e-ISSN (online): 2745-9497

Journal of Anesthesiology and Clinical Research (JACR)

Journal website: https://hmpublisher.com/index.php/jacr

Validity of Acute Physiology and Chronic Health Evaluation (APACHE) IV for the Prediction of Prolonged Intensive Care Unit (ICU) Length of Stay in Dr. Sardjito General Hospital in the COVID Era

Muhammad Mufti Sofyanoor[1], Yunita Widyastuti[2*], Juni Kurniawaty[2], Djayanti Sari[2]

[1]International Undergraduate Program of Medicine, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

[2]Department of Anesthesiology and Intensive Care, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

ARTICLE INFO

Keywords:
Acute physiology and chronic health evaluation-IV
Intensive care unit
Length of stay
Scoring system

***Corresponding author:**
Yunita Widyastuti

E-mail address:
yunita.widya@ugm.ac.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.37275/jacr.v4i2.302

A B S T R A C T

Introduction: APACHE IV was a good predictor of ICU length of stay in the USA and some countries outside the USA but poor in others. It is important to develop a scoring system for the Indonesian population, especially in this scope, Dr. Sardjito General Hospital. To develop such a scoring system, it is reasonable to study the validity of APACHE IV in ICU Dr. Sardjito General Hospital for predicting prolonged length of stay. **Methods:** A retrospective cohort observational study using data from January 1st, 2020, to December 31st, 2020, taken from the ICU of Dr. Sardjito General Hospital. The data are the patient's observed ICU LOS and data required in calculating APACHE IV score and ICU LOS prediction. Discrimination is calculated using the area under (AUC) the receiver operating characteristic curve (ROC) and calibration by the Hosmer-Lemeshow test. **Results:** Samples were 329 patients. APACHE IV ICU length of stay prediction showed moderate discriminatory ability (AUC-ROC: 0.74) and poor calibration ($p < 0.001$) to predict prolonged ICU stay. The APACHE IV score has a strong discriminatory ability (AUC-ROC: 0.83). Using the DeLong method, the AUC from ROC APACHE IV score was greater than the AUC from ROC predicted length of stay in APACHE IV ICU ($p < 0.001$). APACHE IV predicted ICU length of stay overestimated observed ICU length of stay. **Conclusion:** APACHE IV ICU length of stay prediction has moderate discrimination and poor calibration to predict prolonged ICU stay. The APACHE IV score has better discrimination than the APACHE IV ICU length of stay prediction in predicting prolonged ICU stay.

1. Introduction

The use of a scoring system is important in intensive care unit (ICU) settings as the usage of ICU is costly and with limited resources available. The scoring system was first developed as a tool to assess the effectiveness of care[1] to allocate resources more efficiently. The scoring system then is also used to predict patients' outcomes correctly and stratify them by severity as a consideration in taking medical decisions.[2] Other than for stratifying patient severity, scoring systems also have the feature of a probability model, mostly the prediction of mortality rate.[3]

The use of a scoring system to predict patients' length of stay (LOS) in the ICU is limited. Whereas ICU LOS is a significant variable influencing patient outcomes.[4] APACHE IV ICU LOS prediction was modeled and first validated using 131.618 consecutive ICU admissions from 2002 until 2003 in ICUs across the USA.[5] The result of this study was that the difference between the mean observed and mean

predicted ICU LOS was only 1.9 hours. This difference is significant by statistic, but this is only because of the huge sample used, which renders a slight difference as significant. However, a significant amount of variability was found in the calibration curve. This indicates that APACHE IV ICU LOS prediction might not be accurate enough to identify the exact ICU LOS of each patient. Small portions of patients who have prolonged lengths of stay (PLOS) would cost more resources than the other patients with shorter ICU LOS.

Therefore, the ability to be able to predict which patients may have PLOS also important. This approach might be more realistic rather than exactly determining a patient's ICU LOS. By using this approach, a study performed in the USA using data from 2002 until 2007 found that APACHE IV predicted ICU LOS on the day of admission is significantly higher in patients with observed PLOS (defined as ICU LOS of 5 days) compared to non-PLOS patients.[6] The ICU structures, management, and patient care in the USA might be different from other regions.[7] This difference may affect ICU LOS. Therefore, there is a need to perform validation of APACHE IV before using it in another region. Several validations have been done outside the USA, but the result is mostly not good.[8-11]

Therefore, adjustments of APACHE IV to the intended population are performed. For example, a study using the data from 2001 until 2004 in California do recalibration.[12] Before developing such a scoring system, however, it is reasonable first to validate the scoring system. In this case, the performance of APACHE IV predicted ICU LOS in the Indonesian population. Validation of APACHE was rarely done in Indonesia, especially one that evaluates APACHE IV performance in predicting ICU LOS or PLOS. Two studies performed show that APACHE IV underestimates ICU LOS.[13-14] Recently, a study was performed in Sardjito Hospital, the same setting as the current study.[15] This study shows APACHE IV score has weak discrimination and poor calibration to predict PLOS in Sardjito Hospital. Studies previously performed in Indonesia were done before the COVID-19 pandemic. The COVID-19 pandemic might change the case mix, ICU policy and way of conduct, resources available, and overall disease severity in a hospital. Therefore, this study was determined to validate the performance of APACHE IV to predict PLOS in the ICU of Dr. Sardjito General Hospital with patients admitted during 2020, which is already in the COVID-19 pandemic era.

2. Methods

This study is considered an observational retrospective cohort study. The data in this study were collected from the ICU of Dr. Sardjito General Hospital. The data collected are from patients admitted between January 1st, 2020, and December 31st, 2020. This study was approved by the medical and health research ethics committee of Universitas Gadjah Mada, Indonesia (KE/FK/1165/EC/2022).

Subjects are patients who have been admitted to the ICUs of Dr. Sardjito General Hospital between January 1st, 2020, and December 31st, 2020. The method of sampling is consecutive non-probability sampling, where all subjects that have fulfilled the subject inclusion and exclusion criteria will be recruited into the study sequentially. The data of the subjects are collected using consecutive sampling with several conditions; Inclusion criteria: patients of age ≥ 18 and ICU stay of ≥ 24 hours. Exclusion criteria: patients admitted after coronary artery bypass surgery, patients admitted from other ICUs of other hospitals, patients being admitted to other hospital ICUs before being discharged or determined as deceased from ICUs of Dr. Sardjito General Hospital, incomplete data of more than 3 variables or cannot be calculated using the APACHE IV calculator.

APACHE IV score and APACHE IV predicted ICU LOS first would be calculated using web-based calculators. Samples then categorize as PLOS with a definition of ≥ 6 days ICU LOS. Samples with observed ICU LOS less than 6 days but determined to be dead are also included as PLOS. Calculation of discrimination using the area under the receiver operating curve (AUC-ROC) and calibration using the Hosmer-Lemeshow test was then performed. The DeLong method is then used to compare the performance of the APACHE IV score and APACHE IV ICU LOS prediction in predicting PLOS. In addition, as

a secondary outcome, a comparison of mean and median, as well as correlation also performed.

3. Results and Discussion

From the data of 353 ICU patients, 332 patients met the inclusion criteria. Of 21 patients that are not included, 14 are due to the age of < 18 years old, and 7 are due to observed ICU LOS < 24 hours. From 332 patients included, 3 patients are excluded. 2 of which are due to incomplete data of > 3 variables, and the last 1 is due to incomplete data which cannot be calculated using the APACHE IV calculator. Thus, in the end, this study analyzed data from 329 ICU patients.

Table 1. Sample demography.

Characteristics	N (%)
Gender (n = 329)	
Male	162 (49,24)
Female	167 (50,76)
Diagnosis (n = 329)	
Medical	
Cardiovascular	17 (5,17)
Respiratory	41 (12,46)
Digestive	1 (0,3)
Neurologic	22 (6,69)
Metabolic	5 (1,52)
Hematologic	1 (0,3)
Genitourinary	2 (0,61)
Sepsis	57 (17,33)
Trauma	6 (1,82)
Surgical	
Cardiovascular	5 (1,52)
Respiratory	15 (4,56)
Digestive	19 (5,78)
Neurosurgery	92 (27,96)
Genitourinary	36 (10,94)
Trauma	6 (1,82)
Non-traumatic amputation	4 (1,22)
Mean of Age ± SD	49,86 ± 16,02

Most of the samples (67,18%) included the age of < 60 years old, the cut-off age for geriatric in Indonesia.[16] The mean age is 49,86 years old, which is younger than the average age of the original APACHE IV publication, which is 61,45 ± 0,08 years old.[5] This means the population in this study is dominated by relatively younger samples compared to the original study. Males and females are about equal, with 162 females (49,24%) and 167 males (50,76%). This shows that there is no appreciable disparity in numbers between the genders. There are 177 patients (53,8%) who are classified as surgical patients, compared to 152 patients (46,2%) who are classified as medical patients. This contrasts with the initial publication of APACHE IV, where medical patients (69,2%) are more prevalent than surgical ones (30,8%).[5] In medical diagnosis, sepsis is the most prevalent diagnosis with 57 samples (17,33% of the total sample and 37,5% of all medical diagnoses). Surgical diagnosis is dominated by neurosurgery with 92 samples (27,96% of the total sample and 52% of all surgical diagnoses).

The range of APACHE IV score that has the most PLOS is 61-80 with 51 samples (15,5% from the total sample and 36,1% from the total observed PLOS). The percentage of PLOS gradually increase from the least to the highest APACHE IV score. In the original study, ICU LOS gradually increase until a score of 75 before gradually decreasing due to mortality[5]. PLOS in the current study includes mortality as well, which explains PLOS keeps increasing. The highest mortality rate is in the range of 81 – 100 score. Mortality also gradually increases and decreases only in > 100 scores.

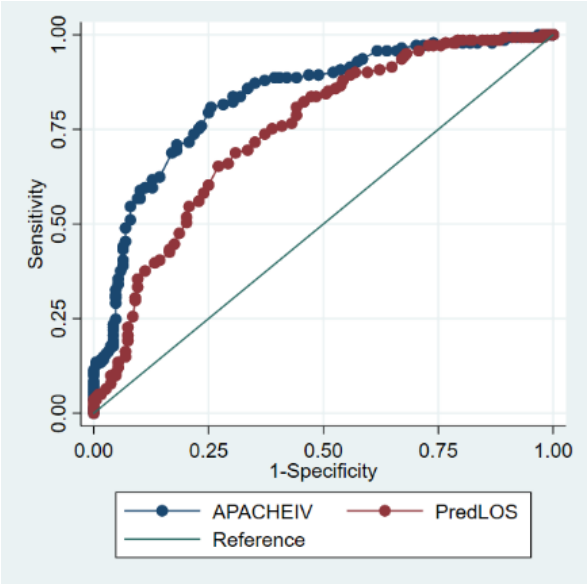

Figure 1. AUC-ROC comparison between APACHE IV predicted ICU LOS and APACHE IV score.

The area under the curve (AUC) of Receiver Operating Characteristics (ROC) with a 95% confidence interval (CI) for APACHE IV predicted ICU LOS against observed PLOS is 0,74 (0,69-0,8). This means that the discriminative power in predicting PLOS using APACHE IV predicted ICU LOS is moderate.[17] The area under the curve (AUC) of Receiver Operating Characteristics (ROC) with a 95% confidence interval (CI) for APACHE IV score against observed PLOS is 0,83 (0,79-0,89). This means that the discriminative power in predicting PLOS using the APACHE IV score is strong.[17] The most optimal APACHE IV score cut-off to predict observed PLOS with 77,2% correct classification is >= 62 (Sn: 80,85%; Sp: 74,47%) or >= 68 (Sn: 70,92%; Sp: 81,91%). This means that patients with scores above this cut-off point will be more likely to experience the outcome of PLOS, and treatments for this kind of patient must be handled with more caution. The DeLong method can calculate whether an AUC is significantly larger than another AUC. The result of the comparison is AUC of APACHE IV score is significantly larger than the AUC of APACHE IV predicted ICU LOS with a p-value < 0.001.

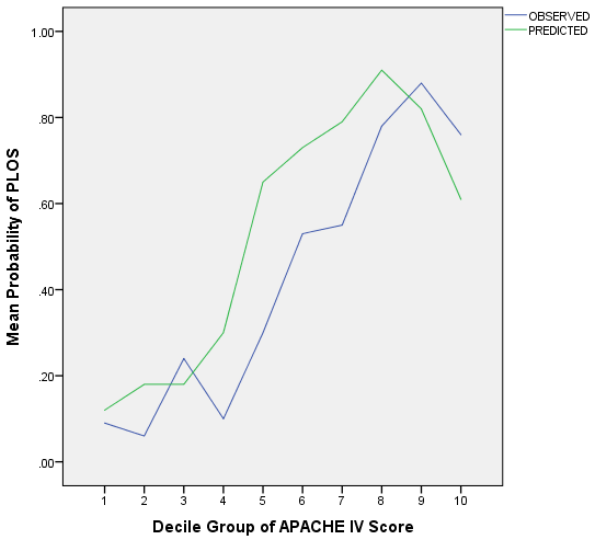

Figure 2. Calibration model between APACHE IV predicted and observed PLOS.

Calibration of APACHE IV predicted PLOS is poor with p-value < 0.001 and X^2 = 57.72. PLOS is being overestimated in the 1st, 2nd, and 4th until the 8th decile. On the other hand, underestimation can be seen in the 3rd, 9th, and 10th decile of the APACHE IV score.

The mean of observed LOS is 5.25 days, and the mean of APACHE IV predicted ICU LOS is 5.93. It can be seen that there is a slight overestimation of 0.68 days from this. This is different from the original study, which results in a slight underestimation of 0.08 days.[5] The difference in the median is quite large (3.1 days), and it is statistically significantly different using Mann–Whitney U test. Several studies performed in Indonesia also show ICU LOS overestimation.[13,14] On the other hand, studies outside Indonesia show underestimation.[8-11]

The mean APACHE IV predicted ICU LOS in the observed non-PLOS population is 5.1, while in the observed PLOS population is 7.2. This difference is statistically significant (p < 0.001). This means that predicted ICU LOS in the observed PLOS population is significantly higher compared to predicted ICU LOS in the observed non-PLOS population.

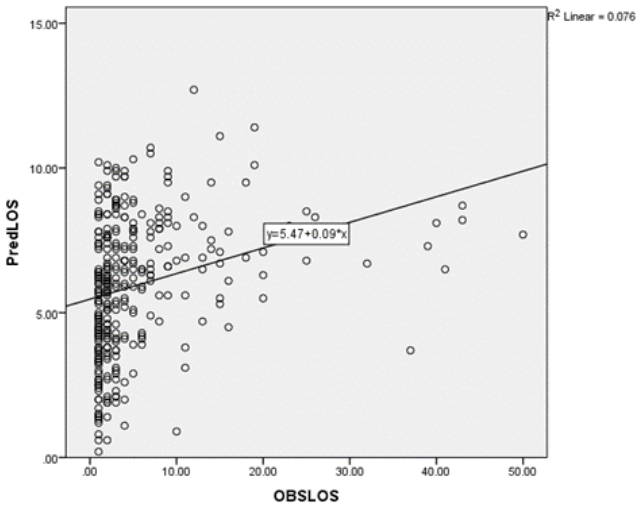

Figure 3. Distribution of observed LOS and predicted LOS.

From the graph above, it is shown that the correlation between predicted and observed LOS is positive. The variation shown in the graph is R2 = 0.076. This means only 7.6% of observed LOS can be explained with the predicted LOS using the linear equation. This number is quite less compared to the original study, which has R2 = 0.215.[5] Spearman correlation score was r = 0.406, a moderate positive correlation with p = < 0.001. This means higher observed LOS corresponds with higher predicted LOS. However, its strength is only moderate.[17]

To understand APACHE IV performance in general, a discussion of secondary outcomes is better to take place first. This is because the most study of LOS prediction using APACHE IV does not categorize LOS as prolonged but use it as a continuous variable. In most studies, APACHE IV predicted ICU LOS performance is presented by the mean difference between predicted and observed LOS. It is interesting to see that the overestimation of APACHE IV predicted ICU LOS could be seen in studies performed in Indonesia only. This is probably caused by a higher rate of mortality in Indonesian studies. ICU LOS overestimation can be seen in high-risk patients that are dead.[5]

Even in the original study, APACHE IV predicted ICU LOS could not accurately predict observed LOS. This is especially true for individual patients. The original study shows that APACHE IV predicted ICU LOS has satisfying results in predicting the mean ICU LOS of the patient group or used for the ICU benchmarking. However, APACHE IV predicted ICU LOS in its original study shows only slight underestimation, although this difference is

statistically significant.[5] In other studies, the difference between observed ICU LOS and APACHE IV predicted ICU LOS is larger. APACHE IV predicted ICU LOS in other studies may not be as accurate as original studies due to differences in ICU condition, policy, facility, and resources. Even in a study performed in California, USA, APACHE IV predicted ICU LOS only acceptable after recalibration was performed.[12]

Seeing from another point of view, another study performed in the USA shows APACHE IV predicted ICU LOS is significantly larger in the patient group with observed PLOS (≥ 5 days) compared to the patient group that does not have PLOS[6]. This aligned with the result from this study that shows the mean APACHE IV predicted ICU LOS in the observed non-PLOS population is 5.1, while in the observed PLOS population is 7.2. This difference is statistically significant ($p < 0.001$). This indicates that APACHE IV predicted ICU LOS might be able to differentiate patients that have PLOS.

The main result of this study is the discrimination and calibration of APACHE IV to predict PLOS. The discrimination power of ICU LOS prediction is considered moderate, while the discrimination power of the APACHE IV score is strong. This difference in discrimination power is then tested with the DeLong method, and the APACHE IV score significantly has better discrimination power compared to APACHE IV predicted ICU LOS to predict observed PLOS. As the discrimination power of APACHE IV score is better than APACHE IV predicted ICU LOS, it needs to be noted that the usage of APACHE IV predicted ICU LOS might not be as good as APACHE IV score in predicting patient with PLOS and need to be used carefully.

Determination of APACHE IV discrimination in predicting PLOS is rarely done. A recent study performed such calculation in the same setting as the current study but with samples from before the pandemic era.[15] In the study, discrimination of APACHE IV score is calculated using the same method as the current study, AUC under ROC, with an AUC of 0.68 (weak discrimination power[17]). The COVID pandemic might cause this difference in APACHE IV performance, especially its discrimination.

Changes in the case mix, which are indicated by how the age and number of comorbidities changed during the pandemic, had been observed.[18] In addition, there has been an increase in the ICU mortality rate of non-COVID patients during the COVID pandemic, especially at the peak of the number of COVID-19 cases.[19,20] The pandemic situation causes the mortality rate to be higher in all patients' severity. These changes might directly or indirectly affect the discrimination ability of APACHE IV predicted ICU LOS.

APACHE IV predicted ICU LOS tends to overestimate PLOS in groups of severe patients that die. This is because severe patients that are predicted to have PLOS but dead usually died early and therefore did not have PLOS. An increase in mortality rate will make severe patients' PLOS overestimation more prevalent. However, because the current study considers mortality as PLOS, PLOS overestimation in severe patients did not happen, and underestimation happened instead. This is because severe patients that are predicted to have PLOS but die early are now included as PLOS as well. This close the gap between observed and predicted PLOS making discrimination higher.

On the other hand, in groups of mild patients that unexpectedly die, an underestimation of PLOS would happen. This is because, in APACHE IV original study, mild patients that were dead had PLOS as well.[5] These dead patients that usually have PLOS was predicted to be non-PLOS, and this caused underestimation. In previous studies in Indonesia, APACHE IV predicted ICU LOS tends to be an overestimation assumingly in any patient's severity, including mild patients. The pandemic situation, which increases the mortality rate, will presumably reduce this overestimation as mortality cause underestimation in a mild patient. Therefore, the current study definition (that accounts for mortality as PLOS) should not affect PLOS prediction in a mild patient, as patients that have mortality should also have PLOS. However, in the current study, many of the mild patients that have mortality died early. Therefore, patients that are categorized as PLOS increase even further, causing even less overestimation. This further closes the gap

between observed and predicted PLOS making discrimination also higher.

Different ICU way of conduct and resources between the current study and the original study is probably what led to the condition that makes mild patient that has mortality die early as well. However, the pandemic situation might also contribute to this. The pandemic presents with unprecedented ICU capacity surge causing health worker workload to increase.[20] When there is a shortage of health workers, health workers may be forced to prioritize the interventions that must be given to patients, such as cutting back on CPR, conducting suction as necessary, and lowering the standard of care to increase the number of interventions.[21] There is also a prioritization of COVID patients.[22] Non-COVID patients are on lack supervision, and the number of patients is limited when there is no emergency.

Despite quite a good result of discrimination, the calibration of APACHE IV in this study is poor (p-value < 0.001 and X2 = 57.72). APACHE IV predicted ICU LOS calibration is poor, even in its original study and a validation study in the USA.[5,12] These studies, however, only showed poor calibration with a low value of R2. Once again, it is claimed that this happened because APACHE IV predicted ICU LOS is intended for a group of patients and not individual patients. The original study claimed that an R2 of 0.215 is high enough considering the nature of ICU LOS, which has many unmeasurable factors affecting it.[5]

The study has several limitations. First, some data of variables are incomplete in several patients, which may alter the result of the APACHE IV score and ICU LOS prediction. This problem, however, has been tackled properly. Second, due to this study being retrospective, some aspects of subjectivity might happen, especially in the process of determining admission diagnosis and completion of missing data.

4. Conclusion

APACHE IV predicted ICU LOS has moderate discrimination in predicting PLOS. APACHE IV score has stronger discrimination in predicting PLOS compared to APACHE IV ICU LOS prediction. The calibration of APACHE IV for predicting PLOS is poor.

5. References

1. Rao MH, Marella P, Kath B. Assessment of severity and outcome of critical illness. 2008.
2. Parsons PE, Wiener-Kronish JP. Critical care secrets. Elsevier; 2007.
3. Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: A compendium. Indian J Crit Care Med. 2014; 18(4): 220–8.
4. Akavipat P, Thinkhamrop J, Thinkhamrop B, Sriraj W. Parameters affecting length of stay among neurosurgical patients in an intensive care unit. Acta Med Indones. 2016; 48(4): 275–81.
5. Zimmerman JE, Kramer AA, McNair DS, Malila FM, Shaffer VL. Parameters affecting length of stay among neurosurgical patients in an intensive care unit. Crit Care Med. 2006; 34(10): 2517–29.
6. Kramer AA, Zimmerman JE. A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. BMC Med Inform Decis Mak. 2010; 10(1): 27.
7. 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).
8. Chattopadhyay A, Chatterjee S. Predicting ICU length of stay using APACHE-IV in persons with severe sepsis – a pilot study. Journal of Epidemiological Research. 2015.
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; 6(2032).
10. Hu Y, Zhang X, Liu Y, Yan J, Li T, Hu A. APACHE IV is superior to MELD scoring system in predicting prognosis in patients after orthotopic liver transplantation. Sun Q, editor. Clin Dev Immunol. 2013; 2013: 809847.
11. Kamal M, Khan A, Ali G. A comparison of APACHE II and APACHE IV scoring systems in predicting outcome in patients with acute lung injury (ALI) and the adult respiratory distress

syndrome (ARDS) in intensive care unit (ICU). Rawal Medical Journal. 2013; 38: 234–8.
12. Vasilevskis EE, Kuzniewicz MW, Cason BA, Lane RK, Dean ML, Clay T, et al. Mortality probability model III and simplified acute physiology score II: Assessing their value in predicting length of stay and comparison to APACHE IV. Chest. 2009; 136(1): 89–101.
13. Zaki WA, Jufan AY, Pratomo BY. Association of APACHE IV scores with mortality prognosis and LOS of patients in the ICU of Dr. Sardjito General Hospital. [Yogyakarta]: Universitas Gadjah Mada; 2019.
14. W HP, Hadisaputro S, Supriyadi S. Comparison of the use of APACHE IV, SAPS 3 and SOFA to predict mortality in critical patients. Jurnal Riset Kesehatan. 2015; 4(1).
15. Widyastuti Y, Arsyad ZW, Widodo U, Yun JA, Yudo PB. Predictive accuracy of the APACHE IV scores on mortality and prolonged stay in the intensive care unit of Dr. Sardjito Hospital. Med Journal Malaysia. 2022; 77(1): 53–8.
16. Ministry of Health of the Republic of Indonesia. Regulation of the Minister of Health of the Republic of Indonesia Number 25 of 2016 concerning the National Action Plan for Elderly Health for 2016-2019. Kementrian Kesehatan Republik Indonesia, NOMOR 25 TAHUN 2016 Indonesia; 2016.
17. Dahlan MS. Statistics for medicine and health. Jakarta: Salemba Medika; 2011.
18. Docherty AB, Mulholland RH, Lone NI, Cheyne CP, De Angelis D, Diaz-Ordaz K, et al. Changes in in-hospital mortality in the first wave of COVID-19: a multicentre prospective observational cohort study using the WHO Clinical Characterisation Protocol UK. Lancet Respir Med. 2021; 9(7): 773–85.
19. Zampieri FG, Bastos LSL, Soares M, Salluh JI, Bozza FA. The association of the COVID-19 pandemic and short-term outcomes of non-COVID-19 critically ill patients: an observational cohort study in Brazilian ICUs. Intensive Care Med. 2021; 47(12): 1440–9.
20. Wilcox ME, Rowan KM, Harrison DA, Doidge JC. Does unprecedented ICU capacity strain, as experienced during the COVID-19 pandemic, impact patient outcome? Crit Care Med. 2022; 50(6).
21. Mailani F, Muthia R, Huriani E, Krisdianto BF, Oktarina E. The challenges of intensive care unit nurses caring for COVID-19 patients in Indonesia: A qualitative study. Nurse Media Journal of Nursing. 2022; 12(2): 233–48.
22. Dhamanti I, Indriani D, Miftahussurur M, Kurniawati E, Engineer CY. Impact of hospital readiness on patient safety incidents during the COVID-19 pandemic in Indonesia: health worker perceptions. BMJ Open. 2022; 12(7).