are the reliable tools estimating the risk of death on ICU admission (to the Department of Intensive Therapy in Krakow). Given good discriminations and high calibrations of both tests, one should consider whether the use of both similar scales is required.

The introduction of SOFA on ICU admission would not improve the efficacy of predicting death.

Considering the specificity of functioning of the Polish health care system and long hospitalisations, it seems necessary to perform additional assessments of patients according to scales during ICU treatment and analyse the results according to the outcomes.

## REFERENCES

1. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985; 13: 818-829.

2. Le Gall JR, 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. doi: 10.1001/jama.1993.03510240069035.

3. Weingarten S, Bolus R, Riedinger MS, Maldonado L, Stein S, Ellrodt AG. The principle of parsimony: Glasgow Coma Scale Score predicts mortality as well as the APACHE II score for stroke patients. Stroke 1990; 21: 1280-1282. doi: https://doi.org/10.1161/01.STR.21.9.1280.

4. 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.

5. Lemeshow S, Le Gall JR. Modeling the severity of illness of ICU patients. A systems update. JAMA 1994; 272: 1049-1055.

6. Zygun DA, Laupland KB, Fick GH, Sandham JD, Doig CJ. Limited ability of SOFA and MOD scores to discriminate outcome: a prospective evaluation in 1,436 patients. Can J Anesth 2005; 52: 302-308.

7. Kajdacsy-Balla Amaral AC, Andrade FM, Moreno R, Artigas A, Cantraine F, Vincent JL. Use of the sequential organ failure assessment score as a severity score. Intensive Care Med 2005; 31: 243-249. doi: 10.1007/s00134-004-2528-6.

8. Cabré L, Mancebo J, Solsona JF, et al. Bioethics Working Group of the SEMICYUC. Multicenter study of the multiple organ dysfunction syndrome in intensive care units: the usefulness of Sequential Organ Failure Assessment scores in decision making. Intensive Care Med 2005; 31: 927-933. doi: 10.1007/s00134-005-2640-2.

9. Juneja D, Singh O, Nasa P, Dang R. Comparison of newer scoring systems with the conventional scoring systems in general intensive care population. Minerva Anestesiol 2012; 78: 194-200.

10. Granholm A, Møller MH, Krag M, Perner A, Hjortrup PB. 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. PLoS One 2016; 11: e0168948, doi: 10.1371/journal.pone.0168948.

11. Ho KM, Lee KY, Williams T, Finn J, Knuiman M, Webb SA. Comparison of Acute Physiology and Chronic Health Evaluation (APACHE) II score with organ failure scores to predict hospital mortality. Anaesthesia 2007; 62: 466-473. doi: 10.1111/j.1365-2044.2007.04999.x.

12. Timsit JF, Fosse JP, Troché G, et al. OUTCOMEREA Study Group, France. Calibration and discrimination by daily Logistic Organ Dysfunction scoring comparatively with daily Sequential Organ Failure Assessment scoring for predicting hospital mortality in critically ill patients. Crit Care Med 2002; 30: 2003-2013. doi: 10.1097/00003246-200209000-00009.

13. Naqvi IH, Mahmood K, Ziaullaha S, Kashif SM, Sharif A. Better prognostic marker in ICU – APACHE II, SOFA or SAP II. Pak J Med Sci 2016; 32: 1146-1151. doi: 10.12669/pjms.325.10080.

14. Faruq MO, Mahmud MR, Begum T, et al. A comparison of severity systems APACHE II and SAPS II in critically ill patients. Bangladesh Crit Care 2013; 1: 27-32. doi: 10.3329/bccj.v1i1.14362.

15. Pettilä V, Pettilä M, Sarna S, Voutilainen P, Takkunen O. Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. Crit Care Med 2002; 30: 1705-1711. doi: 10.1097/00003246-200208000-00005.

16. Livingston BM, Mackenzie SJ, MacKirdy FN, Howie JC. Should the pre-sedation Glasgow Coma Scale value be used when calculating Acute Physiology and Chronic Health Evaluation scores for sedated patients? Scottish Intensive Care Society Audit Group. Crit Care Med 2000; 28: 389-394. doi: 10.1097/00003246-200002000-00017.

17. Sawicka W, Owczuk R, Wujtewicz MA, Wujtewicz M. The effectiveness of the APACHE II, SAPS II and SOFA prognostic scoring systems in patients with haematological malignancies in the intensive care unit. Anaesthesiol Intensive Ther 2014; 46: 166-170. doi: 10.5603/AIT.2014.0030.

18. Pietraszek-Grzywaczewska I, Bernas S, Łojko P, Piechota A, Piechota M. Predictive value of the APACHE II, SAPS II, SOFA and GCS scoring systems in patients with severe purulent bacterial meningitis. Anaesthesiol Intensive Ther 2016; 48: 175-179. doi: 10.5603/AIT.a2016.0030.

19. Sekulic AD, Trpkovic SV, Pavlovic AP, Marinkovic OM, Ilic AN. Scoring systems in assessing survival of critically ill ICU patients. Med Sci Monit 2015; 21: 2621-2629. doi: 10.12659/MSM.894153.

20. Adamski J, Goraj R, Onichimowski D, Gawlikowska E, Weigl W. The differences between two selected intensive care units located in central and northern Europe – preliminary observation. Anaesthesiol Intensive Ther 2015; 47: 117-124. doi: 10.5603/AIT.a2015.0010.

21. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016; 315: 762-774. doi: 10.1001/jama.2016.0288.