

Effectiveness of APACHE II and SAPS II scoring models in foreseeing the outcome of critically ill COPD patients

Yousef Ahmed, Mohamed Adam, Lamees M. Bakkar

Background Acute Physiology and Chronic Health Evaluation II (APACHE II) and Simplified Acute Physiology Score II (SAPS II) scoring systems are the two models that are greatly used by the majority of ICUs to predict clinical consequence.

Objective The aim of the study was to assess the performance of APACHE II and SAPS II scoring methods in foreseeing death among critically ill chronic obstructive pulmonary disease (COPD) patients.

Materials and methods This prospective research included 104 COPD patients who were admitted to the respiratory intensive care unit (RICU) at Assiut University Hospital. The patients were classified as survivors and nonsurvivors. Each scoring system was assessed for its discrimination, calibration, and overall performance.

Results On the basis of the outcome of the study population, 36 (34.6%) patients were non-survivors while 68 (65.4%) patients were survivors. Both APACHE II and SAPS II scores were significantly higher in nonsurvivors. The discriminative power of both models was good as determined by the receiver operating characteristic curve. At a cutoff point greater than 20 for APACHE II and greater than 48 for SAPS II, survival or death can be predicted. The Lemeshow–Hosmer goodness-

of-fit C statistics showed good performance and good calibration for both models. APACHE II score had the least Brier score and reliability but had the highest resolution.

Conclusion The conclusions made were first, APACHE II and SAPS II have nearly similar performance in predicting mortality among COPD patients but with some preference for APACHE. Second, Both models have good discrimination and good calibration.

Egypt J Bronchol 2019 13:654–659

© 2020 Egyptian Journal of Bronchology

Egyptian Journal of Bronchology 2019 13:654–659

Keywords: Acute Physiology and Chronic Health Evaluation II, Assiut, chronic obstructive pulmonary disease, respiratory intensive care unit, Simplified Acute Physiology Score II

Chest Department, Faculty of Medicine, Assiut University, Assiut, Egypt

Correspondence to Yousef Ahmed, MD, Chest Department, Faculty of Medicine, Assiut University Hospital, Assiut, 71515, Egypt.

Tel: +20 102 503 3083; fax +2088233327;

e-mail: yousef_ahmed1972@yahoo.com

Received: 23 August 2019 **Accepted:** 24 November 2019

Published: xx xx 2020

Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive and debilitating airway disease that results in a large burden, both medically and financially. It affects millions of people around the world and causes great rates of morbidity and mortality. This burden is anticipated to increase with an estimated 5.8 million deaths annually by 2030 [1]. A large proportion of patients with COPD usually require admission to the ICU and it may be helpful to recognize patients at the time of admission who are probable to have bad consequence, so that these patients can be managed violently [2]. There are many ICU scoring models, and numerous new systems are being progressed to assess severity of illness in ICU patients. The use of scoring models particularly developed for patient evaluation at the time of ICU entry has decreased many troubles and helped therapy delineation. Furthermore, these methods aid in assessing and comparing the goodness and magnitude of care between different health-care academies [3,4]. Acute Physiology and Chronic Health Evaluation II (APACHE II) and Simplified Acute Physiology Score II (SAPS II) scoring systems are the two models that are greatly used by the majority of ICUs to forecast the clinical consequence [5]. The aim of our study was to assess the performance of APACHE II and SAPS II scoring

methods in forecasting death among critically ill COPD patients admitted to the respiratory intensive care unit (RICU) at Assiut University Hospital.

Materials and methods

This prospective, descriptive, comparative research was performed from January 2018 to March 2019 and included 104 COPD patients who were admitted to the RICU with severe exacerbation requiring admission to the RICU (severe dyspnea that responds inadequately to initial emergency therapy, changes in mental status, persistent or worsening hypoxemia, persistent or worsening respiratory acidosis, the need for invasive mechanical ventilation, and/or hemodynamic instability). The diagnosis of COPD was based on the patient's medical history obtained from the patient himself and/or the family of the patient, consistent physical findings, previous spirometry and/or evidence of hyperinflation on current or previous chest radiograph. Excluded from this study were COPD

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

patients admitted to the ICU because of any other underlying problem such as those with acute cardiac situation, patients with respiratory failure due to other diseases along with COPD, patients who lasted less than 24 h in the ICU, and those who died before the completion of data collection. Our study was accepted by the Scientific Ethics Committee of Faculty of Medicine of Assiut University. Before participation, informed written consent to deal with the patient's data for scientific purposes was obtained from the patients or persons in charge of them. We made sure to preserve patient privateness by not publishing identifying data. All patients involved have undergone full history with special stress on age, gender, special habits, and associated diseases. Information such as the patient's need for mechanical ventilation as well as the duration of lodging in the ICU was entered. Routine laboratory variables were also registered. Patients were followed up until their outcome was determined. The outcome was decided according to the mortality within the ICU and recorded as survivors and nonsurvivors. Scores of APACHE II and SAPS II were calculated as illustrated in the previously published works [6,7].

Statistical analysis

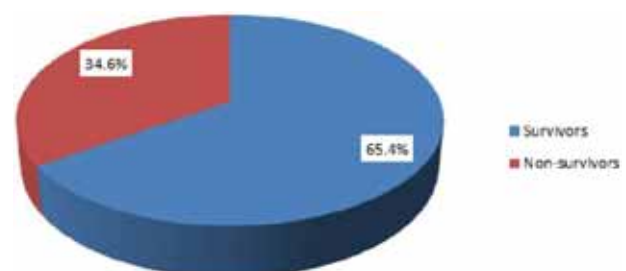
Statistical Package for the Social Sciences (Version 20; IBM, Armonk, New York, USA) was used for analyzing the collected data. The nominal data was expressed in the form of frequency (percentage) and continuous data were expressed in the form of mean \pm SD or range. Student's *t*-test was used to compare between continuous data while the nominal data was compared by χ^2 -test. Each scoring system was assessed for its discrimination (the ability of the scoring system to differentiate patients who die in the ICU from those who survive) and calibration (the degree of compatibility between calculated expectation of death produced by the scoring system and the true mortality). The discriminatory power was assessed by calculating the area under curve (AUC) in the receiver operating characteristic (ROC) curve (when this area approximates 1.0, the system becomes more ideal whereas when this area is in the direction of about 0.5, system performance becomes unexpected and inaccurate). Calibration of the model was assessed by Lemeshow-Hosmer goodness-of-fit C-statistic and calibration curve. (Patients were ranked in order of probability of death and were divided into 10 groups of approximately equal number of observations; predicted and observed deaths in these groups were compared using Pearson's statistics (c). The model with the least c and the highest *P* value shows the best agreement between the observed and the predicted number of deaths.) The calibration curve, also named calibration

plot, intends to provide complementary information. (If the model calibrates well, there will not be a substantial deviation from the 45° line of perfect fit or bisector. On the contrary, miscalibration of the model will be a function of expected probability.) To calculate the standardized mortality ratio (SMR), the observed mortality was divided by the predicted mortality. (If the SMR is equal to 1, then this means the number of observed deaths equals that of expected cases. If higher than 1, then there is a higher number of deaths than is expected.) The overall performance was finally evaluated by Brier score in order that estimating the predictive accuracy of the scoring system. It measures the average squared difference between the forecasted probabilities of consequences. A lower score represents a higher precision [8–10].

Results

This study enrolled 104 patients with acute exacerbation of COPD that required admission to the RICU. Based on the outcome of the study population, 36 (34.6%) patients were non-survivors while 68 (65.4%) patients were survivors (Fig. 1). Table 1 shows baseline data of the studied patients based on the outcome. These data include age, sex, smoking, associated diseases, exposure to mechanical ventilation, duration of ICU stay, APACHE II score, and SAPS II score. The mean age of survivors was 62.67 \pm 10.61 years, while the mean age of nonsurvivors was 63.87 \pm 8.75 years. Majority of the studied patients were men. The proportion of smoking and comorbidities is higher in nonsurvivors but not statistically significant. The percentage of mechanically ventilated patients among the nonsurvivor group was significantly higher than those among survivors (80.6 vs 53%; *P*=0.03). It was noticed that the duration of ICU stay was significantly higher in non-survivors in comparison with survivors (14.36 \pm 5.94 vs 11.22 \pm 5.71 days; *P*=0.04). Both APACHE II and SAPS II scores were significantly higher in non-survivors (APACHE II: 24.61 \pm 6.96 vs 17.22 \pm 5.91; *P*<0.001 and SAPS II:

Figure 1

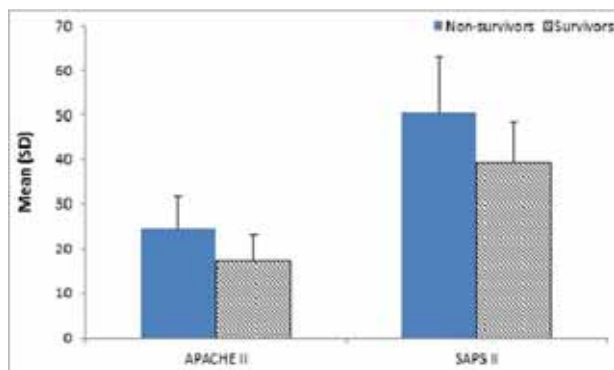


Outcome of the studied patients.

Table 1 Baseline data of the studied patients based on the outcome

	Nonsurvivors (n=36)	Survivors (n=68)	P value
Age (years)	63.87±8.75	62.67±10.61	0.53
Sex			
Male	29 (80.6)	52 (76.5)	0.39
Female	7 (19.4)	16 (23.5)	
Smoking	30 (83.3%)	52 (76.5%)	0.09
Comorbidities	13 (38.2)	23 (33.8)	0.11
Mechanically ventilated	29 (80.6)	36 (53)	0.03
ICU's stay (days)	14.36±5.94	11.22±5.71	0.04
APACHE II	24.61±6.96	17.22±5.91	<0.001
SAPS II	50.44±12.77	39.26±9.06	<0.001

Data expressed as frequency (percentage) and mean±SD. APACHE II, acute physiology and chronic health evaluation version II; SAPS II, simplified acute physiology score version II. P value was significant if <0.05.

Figure 2

Mean APACHE II and SAPS II scores among survivors and non-survivors. APACHE II, Acute Physiology and Chronic Health Evaluation II; SAPS II, Simplified Acute Physiology Score II.

Table 2 Diagnostic accuracy of APACHE II and SAPS II scoring systems in predicting mortality among the studied patients

	APACHE II (%)	SAPS II (%)
Sensitivity	72	61
Specificity	79	88
Positive predictive value	65	73
Negative predictive value	84.4	81
Accuracy	76.9	78.8
Cutoff point	>20	>48
Area under the curve	0.80	0.74

APACHE II, Acute Physiology and Chronic Health Evaluation version II; SAPS, simplified acute physiology score version II.

50.44±12.77 vs 39.26±9.06; $P<0.001$). Further illustration is shown in Fig. 2 for both APACHE II and SAPS II scores between survivors and nonsurvivors. Regarding the diagnostic accuracy of the severity scoring models used in our study, it can be explored from Table 2 and Fig. 3 where at a cutoff greater than 20, APACHE II

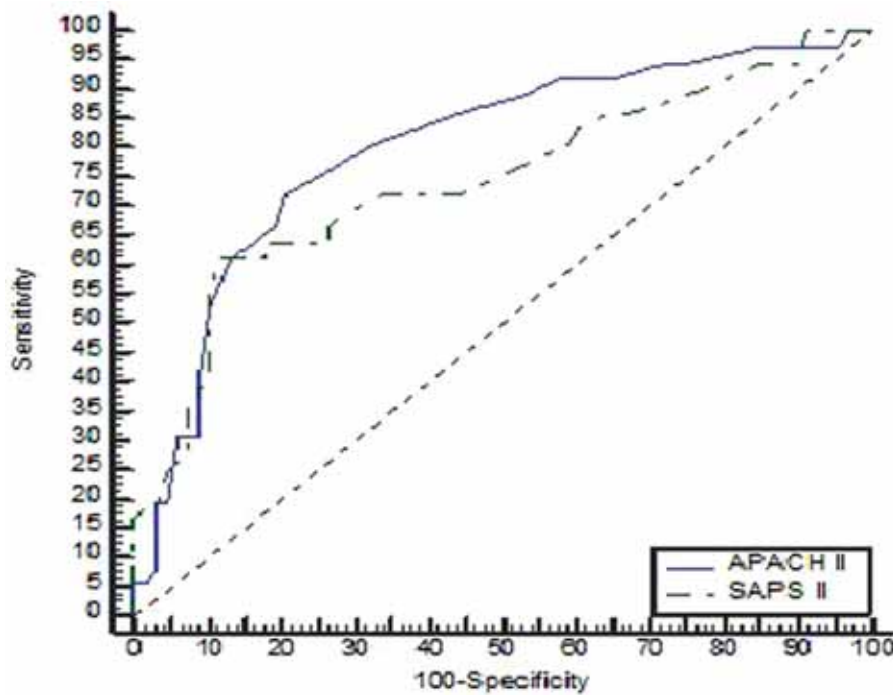
had a sensitivity of 72% and a specificity of 79% for prediction of mortality with an overall accuracy of 76.9%, while at a cutoff greater than 48, SAPS II had 61% sensitivity and 88% specificity for prediction of mortality with overall accuracy being 78.8%. Tables 3 and 4 detail the Lemeshow–Hosmer goodness-of-fit statistics for APACHE II and SAPS II, respectively. It was noticed that the SMR for APACHE II was equal to 1, which denotes that the expected mortality by this score was equal to the observed mortality, while SMR for SAPS II equals 0.97 which denotes that the expected mortality by SAPS II score was more than the observed mortality and so, SAPS II overestimated in-ICU mortality but this was not statistically significant. Lemeshow–Hosmer goodness-of-fit statistics for SAPS II was 4.30. Both models calibrated well on formal goodness-of-fit testing, and had low C-statistic and low P values. Although APACHE II calibrated the best, insignificant difference was found between the two models (Fig. 4). As regards the overall performance of both models based on Brier score, it was observed that APACHE II score had the least Brier score and reliability but had the highest resolution, so it is considered the best score in the present study for prediction of mortality (Table 5).

Discussion

The ICU in the chest department at Assiut University Hospital is the most important ICU for referral for respiratory diseases in Upper Egypt. It has 26 well-equipped beds and accommodates critically ill adult patients with acute or exacerbated respiratory failure caused by a disease that is primarily respiratory. To our knowledge, this is the first research that aims primarily to evaluate the performance of severity scoring models in predicting mortality among COPD patients admitted to our RICU. In this study, we evaluated the two most common models, APACHE II and SAPS II, in critically ill COPD patients who are the most commonly admitted cases to the RICU [11–13]. These two scoring models were compared in various previous studies that yielded somewhat contradictory results. However, we like to draw attention to the fact that we have not been able to obtain any previous studies comparing the performance of these models in predicting ICU mortality among patients with COPD in particular, but what we got is research on intensive care patients without allocating a specific disease or studies that have generally looked at predicting outcomes in critically ill COPD patients, so it may be difficult to compare our results with those of other studies.

This study enrolled 104 patients, based on the outcome of the study population; 36 (34.6%) patients were non-

Figure 3



Diagnostic accuracy of APACHE II and SAPS II scoring systems in predicting mortality among the studied patients. APACHE II, Acute Physiology and Chronic Health Evaluation II; SAPS II, Simplified Acute Physiology Score II.

Table 3 Lemeshow–Hosmer goodness-of-fit statistics for APACHE II

Probability groups	Survivors		Nonsurvivors		Total
	Observed	Expected	Observed	Expected	
1	10	10.153	1	0.847	11
2	9	8.883	1	1.117	10
3	10	9.284	1	1.716	11
4	9	8.674	2	2.326	11
5	8	7.475	2	2.525	10
6	8	7.71	3	3.29	11
7	5	5.39	4	3.61	9
8	3	4.941	7	5.059	10
9	3	3.721	8	7.279	11
10	3	1.77	7	8.23	10
Observed mortality			36		
Expected mortality			35.99		
SMR			1		
Lemeshow–Hosmer C test			3.46		
P value			0.90		

APACHE II, Acute Physiology and Chronic Health Evaluation version II; SMR, standardized mortality ratio.

survivors while 68 (65.4%) patients were survivors. We found that both APACHE II and SAPS II scores were significantly higher in non-survivors than in survivors. Mohan *et al.* [14], Timmers *et al.* [15], and El-Shahat *et al.* [12] reported findings that are consistent with our results. Mean APACHE II and SAPS II scores among nonsurvivors were 24.61 and 50.44, respectively. In the

previous reports, the mean scores among this group of patients were 23.17 and 51.55 in a study by Aminiahidashti *et al.* [16], 23.91 and 55.14 in a study by Faruq *et al.* [5], 31.5 and 63.9 in a study by Godinjak *et al.* [17], and 23.14 and 46.14 in a study by Fadaizadeh *et al.* [4]. The findings of our study showed that 20 and 48 could be considered as reasonable cutoff points for APACHE II and SAPS II, respectively, which could predict survival or death. Previously recorded cutoff values for both models, respectively, were 27.5 and 50 in a study by Godinjak *et al.* [17], 13 and 44 in a study by Haq *et al.* [18], 13.5 and 27.5 in a study by Fadaizadeh *et al.* [4], and 14 and 26 in a study by Kandil *et al.* [19]. To assess the discriminative power of APACHE II and SAPS II, ROC curve and AUC must be studied. In our study, these areas were 0.80 and 0.74 for APACHE II and SAPS II, respectively, which means that these two scoring systems are almost similar for AUC and have good discrimination power, although the discrimination was better slightly for APACHE II than for SAPS II. Previously reported area under the ROC curve of APACHE II and SAPS II included 0.83 and 0.87 in a study by Katsaragakis *et al.* [20], 0.78 and 0.81 in a study by Moreno *et al.* [21], 0.83 and 0.79 in a study by Arabi *et al.* [22], 0.81 and 0.84 in a study by Noura *et al.* [23], and 0.88 and 0.87 in a study by Tan [24]. The Lemeshow–Hosmer goodness-of-fit C statistics revealed good performance

and good calibration for both models in our study. Although APACHE II calibrated the best, little difference was found between the two models. This was similar to the findings reported in studies by Naqvi

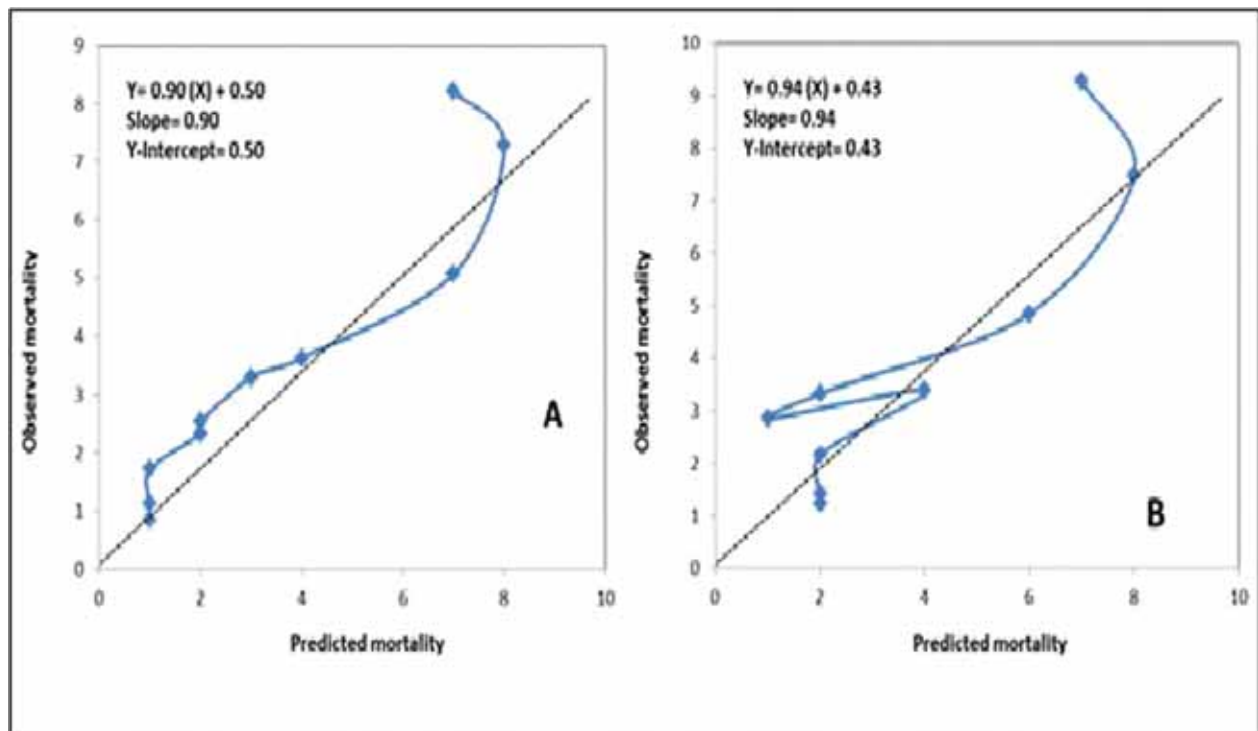
et al. [25] and Fadaizadeh *et al.* [4]. In contrast, calibration of the two models had some amount of lack of fit in studies by Aminiahidashti *et al.* [16] and Khwannimit and Geater [26]. In this study, we observed that the SMR for APACHE II was equal to 1 and SMR for SAPS II was 0.97. This result indicates that APACHE II model predicted accurately the ICU mortality while SAPS II score slightly overestimated the mortality. Del Bufalo *et al.* [27] concluded that APACHE II score was a good predictor of ICU mortality and better than SAPS II, with a ratio between actual and predicted mortality being 86% for APACHE II and 83% for SAPS II. In a study by El-Shahat *et al.* [12] SMR with APACHE II was 90.3% and thus mortality was overestimated to a lesser extent. However, SMR with SAPS II was 119.6% and therefore it reduces the mortality to a higher extent. Regarding overall performance, we found that the Brier score of APACHE II and SAPS II was 0.47 and 0.53, respectively. In addition, the reliability of both models was 0.01 and 0.03 while the resolution was 0.23 and 0.20. In a study by Aminiahidashti *et al.* [16] the Brier score of APACHE II and SAPS II were 0.21 and 0.20, respectively. In addition, the reliability was 0.02 and 0.01 but this study did not comment on the resolution.

Table 4 Lemeshow–Hosmer goodness-of-fit statistics for SAPS II

Probability groups	Survivors		Nonsurvivors		Total
	Observed	Expected	Observed	Expected	
1	8	10.767	2	1.233	10
2	7	7.584	2	1.416	9
3	10	9.825	2	2.175	12
4	8	11.61	4	3.39	12
5	8	8.149	1	2.851	9
6	8	6.666	2	3.334	10
7	5	6.158	6	4.842	11
8	4	4.528	8	7.472	12
9	3	2.712	7	9.288	10
10	7	10.767	2	1.233	9
Observed mortality			36		
Expected mortality			37.23		
SMR			0.97		
Lemeshow–Hosmer C test			4.30		
P value			0.77		

SAPS, simplified acute physiology score version II; SMR, standardized mortality ratio.

Figure 4



Calibration curve of (a) APACHE II, it shows that if predicted mortality increases by 1, the observed mortality increases by 0.90 and (b) SAPS II, it shows that if the predicted mortality increases by 1, the observed mortality increases by 0.94. APACH II, Acute Physiology and Chronic Health Evaluation II; SAPS II, Simplified Acute Physiology Score II.

Table 5 Overall performances of different scores based on Brier score

	Brier score	Resolution	Reliability
APACHE II	0.47	0.23	0.01
SAPS II	0.53	0.20	0.03

APACHE II, Acute Physiology and Chronic Health Evaluation version II; SAPS, simplified acute physiology score version II.

It was not strange to find different results among many studies regarding the performance of APACHE II and SAPS II. On the contrary, it was expected, perhaps for several reasons, including, first, varying quality of the ICUs across the world; second, the number of cases enrolled in the studies and their diagnosis; third, the difference in approach used to scoring the Glasgow coma scale in sedated patients; and fourth, the differences in the treatment received by the patients during the time between the hospital admission and their admission to the ICU or what is called admission time delay.

What we consider a limitation in our study is that we did not take into account the quality of care and type of treatment received by the patient during time before ICU admission. However, by looking at the results of our study, we can conclude that, first, APACHE II and SAPS II have nearly similar value in predicting mortality among COPD patients but with some preference for APACHE; and second, both models have good discrimination and good calibration.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- May SM, Li JT. Burden of chronic obstructive pulmonary disease: healthcare costs and beyond. *Allergy Asthma Proc* 2015; **36**:4–10.
- Khilnani GC, Banga A, Sharma SK. Predictors of mortality of patients with acute respiratory failure secondary to chronic obstructive pulmonary disease admitted to an intensive care unit: a one year study. *BMC Pulm Med* 2004; **4**:12.
- Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: a compendium. *Indian J Crit Care Med*. 2014; **18**:220–228.
- Fadaizadeh L, Tamadon R, Saeedfar K, Jamaati H. Performance assessment of Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II in a referral respiratory intensive care unit in Iran. *Acta Anaesthesiol Taiwanica* 2012; **50**:59–62.
- Faruq MO, Mahmud MR, Begum T, Ahsan TS, Fatema K, Ahmed F, et al. A comparison of severity systems APACHE II and SAPS II in critically ill patients. *Bangladesh Crit Care J* 2013; **1**:27–32.
- Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. *Crit Care Med* 1985; **13**:818–829.
- 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.
- Hnley J, McNeil B. The meaning and the use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**:29–36.
- Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982; **115**:92–106.
- Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 1950; **78**:1e3.
- Ghoneim A, Hussein R, El-Ghamry R, Mahmoud L. Patterns of admitted cases to Respiratory Intensive Care Unit at Zagazig University Hospitals. *Egypt EJCT* 2013; **62**:661–668.
- El-Shahat H, Salama S, Wafy S, Bayoumi H. Risk factors for hospital mortality among mechanically ventilated patients in respiratory ICU. *Egypt J Bronchol* 2015; **9**:231–237.
- Khalil MM, Salem HR, El Tohamy MF. Characteristics and clinical outcome of patients treated in the respiratory ICU of Abbassia Chest Hospital. *Egypt J Bronchol* 2019; **13**:93–99.
- Mohan A, Bhatt SP, Mohan C, Arora S, Luqman-Arafath TK, Guleria R. Derivation of a prognostic equation to predict in-hospital mortality and requirement of invasive mechanical ventilation in patients with acute exacerbation of chronic obstructive pulmonary disease. *Indian J Chest Dis Allied Sci* 2008; **50**:335–342.
- Timmers TK, Verhofstad MHJ, Moons KGM, Leenen LPH. Validation of six mortality prediction systems for ICU surgical populations. *Neth J Crit Care* 2011; **15**:118–130.
- Aminiahdashti H, Bozorgi F, Montazer SH, Baboli M, Firouzian A. Comparison of APACHE II and SAPS II Scoring systems in prediction of critically ill patients' outcome. *Emergency* 2017; **5**:e4.
- Godinjak A, Igllica A, Rama A, Tančica I, Jusufović S, Ajanović A. Predictive value of SAPS II and APACHE II scoring systems for patient outcome in medical intensive care unit. *Acta Med Acad* 2016; **45**:89–95.
- Haq A, Patil S, Parcels AL, Chamberlain RS. The Simplified Acute Physiology Score III Is Superior to the Simplified Acute Physiology Score II and Acute Physiology and Chronic Health Evaluation II in Predicting Surgical and ICU Mortality in the 'Oldest Old'. *Curr Gerontol Geriatr Res* 2014; **2014**:934852.
- Kandil A, Kenawi M, Samir A, Hussein K. Traumatic brain injury predictive value of common intensive care severity scores. *Res Opinion Anesth Intensive Care* 2017; **4**:124–128.
- Katsaragakis S, Papadimitropoulos K, Antonakis P, Stregiopoulos S, Konstadoulakis MM, Androulakis G. Comparison of acute physiology and chronic health evaluation II (APACHE II) and simplified acute physiology score II (SAPS II) scoring systems in a single Greek intensive care unit. *Crit Care Med* 2000; **28**:426–432.
- Moreno R, Morais P. Outcome prediction in intensive care: results of a prospective, multicentre, Portuguese study. *Intensive Care Med* 1997; **23**:177–186.
- Arabi Y, Haddad S, Goraj R, Al-Shimemeri A, Al-Malik S. Assessment of performance of four mortality prediction systems in a Saudi Arabian intensive care unit. *Crit Care* 2002; **6**:166–174.
- Nouira S, Belghith M, Elatrous S, Jaafoura M, Ellouzi M, Boujdaria R. Predictive value of severity scoring systems: comparison of four models in Tunisian adult intensive care units. *Crit Care Med* 1998; **26**:852.
- Tan IK. APACHE II and SAPS II are poorly calibrated in a Hong Kong intensive care unit. *Ann Acad Med Singapore* 1998; **27**:318–322.
- Naqvi IH, Mahmood K, Ziaullah S, Kashif SM, Sharif A. Better prognostic marker in ICU-APACHE II, SOFA or SAP III. *Pak J Med Sci* 2016; **32**:PMC5103123.
- Khwannimit B, Geater A. A comparison of APACHE II and SAPS II scoring systems in predicting hospital mortality in Thai adult intensive care units. *J Med Assoc Thai* 2007; **90**:643–652.
- Del Bufalo C, Morelli A, Bassein L, Fasano L, Quarta CC, Pacilli AM, et al. Severity scores in respiratory intensive care: APACHE II predicted mortality better than SAPS II. *Respir Care* 1995; **40**:1042–1047.