Mortality Prediction Using Acute Physiology and Chronic Health Evaluation II and Acute Physiology and Chronic Health Evaluation IV Scoring Systems: Is There a Difference?

Ramesh Venkataraman, Vijayaprasad Gopichandran[1], Lakshmi Ranganathan, Senthilkumar Rajagopal, Babu K Abraham, Nagarajan Ramakrishnan

Department of Critical Care Medicine, Apollo Hospitals, [1]Department of Community Medicine, ESIC Medical College and PGIMSR, Chennai, Tamil Nadu, India

Abstract

Background: Mortality prediction in the Intensive Care Unit (ICU) setting is complex, and there are several scoring systems utilized for this process. The Acute Physiology and Chronic Health Evaluation (APACHE) II has been the most widely used scoring system; although, the more recent APACHE IV is considered an updated and advanced prediction model. However, these two systems may not give similar mortality predictions. **Objectives:** The aim of this study is to compare the mortality prediction ability of APACHE II and APACHE IV scoring systems among patients admitted to a tertiary care ICU. **Methods:** In this prospective longitudinal observational study, APACHE II and APACHE IV scores of ICU patients were computed using an online calculator. The outcome of the ICU admissions for all the patients was collected as discharged or deceased. The data were analyzed to compare the discrimination and calibration of the mortality prediction ability of the two scores. **Results:** Out of the 1670 patients' data analyzed, the area under the receiver operating characteristic of APACHE II score was 0.906 (95% confidence interval [CI] – 0.890–0.992), and APACHE IV score was 0.881 (95% CI – 0.862–0.890). The mean predicted mortality rate of the study population as given by the APACHE II scoring system was 44.8 ± 26.7 and as given by APACHE IV scoring system was 29.1 ± 28.5. The observed mortality rate was 22.4%. **Conclusions:** The APACHE II and IV scoring systems have comparable discrimination ability, but the calibration of APACHE IV seems to be better than that of APACHE II. There is a need to recalibrate the scales with weights derived from the Indian population.

Keywords: Acute Physiology and Chronic Health Evaluation II, Acute Physiology and Chronic Health Evaluation IV, mortality prediction, mortality rate, scoring systems

INTRODUCTION

The outcomes of patients admitted to the Intensive Care Units (ICUs) depend on numerous factors including the age, sex, type, and severity of underlying illness and various clinical and laboratory parameters. Several scoring systems have been developed over the years to assess and describe the severity of the illness and predict the mortality rate of patients admitted to the ICU. One of the earliest valid mortality prediction tools is the Acute Physiology and Chronic Health Evaluation (APACHE) II system which was described in 1981.[1] The tool is administered on a patient admitted to the ICU within 24 h of admission wherein various parameters including the patient demographics, clinical features, and laboratory features are entered, and an APACHE score ranging between 0 and 71 is computed. Although APACHE scores may not be useful to determine outcomes of individual patients, it can work as a good guide to prognosticate in the cohort of patients admitted to ICU. Moreover, comparison of the actual mortality rates in an ICU with the predicted mortality rate (PMR) can be used to indicate the performance of an ICU and to compare outcomes across different ICUs. In 1991, APACHE II was modified to the APACHE III score which was more elaborate with 20 variables and used several additional parameters with better predictive ability.[2]

Address for correspondence: Dr. Ramesh Venkataraman, Department of Critical Care Medicine, Apollo Hospitals, 21 Greams Lane, Chennai - 600 006, Tamil Nadu, India. E-mail: ccmramesh@gmail.com

In 2006, the APACHE scoring system was further refined (APACHE IV) adding more predictor variables and revising the statistical modeling technique. The APACHE IV system was standardized and benchmarked for use in the United States ICUs as it had good predictive validity calibration.[3] It is well known that the predictive accuracy of these scoring systems is dynamic and need to be constantly revised and re-examined to account for the improvements in care delivery. If applied inappropriately the models may lead to wrong prediction and inappropriate use of time and resources.[4]

The APACHE II and IV scoring systems are widely used all over the world even though they were both originally developed among ICU patients in the United States and lack validation in all patients of all races and health-care systems. Despite the development and deployment of APACHE IV most major clinical studies and several ICUs to date use APACHE II score as their severity scoring system and use it for both patient prognostication and quality monitoring. It is unclear whether APACHE II and APACHE IV will yield concordant results in our patient population. Several studies in other countries comparing APACHE II and APACHE IV have provided conflicting results. They were all underpowered, evaluated a narrow spectrum of patients and/or had several limitations.[5-12]

In summary, there is a dearth of literature comparing the performance of APACHE II and APACHE IV in a general ICU among a wide variety of patients in the Indian setting. In this context, we sought to compare the mortality prediction ability of APACHE II and APACHE IV scoring systems among patients admitted to a multidisciplinary tertiary care ICU in India.

METHODS

This prospective longitudinal follow-up study was conducted in a tertiary care private hospital in Chennai, India. All critically ill patients admitted to the multidisciplinary ICU between April 2014 and December 2015 were included in this study. The data of all the included patients were extracted from their hospital records after the first 24 h of admission. Online APACHE calculator was used to calculate APACHE II and IV scores from the worst noted variables in 24 h of ICU admission. The patients were followed up until outcome (death or discharge) and the outcome documented. Patients who were discharged against medical advice or at request were excluded from analysis as the outcome could not be ascertained in these patients. For both scoring systems observed mortality rate (OMR) was compared with PMR. The standardized mortality ratios (SMRs) were also determined from the calculated PMR and OMR.

The data analysis was performed using SPSS Statistical Package version 17 (SPSS, IBM, Chicago, USA). Descriptive statistics on the characteristics of the patients was analyzed. The receiver operating characteristic (ROC) curves for the APACHE II and APACHE IV for predicting mortality were plotted. The area under the ROC (AUC) for the two scoring systems were compared. Further, using the various sets of sensitivities and specificities derived from the ROC, the cutoff values of the scores which provided the optimal sensitivity and specificity were identified. Kruskal–Wallis test was performed to compare the mortality rate predicted by APACHE II and APACHE IV scores.

RESULTS

In the study period, a total of 1859 patients were admitted to the multidisciplinary ICU. A total of 189 (10%) were excluded from the study because their outcomes were not available due to discharge at the request to other facilities or discharge against medical advice. Data of 1670 patients admitted in the ICU and followed up to outcome were analyzed. The mean age of these patients was 55 years with a predominance of males (66%). Majority of them (66%) were direct admissions from the emergency room, and 81% were admitted for medical causes. About 70% of the patients were ventilated during their ICU stay. These characteristics of the study population are depicted in Table 1.

These patients were admitted in the ICU for a wide variety of reasons including respiratory, septic, cardiovascular, neurological, renal, trauma-related, and hepato-pancreatic issues which are depicted in Table 2.

Figure 1 depicts the ROC of APACHE II and APACHE IV scores with the scores as continuous variables and the dependent variable of death or discharge. It is seen that both APACHE II and APACHE IV have almost similar, overlapping curves with APACHE II score having a slightly greater AUC than APACHE IV score.

Table 1: Characteristics of the study participants

Characteristic	Categories	Number of patients (%)
Mean age (years)	55.6±16.9	
Sex	Males	1113 (66.6)
	Females	557 (33.4)
Source of admission	Direct ICU admission	395 (23.7)
	Emergency room	1103 (66.0)
	High dependency unit	22 (1.3)
	Ward	150 (9.0)
Type of illness	Medical	1354 (81.1)
	Surgical	316 (18.9)
Ventilation status	Ventilated	1174 (70.3)
	Not ventilated	496 (29.7)
Outcome	Discharged	1296 (77.6)
	Died	374 (22.4)
Mean duration of ventilation (days)	3.46±4.05	
Mean ICU length of stay	7±18.9	
Mean APACHE II score	22.7±10.09	
Mean APACHE IV score	68.85±34.79	

ICU: Intensive Care Unit; APACHE: Acute Physiology and Chronic Health Evaluation

The AUC of APACHE II score is 0.906 (95% confidence interval [CI] – 0.890–0.992) and the AUC of the APACHE IV score is 0.881 (95% CI – 0.862–0.890). Table 3 compares the predictive validity of the two scoring systems.

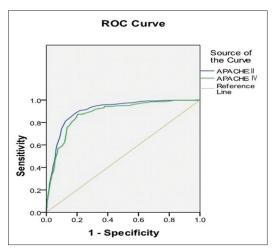The mean PMR of the study population as given by the APACHE II scoring system was 44.8 ± 26.7, and the mean

Figure 1: Receiver operating characteristic curves of Acute Physiology and Chronic Health Evaluation II and Acute Physiology and Chronic Health Evaluation IV scores in predicting Intensive Care Unit mortality

Table 2: Reasons for Intensive Care Unit admission among the study patients

Diagnosis	*n* (%)
Respiratory causes (acute respiratory distress, acute lung injury, chronic lung disease, etc.)	284 (17)
Sepsis	190 (11.3)
Cardiovascular causes (acute coronary event, congenital heart disease, rheumatic heart disease, etc.)	181 (10.8)
Neurological causes (cerebrovascular events, seizure disorders, etc.)	175 (10.4)
Renal causes (acute kidney failure, chronic kidney failure, etc.)	155 (9.2)
Trauma (road traffic crashes, poisoning, drowning, etc.)	127 (7.6)
Hepato-pancreatic causes (acute hepatic failure, chronic hepatic failure, pancreatitis, etc.)	102 (6.1)
Endocrine causes (diabetes, thyroid, acid-base imbalances, etc.)	93 (5.5)
Febrile illnesses (H1N1, HIV/AIDS, dengue, etc.)	87 (5.2)
Cancers	72 (4.3)
Other causes	204 (12.6)

Influenza A (H1N1) virus; HIV: Human Immunodeficiency Virus, AIDS: Acquired Immunodeficiency Syndrome

Table 3: Predictive validity of Acute Physiology and Chronic Health Evaluation II and Acute Physiology and Chronic Health Evaluation IV scoring systems

Scoring systems	Cut off value	Sensitivity (%)	Specificity (%)
APACHE II	25.5	88	81.5
APACHE IV	72.5	88	75.4

APACHE: Acute Physiology and Chronic Health Evaluation

PMR by the APACHE IV scoring system was 29.1 ± 28.5. The OMR was 22.4%. Thus, the SMR for the APACHE II score was 0.5 and for APACHE IV score was 0.70. The Kruskal–Wallis test of significance of the difference between the mean PMR of these two groups showed a $P < 0.001$.

DISCUSSION

This prospective longitudinal follow-up study of 1670 patients admitted to a multidisciplinary ICU in a private tertiary care hospital in Chennai, showed that APACHE II scoring system has a comparable AUC of the ROC curves to the APACHE IV score in predicting mortality in the ICU. Furthermore, the APACHE II score showed a greater PMR of 44.8% compared to the APACHE IV score which showed a PMR of 29.1%. For the same level of sensitivity, the APACHE II score had greater specificity than the APACHE IV score.

Both the APACHE II and APACHE IV scores are robust scoring systems for predicting mortality in the ICU setting. However, the application of both these systems is strongly context specific. The predictive validity of the systems depends on the demographics of the patients, the disease condition and severity, and the infrastructure and facilities of the ICU. There are several reasons why the APACHE scoring system performance is different in different settings. Mortality prediction models at their best estimate the mortality risk only for the population in which the model is developed.[13] Both the APACHE II and IV scoring models were developed for the North American population. Therefore, their validity in the Indian setting is likely to be different. Moreover, the relative performance of these two scoring systems is also likely to be different in the Indian setting. The other important consideration in the comparison of APACHE II and APACHE IV scores in this study is the method by which the scores were computed. The scores were computed using the internet based software into which the variables were entered. Previously studies have shown that such manual charting of data computes different scores compared to automated charting from health management information systems.[14] Previous studies have shown that APACHE IV performed better than APACHE II in conditions such as acute lung injury and neurological damage.[6-9] However, for conditions such as pancreatitis and sepsis APACHE II performed better than APACHE IV.[10-11] We chose to compare these scores in patients with a wide range of diagnoses admitted to a medical-surgical ICU. This could have led to the disparity of outcomes projected by these two scoring systems. The calibration of these scoring systems is also dynamic and dependent on the population on which the scoring is applied. The weightage given to the different variables in the APACHE II and APACHE IV scores are also dependent on the population characteristics.[15] Whereas APACHE II weights were given by a panel of experts, the APACHE IV weights were given by multiple logistic regression analysis. Therefore, APACHE IV scoring system weightage is more susceptible to change in different contexts than the APACHE II scoring system and

more likely to be dependent on the population characteristics than APACHE II.[16] This gain could lead to differences in outcome projections between the two scores.

Important questions that face the intensivist in the ICU setting are as follows:

What is the meaning of different PMR given by the two different scoring systems? As the discrimination ability of both the APACHE II and IV scoring systems are almost identical, does this mean that both predict mortality accurately? If so, why are the PMR given by the two scores different? As the two scoring systems give different PMR and SMR, which one should an ICU use routinely for assessing their performance?

For all the scales that are used to predict mortality in an ICU setting, there are two important characteristics namely model calibration and model discrimination. Calibration refers to the level of agreement between the estimated probability of mortality and the observed probability. If a model has good calibration, it means that the model can predict mortality well. Model discrimination shows the ability of the model to discriminate between those who will die and those who will survive the ICU admission. This is indicated by the AUC of the ROC curve of the mortality prediction scores.[17] Thus, lack of concordance of the PMR between the two scales indicates that the scale calibration is better for the APACHE IV scoring system, whereas the discrimination is marginally better for the APACHE II system. From this finding, it is not possible to conclude which of the two scoring systems can be used routinely in an ICU setting. As suggested by a previous review, these scoring systems are not competitive. They are complementary to each other and should be used in combination. For a given ICU setting the factors which largely drive the decision on which scoring system to use would be the ease of its application, the ability to standardize the measurements, the comfort level of the intensivists using the scale and the level of calibration that is possible to achieve in that setting.

The strength of this study is that it examines the comparability of the APACHE II and APACHE IV scores in a large sample size with diverse admission diagnoses in a multidisciplinary ICU. This is also one of the few studies comparing these scores in the Indian setting. The limitation of the study is that the APACHE II and IV scores were computed using online software. The weightage and calibration were based on the scoring system available from a North American patient population. The weightage was not obtained specifically for the population under study.

CONCLUSIONS

Both the APACHE II and APACHE IV scores have an almost equal discrimination ability but have different levels of calibration with the APACHE IV having a PMR closer to the OMR compared to APACHE II. Therefore, there is a need to recalibrate the scoring systems to different populations in different settings to make more meaning out of the PMRs.

REFERENCES

1. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: A physiologically based classification system. Crit Care Med 1981;9:591-7.
2. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, *et al.* The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991;100:1619-36.
3. 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;34:1297-310.
4. Rapsang AG, Shyam DC. Scoring systems in the Intensive Care Unit: A compendium. Indian J Crit Care Med 2014;18:220-8.
5. Ayazoglu TA. Validation of the APACHE IV scoring system in patients with stroke: A comparison with the APACHE II system. Anaesth Pain Intensive Care 2011;15:7-12.
6. Lee H, Shon YJ, Kim H, Paik H, Park HP. 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;67:115-22.
7. Ilker I, Mehmet K, Mehmet A, Aysenur D, Fesih K, Aysenur S, *et al.* Study of effectiveness of the SAPS II-III, APACHE II-IV, and MPM II scores in the determination of prognosis of the patients in reanimation Intensive Care Unit. Acta Med Meditterranea 2015;31:127.
8. Saleh A, Ahmed M, Sultan I, Abdel-Lateif A. Comparison of the mortality prediction of different ICU scoring systems (APACHE II and III, SAPS II, and SOFA) in a single-center ICU subpopulation with acute respiratory distress syndrome. Egypt J Chest Dis Tuberc 2015;64:843-8.
9. Kamal M, Khan AN, 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 Med J 2013;38:234-8.
10. Basu S, Bhattacharya M, Chatterjee TK, Chaudhuri S, Todi SK, Majumdar A, *et al.* Microalbuminuria: A novel biomarker of sepsis. Indian J Crit Care Med 2010;14:22-8.
11. Kakar P, Govil D, Gupta S, Srinivasan S, Mehta P, Prakash O, *et al.* Validation of APACHE IV in patients with severe acute pancreatitis. Crit Care 2008;12 Suppl 2:P500.
12. Parajuli BD, Shrestha GS, Pradhan B, Amatya R. Comparison of acute physiology and chronic health evaluation II and acute physiology and chronic health evaluation IV to predict Intensive Care Unit mortality. Indian J Crit Care Med 2015;19:87-91.
13. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, *et al.* SAPS 3 – From evaluation of the patient to evaluation of the Intensive Care Unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med 2005;31:1345-55.
14. Bosman RJ, Oudemane van Straaten HM, Zandstra DF. The use of intensive care information systems alters outcome prediction. Intensive Care Med 1998;24:953-8.
15. Sirio CA, Shepardson LB, Rotondi AJ, Cooper GS, Angus DC, Harper DL, *et al.* Community-wide assessment of intensive care outcomes using a physiologically based prognostic measure: Implications for critical care delivery from Cleveland Health Quality Choice. Chest 1999;115:793-801.
16. Vincent JL, Moreno R. Clinical review: Scoring systems in the critically ill. Crit Care 2010;14:207.
17. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. Contin Educ Anaesth Crit Care Pain 2008;8:181-5.