Predictive accuracy of the APACHE IV scores on mortality and prolonged stay in the intensive care unit of Dr Sardjito Hospital

Yunita Widyastuti, PhD, Wildan Arsyad Zaki, MD, Untung Widodo, PhD, Akhmad Yun Jufan, MD, Bhirowo Yudo Pratomo, MD

Department of Anesthesiology and Intensive Therapy Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Indonesia

ABSTRACT

Introduction: Acute Physiology and Chronic Health Evaluation (APACHE) is the most widely used scoring system in the intensive care unit (ICU). The APACHE IV showed a good level of discrimination and calibration on predicting mortality and prolonged stay (PLOS) in some countries. This study is aimed to determine the predictive accuracy of the APACHE IV score on mortality and PLOS at the ICU of Dr Sardjito General Hospital (SGH).

Materials and Methods: This study involved all adult patients at the ICU of SGH during 2018 that met the inclusion criteria. The discrimination of APACHE IV scores on mortality and PLOS was analyzed with Receiver Operating Characteristic Curve, and the optimal cut-off point was assessed with the Youden Index. The calibration of the APACHE IV score was assessed with the Hosmer-Lemeshow goodness-of-fit test, and a *p*-value of >0.05 is considered a good calibration.

Results: From the data of 742 patients, only 476 were included. The overall mortality and PLOS rate was 25.4 % and 15.1 %, respectively. The mean of APACHE IV score was 66.27±27.7. The area under the receiving curve with a 95% confidence interval for mortality is 0.99(0.97–1.00) and for PLOS was 0.68(0.62–0.74). The optimal cut-off point of the APACHE IV score for mortality was 78.9, with a sensitivity of 0.96 and a specificity of 0.96. The optimal cut-off point of the APACHE IV score for PLOS is 62.5 (in the 6th percentiles), with a sensitivity of 0.72 and a specificity of 0.61. The calibration is good for mortality prediction (*p*=0.98) but is poor for PLOS prediction (*p*=0.01).

Conclusion: APACHE IV score has excellent accuracy for mortality prediction but is poor for PLOS prediction in patients in the ICU of SGH.

KEYWORDS:
APACHE IV score, ICU, Mortality, PLOS, Prediction

INTRODUCTION

The scoring system to predict the outcome of critically ill patients is needed for consideration of decision making, resources allocation, benchmarking, and stratification for clinical trials.[1] The accuracy of the risk prediction models are measured by their good calibration, discrimination, and generalizability (good reproducibility of the score and transportability across geographic, time, and methodology). Some scores have excellent performance in the population they were originally developed. Still, some lose the accuracy when applied to different populations because of the differences in the population characteristic from where the score was initially created.[2] Population characteristics, the severity of disease, and health policy differ within and between countries from time to time; therefore, the risk prediction models require regular validation and refinement (for example, by recalibration).[3]

The intensive care unit (ICU) scoring system for intensive care has been developed since 1980 in response to demands for the evaluation and monitoring of health services. The system allows for a comparative audit and evaluation of intensive service research. Several assessment systems have been developed for critically ill patients to predict the likelihood of patients surviving in the hospital. Mortality and prolonged stay (PLOS) prediction in critically ill can benefit health services for quality improvement.[1]

The Acute Physiology and Chronic Health Evaluation (APACHE) and Simplified Acute Physiology Score (SAPS) are the most widely used scoring systems in the ICU.[4] The APACHE IV was introduced in 2006. The development of the APACHE IV system is based on 104 ICUs and 131,618 patients from the United States of America (USA).[5]

The APACHE IV score has a good level of discrimination and calibration with a broadly validated sample size and receiver operating characteristics (ROC) of 0.88.[5] The APACHE IV score can be used as an ICU benchmark by using the standardized number of mortality ratio to evaluate groups of patients.[6]

This study is aimed to assess the predictive accuracy of the APACHE IV score on mortality and PLOS of patients in a single tertiary hospital.

MATERIAL AND METHODS

This retrospective study was based on the data of patients who were admitted to the ICU in SGH from January 1 to December 31, 2018. The study was conducted after the

Corresponding Author: Yunita Widyastuti
Email: yunita.widya@ugm.ac.id

Table I: Demographic Data of Research Participants

Variable	TOTAL n(%)	Dead (n=119) n(%)	PLOS(70) n(%)
Gender			
Female	256(53.4)	45(37.8)	30(42.9)
Male	220(46.6)	74(62.2)	40(57.1)
Age (years)			
< 20	12(2.5)	3(2.5)	2(2.9)
21-40	144(30.3)	25(20.0)	17(24.3)
41-60	210(44.1)	32(26.9)	32(45.8)
60-80	100(21.0)	52(43.7)	16(22.9)
>80	10(2.1)	7(5.9)	3(4.3)
Diagnosis at Admission			
Medical	108(22.7)	63 (52.9)	38(54.3)
Post Neurosurgery	172(36.1)	16 (13.4)	18(25.7)
Post Orthopedic surgery	31(6.5)	1 (0.8)	0(0)
Post Urologic surgery	17(3.6)	1 (0.8)	0(0)
Post Digestive surgery	43(9.0)	23 (19.3)	5(7.1)
Post Oncologic surgery	10(2.1)	1 (0.8)	1(1.4)
Post-Obstetric-gynecologic surgery	47(9.9)	7 (5.9)	5(7.1)
Post-Thorax/Vascular surgery	48(10.1)	7 (5.9)	3(4.3)

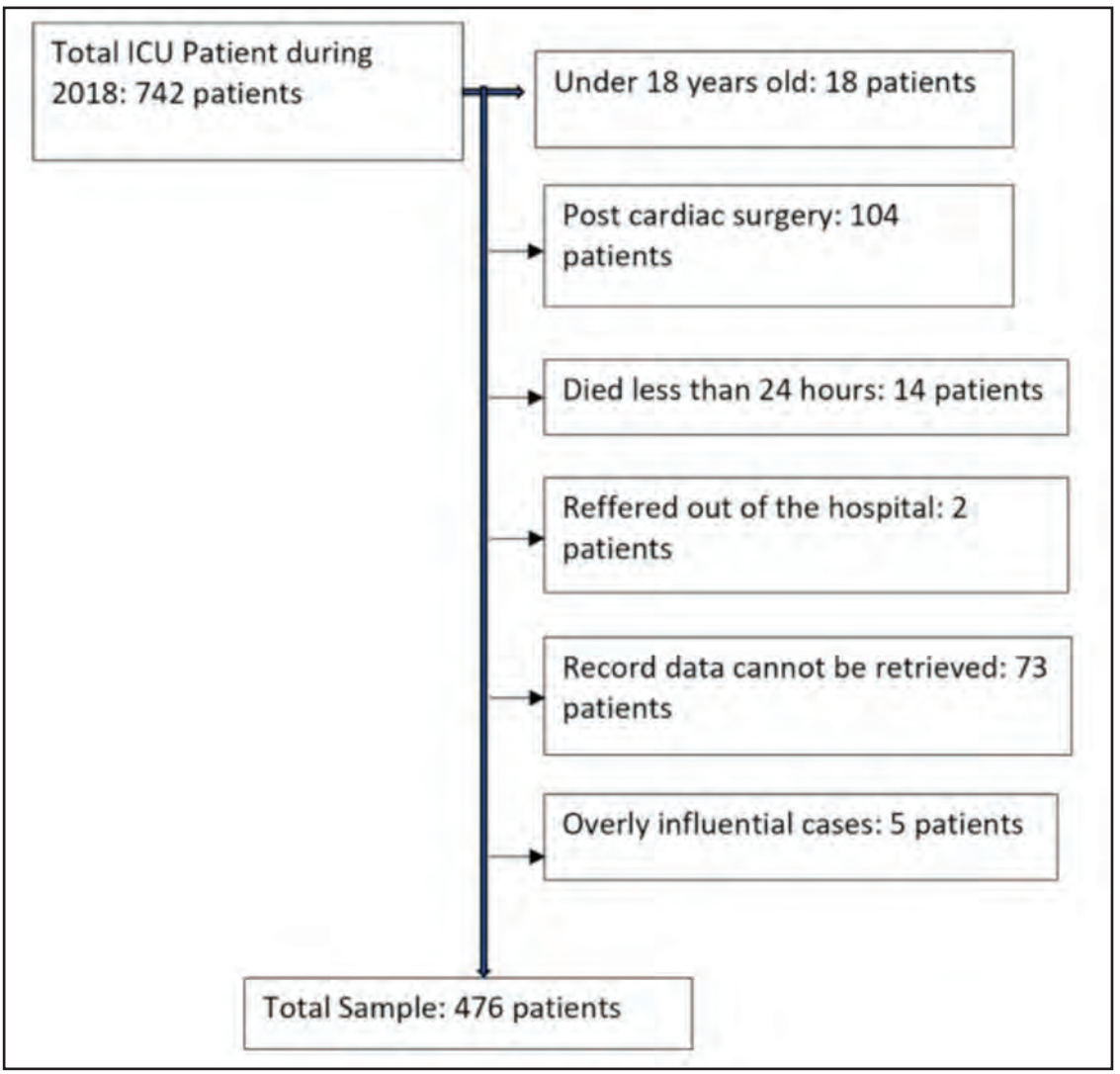

Fig. 1: The study sampling

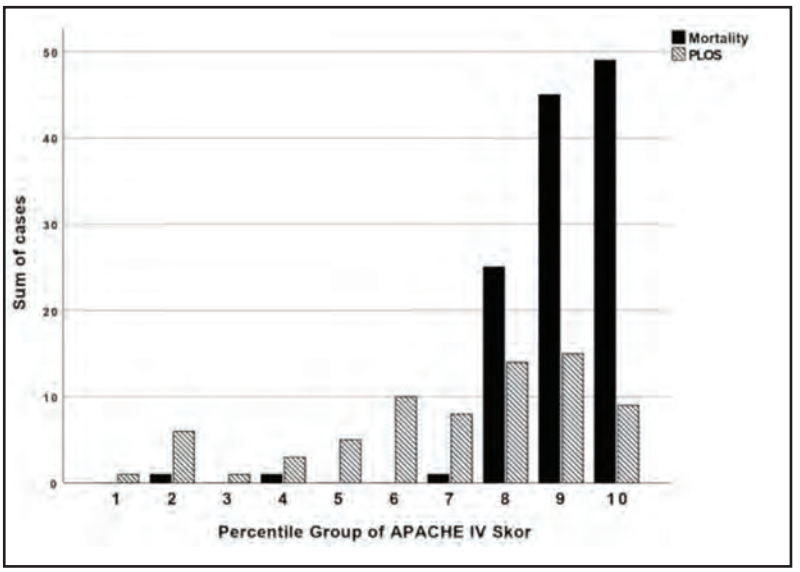

Fig. 2: The outcome based on the deciles of the acute physiology and chronic health evaluation (APACHE) IV.

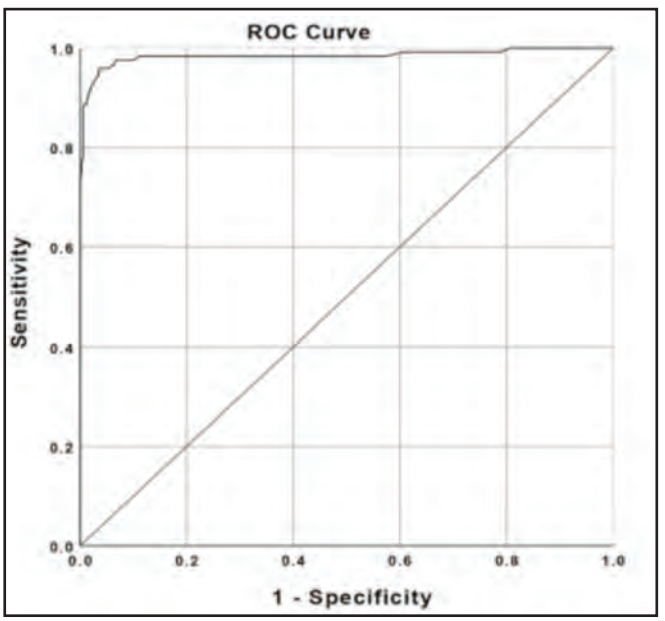

Fig. 3a: The discrimination of the acute physiology and chronic health evaluation (APACHE)IV score for mortality, the area under the curve (AUC) of Receiver Operating Characteristics (ROC) with 95% confidence interval (CI) for mortality: 0.99(0.97–1.00).

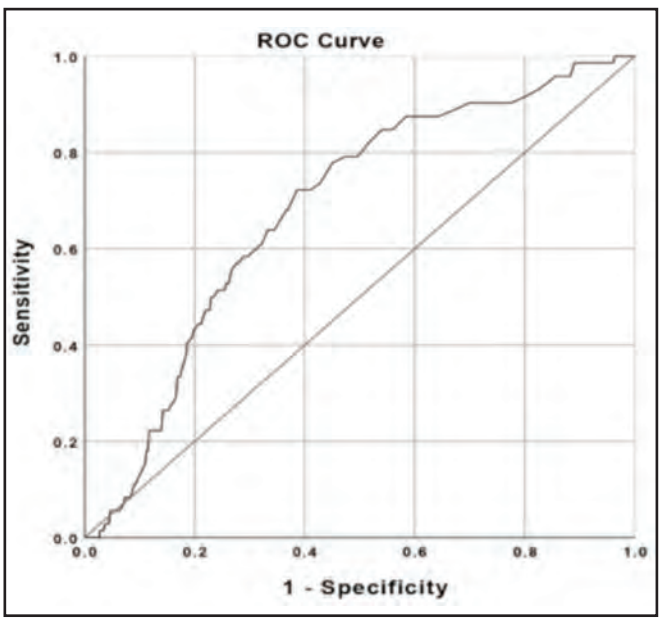

Fig. 3b: The discrimination of the acute physiology and chronic health evaluation (APACHE) IV score for PLOS, the area under the curve (AUC) of Receiver Operating Characteristics (ROC) with 95% confidence interval (CI) for a prolonged length of stay: 0.68 (0.62–0.74).

approval from the Ethical Committee No KE/FK/0684/EC/2019. The inclusion criteria in this study were all adult (≥18years old) patients admitted to the ICU of SGH. The exclusion criteria include the following: post-cardiac surgery, treatment <24 h, referred out of the hospital, and patients whose medical record data could not be collected during the sampling period. The dependent variable is mortality,defined as ICU mortality and PLOS as ICU stay >7 days. The independent variable in this study is the APACHE IV score.

The APACHE IV score was calculated according to the website (http://www.mecriticalcare.net/icu_scores/apacheIV.php), based on the patient's data in the first 24 h. The ROC producing an area under the curve (AUC) with 95% confidence intervals (CIs) was used to assess discrimination power. The optimal cut-off points of sensitivity and specificity were calculated using Youden Index. The calibration of the APACHE IV score was assessed with the Hosmer-Lemeshow goodness-of-fit test, and a p-value of >0.05 is considered a good calibration.

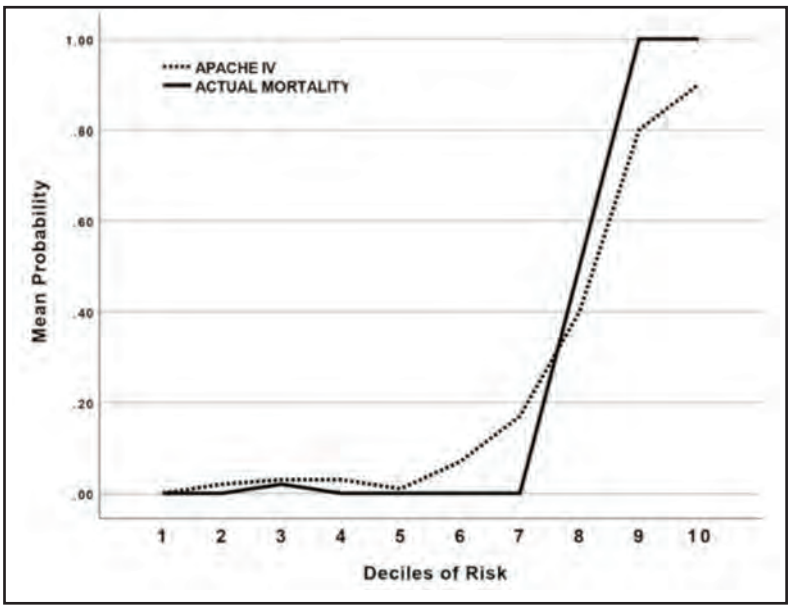

Fig. 4: The Observed versus Predicted Mortality by the acute physiology and chronic health evaluation (APACHE) IV score.

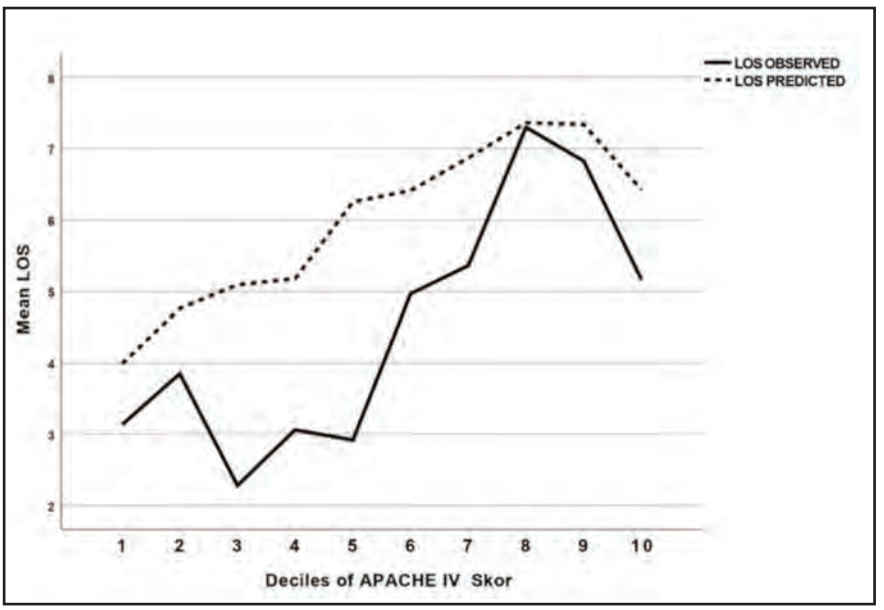

Fig. 5: The Observed versus Predicted prolonged length of stay by the acute physiology and chronic health evaluation (APACHE) IV score.

The ROC and Hosmer-Lemeshow tests were calculated using the Statistical Package for the Social Sciences (version 27.0 SPSS Inc., Chicago, IL, USA). Descriptive statistics were calculated with continuous data presented as mean ±SD and categorical data presented as percentages. Overly influential variables were removed from the data.

RESULTS

The study was conducted on the medical records of 742 patients, but only 476 patients were included in the study sample (Fig. 1).

The baseline characteristics of the study group are presented in Table I. Of 476 patients, there were 220 males (46.6%), with an average age of 48±16 years. The mortality rate was 25.4%, the PLOS rate was 15.1%, and the APACHE IV score was underestimated at 23.4% and 6%, respectively. Male patients have a higher incidence of mortality (62.3%) and PLOS (58.3%). Patients aged 61–80 years had the highest mortality rate (44.1%), whereas 41–60 mostly experienced PLOS (45.8%).

Patients admitted to the ICU were divided into two large groups, namely, 108(22.7%) medical cases and 368 (77.3%) surgical cases, and most patients who died had a medical

admission diagnosis (52.9%) followed by digestive surgery 23(19.3%). Most patients with PLOS were diagnosed with a medical condition (52.8%), followed by a neurosurgery procedure 18 (25.7%) (Table 1).

Fig. 2 shows the outcome distribution of APACHE IV scores into deciles, which led that patients who died mainly in the APACHE score range 81–100 and patients who got PLOS especially were 61–80. The percentage of patients living was 100% in patients with APACHE IV scores of <40, whereas 100% of patients were dead in scores of>100,which reveal the fact that the greater the APACHE IV score, the higher the mortality rate, but not the same for PLOS.

The accuracy of the APACHE scores to discriminate the patients who died was excellent with AUC with a 95% CI for mortality: 0.99(0.97–1.00) (Fig. 3a). The AUC (95%CI) for PLOS was 0.68 (0.62–0.74), which is considered poor (Fig.3b). The optimal cut-off point of the APACHE IV score for mortality was 78.9 (in the 8th percentiles), with a sensitivity of 0.96 and specificity of 0.96. The optimal cut-off point for PLOS prediction was 62.5 (in the 6th percentiles), with a sensitivity of 0.72 and specificity of 0.61.

The calibration of the APACHE IV score for mortality was good with a p-value of 0.98, but poor for PLOS with a p-value=0.01. The APACHE score slightly overestimated the mortality in the low risk and underestimated the high risk (Fig.4). For PLOS prediction, the APACHE IV score tends to overestimate the length of stay (Fig. 5).

DISCUSSION

This study revealed that the APACHE IV score consistently gives good discrimination in predicting mortality. The accuracy of predicting mortality was higher than the accuracy of predicting PLOS. This finding is consistent with another study.[5-14] The discrimination of the APACHE IV score for mortality prediction in our research is as good as in the original population, despite the different characteristics of a population. In our study, the age of ICU patients was younger, and the mean of the APACHE score was higher (66.27 vs 46.43).[5] The number of patients who died with the APACHE IV score of >100 was 59% in our study, whereas the original APACHE IV score was only 47%, and the remaining 39.7% had the APACHE IV score of 80–100.

The APACHE IV score is a physiology-based classification system that measures disease severity in a group of patients with a critical illness. The accuracy of the APACHE IV score plays a role in improving patient services. Its ability to provide mortality prognosis has been tested in USA and parts of Europe and Asia.[5,7–9]

The calibration of the APACHE IV score in predicting mortality is good. A study by Choi et al. also revealed the superiority of the APACHE IV calibration (p=0.905) compared to the APACHE II score (p=0.805).[9] Costa et al. (2011) prospectively examined APACHE IV, SAPS III, and MPMIII scores in critical patients with acute renal failure and showed good discrimination (AUROC 0.74) and calibration (p-value:0.574).[14] The calibration of the APACHE IV on mortality, especially in high-risk patients, is underestimated. However, the overall calibration on mortality is still good. However, some studies showed that the mortality prediction of the APACHE IV was not accurate. The differences in patient characteristics, clinical practice, assurance, quality, and services of health care systems may make different outcomes. The study by Chan et al.[15] reported that the accuracy of the APACHE IV score for mortality prediction was poor. They conducted a retrospective study of a group of postoperative patients with surgical abdominal sepsis and showed that the APACHE IV score poorly predicted mortality. The other study also reported that the APACHE IV score did not accurately predict mortality in patients with trauma.[16]

Recently, the accuracy of the APACHE IV score was compared to the APACHE II and Sequential Organ Failure Assessment (SOFA) score for mortality in patients with Coronavirus disease in the ICU and showed that all scores had bad discrimination on overall population (the APACHE IV AUROC 0.67 vs 0.63 of APACHE II vs 0.53 of SOFA score). On the other hand, the APACHE IV had the best discriminative power of the three scoring systems in the subgroup of patients with low molecular weight heparin (APACHE IV AUROC 0.82 vs 0.7 of APACHE II vs 0.49 of SOFA score).[17]

The APACHE IV score lacked both discrimination and calibration in predicting PLOS and tends to overestimate prediction. Patients with PLOS in our population mainly were patients with a medical diagnosis, sepsis, and more than one organ failure. Thus, they tended to have a shorter length of stay because they died earlier than patients with a lower score. Another factor determining the inaccuracy of the APACHE IV score in predicting PLOS is the differences in time units to calculate the length of treatment from the time patients enter until discharge. The original study of APACHE scores used hours, whereas our study used days.[6] The APACHE IV scores have poor performance in patients with severe sepsis[6] and have poor accuracy in the length of stay prediction, especially in severe sepsis and trauma cases.[11,18]

The difference in hospital policy related to the patient's end-of-life status, case-mix differences, insurance policies, step-down policy, and differences in clinical practice between USA and Indonesia affected the accuracy in predicting PLOS. This study has several limitations. Firstly, 9.8% of medical records cannot be taken and secondly only data for1-year was taken, and more extensive data may give a different result.

CONCLUSION

APACHE IV score has excellent accuracy for mortality prediction but poor accuracy for PLOS prediction in patients admitted in the ICU of SGH.

ACKNOWLEDGMENT

This research is funded by RSUP Dr Sardjito grant no: HK.02.03/XI.2/8268/2019.

CONFLICT OF INTEREST

None to declare.

REFERENCES

1. Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: A compendium. Indian J Crit Care Med 2014; 18(4): 220–8.
2. Myers PD, Ng K, Severson K, Kartoun U, Dai W, Huang W, et al. Identifying unreliable predictions in clinical risk models. npj Digit Med. 2020; 3(1): 1-8.
3. Shrestha GS. Developing a feasible and valid scoring system for critically ill patients in resource-limited settings. Crit Care 2018; 22(1): 1–2.
4. 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(6): 798–803.
5. Zimmerman JE, Kramer AA, McNair DS,Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. Crit Care Med 2006; 34(5): 1297–310.
6. Zimmerman JE, Kramer AA, McNair DS, Malila FM, Shaffer VL. Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. Crit Care Med 2006; 34(10): 2517–29.
7. Shrope-Mok SR, Propst KA, Iyengar R. APACHE IV versus PPI for predicting community hospital ICU mortality. Am J Hosp Palliat Care 2010; 27(4): 243–7.
8. Choi JW, Park YS, Lee YS, Park YH, Chung C, Park DI, et al. The Ability of the Acute Physiology and Chronic Health Evaluation (APACHE) IV score to predict mortality in a single tertiary hospital. Korean J Crit Care Med 2017; 32(3): 275–83.
9. Ghorbani M, Ghaem H, Rezaianzadeh A, Shayan Z, Zand F,Nikandish R. A study on the efficacy of APACHE-IV for predicting mortality and length of stay in an intensive care unit in Iran. F1000Res 2017; 20(6): 2032.
10. Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, de Jonge E, Bosman RJ, Peelen L, et al. External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. J Crit Care 2011; 26(1): 105.e11–8.
11. Keegan MT, Gajic O, Afessa B. Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and influence of resuscitation status on model performance. Chest 2012; 142(4): 851–8.
12. Lee H, Shon Y-J, Kim H, Paik H,Park H-P. Validation of the APACHE IV model and its comparison with the APACHE II, SAPS 3, and Korean SAPS 3 models for the prediction of hospital mortality in a Korean surgical intensive care unit. Korean J Anesthesiol 2014; 67(2): 115–22.
13. Costa e Silva VT, de Castro I, Liano F, Muriel A, Rodriguez-Palomares JR, Yu L. Performance of the third-generation models of severity scoring systems (APACHE IV, SAPS 3 and MPM-III) in acute kidney injury critically ill patients. Nephrol Dial Transplant 2011; 26(12): 3894–901.
14. Chan T. Evaluation of APACHE-IV predictive scoring in surgical abdominal sepsis: A retrospective cohort study. J Clin Diagn Res 2016; 10(3): PC16–8.
15. Wong RS, Ismail NA,Tan CC. An external independent validation of APACHE IV in a Malaysian intensive care unit. Ann Acad Med 2015; 44(4): 6.
16. Toker MK. SAPS 3 or APACHE IV: Which score to choose for acute trauma patients in intensive care unit? Ulus Travma Acil Cerrahi Derg 2018; 25(3): 247–52
17. Vandenbrande J, Verbrugge L, Bruckers L, Geebelen L, Geerts E, Callebaut I, et al. Validation of the Acute Physiology and Chronic Health Evaluation (APACHE) II and IV Score in COVID-19 Patients. Crit Care Res Pract 2021; 2021(19): 1–9.
18. Chattopadhyay A and Chatterjee S. Predicting ICU length of stay using APACHE-IV in persons with severe sepsis – a pilot study. J Epidemiol Res 2016; 2(1): 1–8.