

# Predicting Mortality of Patients With Sepsis: A Comparison of APACHE II and APACHE III Scoring Systems

Farid Sadaka<sup>a, b</sup>, Cheikh EthmaneAbouElMaali<sup>a</sup>, Margaret A. Cytron<sup>a</sup>, Kimberly Fowler<sup>a</sup>, Victoria M. Javaux<sup>a</sup>, Jacklyn O'Brien<sup>a</sup>

## Abstract

**Background:** Acute Physiology, Age and Chronic Health Evaluation (APACHE) II and III scores were developed in 1985 and 1991, respectively, and are used mainly for critically ill patients of all disease categories admitted to the intensive care unit (ICU). They differ in how chronic health status is assessed, in the number of physiologic variables included (12 vs. 17), and in the total score. These two scoring systems have not been compared in predicting hospital mortality in patients with sepsis.

**Methods:** We retrospectively identified all septic patients admitted to our 54-bed medical-surgical ICU between June 2009 and February 2014 using the APACHE outcomes database. We calculated correlation coefficients for APACHE II and APACHE III scores in predicting hospital mortality. Receiver-operating characteristic (ROC) curves were also used to assess the mortality predictions.

**Results:** We identified a total of 2,054 septic patients. Average APACHE II score was  $19 \pm 7$ , and average APACHE III score was  $68 \pm 28$ . ICU mortality was 11.8% and hospital mortality was 18.3%. Both APACHE II ( $r = 0.41$ ) and APACHE III scores ( $r = 0.44$ ) had good correlations with hospital mortality. There was no statistically significant difference between the two correlations ( $P = 0.1$ ). ROC area under the curve (AUC) was 0.80 (95% confidence interval (CI): 0.78 - 0.82) for APACHE II, and 0.83 (95% CI: 0.81 - 0.85) for APACHE III, suggesting that both scores have very good discriminative powers for predicting hospital mortality.

**Conclusions:** This study shows that both APACHE II and APACHE III scores in septic patients were very strong predictors of hospital mortality. APACHE II was as good as APACHE III in predicting hospital mortality in septic patients.

**Keywords:** Sepsis; Epidemiology; Acute physiologic and chronic health evaluation; APACHE; Mortality; Outcome

## Introduction

Sepsis is a life-threatening condition that affects more than 1 million patients a year in the United States and even more patients around the globe and is one of the leading causes of death. New definitions for sepsis and septic shock (sepsis-3) were published [1]. Sepsis is now defined as life-threatening organ dysfunction caused by a dysregulated host response to infection. Septic shock is a subset of sepsis with circulatory and cellular/metabolic dysfunction associated with a higher risk of mortality. Mortality from septic shock in the intensive care unit (ICU) is estimated to range between 45% and 63% in observational studies [1-7]. Many different scoring systems are used to analyze the severity of disease and predict mortality in patients admitted to the ICU including patients with sepsis, such as the Acute Physiology, Age and Chronic Health Evaluation (APACHE), the simplified acute physiology score (SAPS), and the mortality probability model (MPM) [8-14]. Predictive scoring systems are measures of disease severity that are used to predict outcomes, typically mortality, of patients in the ICU. They are helpful for standardizing research and comparing the quality of patient care across ICUs. Scoring systems are typically developed using prospectively collected data from a large number of patients from several ICUs. The individual components and the point in time for data collection vary among the predictive scoring systems, which can lead to differences in efficacy and prediction validity. No single instrument has been adequately proven to be superior to another in its ability to predict death. When choosing a predictive scoring system, it is important to use a score that was appropriately validated and that accurately predicts the outcomes in the population of interest (i.e. sepsis). Other factors that should be taken into consideration include feasibility, ease of use, and cost. APACHE II and III scores were developed in 1985 and 1991, respectively, and are used mainly for critically ill patients of all disease categories admitted to the ICU. They differ in how chronic health status is assessed, in the number of physiologic variables included (12 vs. 17), and in the total score. Compared with the APACHE II, the APACHE III is more complex and time-consuming. APACHE II and APACHE III have been compared in patients with liver cirrhosis [15], acute head injury [16], patients undergoing surgery [17], patients with respiratory failure [18], and in all ICU patients [19]. These two scoring systems have not been compared in predicting hospital

Manuscript submitted July 8, 2017, accepted July 28, 2017

<sup>a</sup>Mercy Hospital St Louis, St Louis University, St. Louis, MO, USA

<sup>b</sup>Corresponding Author: Farid Sadaka, 625 S. New Ballas Rd, Suite 7020, St. Louis, MO 63141, USA. Email: farid.sadaka@mercy.net

doi: <https://doi.org/10.14740/jocmr3083w>

**Table 1.** Comparison of APACHE II and APACHE III Scoring Systems in Patients With Sepsis

|  | APACHE II          | APACHE III         |
|--|--------------------|--------------------|
| Number of patients                     | 2,054              | 2,054              |
| Average score (SD)                     | 19 (7)             | 68 (28)            |
| Correlation with hospital mortality, r | 0.41               | 0.44               |
| ROC (95% CI)                           | 0.80 (0.78 - 0.82) | 0.83 (0.81 - 0.85) |

APACHE: Acute Physiology, Age and Chronic Health Evaluation; SD: standard deviation; ROC: receiver-operating characteristic curve; CI: confidence interval.

mortality in patients with sepsis.

## Methods

We retrospectively identified all septic patients admitted to our 54-bed medical-surgical ICU between June 2009 and February 2014 using the APACHE outcomes database. APACHE outcome database is free, web-based offering from Cerner Corporation that provides users the ability to calculate and report on outcomes data based upon the APACHE predictions available in the public domain. The system may then be used either as an on-line calculator to quickly obtain scores and predictions for individual patients or subsets of patients on an as needed basis, or as a severity-adjusted outcomes measurement system for an individual ICU (or group of ICUs) to assess quality of care and identify opportunities for improvement. The diagnosis of sepsis is based on surviving sepsis campaign (SSC) definition [1]. Data were entered using a software program that included computerized pick lists and automated calculation of physiological means and gradients and error checking. Patient identifiers were removed from the database, and informed consent was waived by our institutional review board. Mean, standard deviation and P values were reported for comparisons. We calculated Pearson correlation coefficient (PCC) for APACHE II

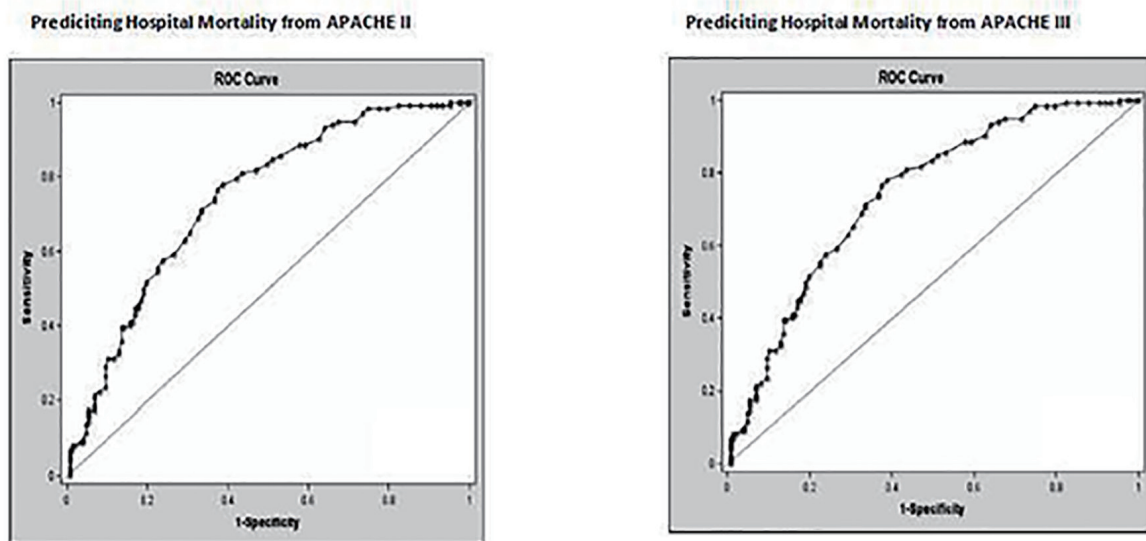
and APACHE III scores in predicting hospital mortality. Receiver-operating characteristic (ROC) curves were also used to assess the mortality predictions.

## Results

We identified a total of 2,054 septic patients. Average APACHE II score was  $19 \pm 7$ , and average APACHE III score was  $68 \pm 28$ . ICU mortality was 11.8% and hospital mortality was 18.3%. Both APACHE II ( $r = 0.41$ ) and APACHE III scores ( $r = 0.44$ ) had good correlations with hospital mortality. There was no statistically significant difference between the two correlations ( $P = 0.1$ ). ROC area under the curve (AUC) was 0.80 (95% confidence interval (CI): 0.78 - 0.82) for APACHE II, and 0.83 (95% CI: 0.81 - 0.85) for APACHE III, suggesting that both scores have very good discriminative powers for predicting hospital mortality (Table 1, Fig. 1).

## Discussion

This study shows that both APACHE II and APACHE III scores in septic patients were very strong predictors of hospital mortality. APACHE II was as good as APACHE III in predict-



**Figure 1.** Receiver operating characteristic curves for APACHE II score and APACHE III score for prediction of hospital mortality.

ing hospital mortality in septic patients. In 200 patients with acute head injury, APACHE II and III were compared to Glasgow coma scale (GCS). APACHE III yielded the best results of correct prediction of functional outcome, but not mortality [16]. In another study, APACHE II was compared to APACHE III in 844 consecutive patients in the surgical ICU. Both scoring systems overestimated mortality, but estimations made by APACHE III were significantly ( $P < 0.01$ ) higher. The authors concluded that in situations where APACHE II underestimated mortality, APACHE III may be corrective. However, the differences were subtle and may be difficult to detect in smaller studies [17]. This study addressed postoperative patients in a surgical ICU, a patient population that could differ significantly from a sicker septic population like our study. In addition, this included a much smaller number of patients than our study, which could increase possibility of type I error. Another retrospective study compared the capability of the APACHE II and APACHE III scoring systems to predict outcome for patients with respiratory failure in a medical ICU. APACHE III had a higher discriminative power (area 0.7462) than APACHE II (area 0.6856;  $P < 0.05$ ) for predicting in-hospital mortality [18]. According to the authors, the extent of physiologic derangement in these patients is likely to have been less marked at the time of ICU admission, which may explain why the predicted mortality rate calculated by the APACHE II model was lower than the actual mortality rate (lead-time bias). In addition, patients with isolated respiratory organ failure are different than septic patients who may have multiple organ failures and/or compromise, like our patient population. A multicenter, multinational study comparing the severity of illness in ICU patients showed that APACHE III had a better ROC area than APACHE II. In 12,899 patients, the improvement in the ROC curve area of APACHE III compared with APACHE II was highly significant ( $P < 0.0001$ ). APACHE II and APACHE III scores were highly correlated ( $r^2 = 0.827$ ) [19]. The area under the ROC curve provides an adequate measure to compare discrimination for each scoring system [20]. Our results demonstrated that APACHE III was not better than APACHE II in predicting hospital mortality, although the discriminative power of APACHE III may be increased by the inclusion of more physiologic variables. The possible reasons behind our results could be multifactorial. For instance, the variables used to assess comorbid conditions in APACHE III and in APACHE II are similar. Even though APACHE III has more independent variables than the APACHE II, the common variables in APACHE II and III systems are mean arterial blood pressure (MBP), respiratory rate, arterial oxygen pressure ( $\text{PaO}_2$ ), oxygen gradient between alveoli and artery ( $\text{PA-aO}_2$ ), and serum creatinine concentration (for acute renal failure in APACHE III). These variables may account for most prediction of outcome in patients with sepsis. This may explain why APACHE II and APACHE III were not different in our study. In addition, many other factors not accounted for in APACHE II or APACHE III can influence the assessment of outcome in septic patients, including case mix, severity of illness, number of progressive organ failures, quality of care, source of ICU admission (emergency room, hospital floor, and another hospital) [21], as well as other potential biases, such as lead-time bias [22].

Our study has several limitations. It is a retrospective and a single center study. However, it encompasses a large cohort of patients, and the outcomes of septic patients in our study are similar to other larger multicenter studies [2-7]. Pearson correlation, used to compare APACHE II and APACHE III in our study, also has its limitations. It assumes a linear relationship which may not be the case in our study. In addition it is unduly affected by extreme values, and a high degree of correlation (like our study) does not necessarily mean a very close relationship. APACHE II and APACHE III systems have their own limitations. Accurate discrimination and calibration are two key characteristics that should be met by all predictive scoring systems. Discrimination describes the accuracy of a given prediction, in our case, its ability to discriminate between survivors and non-survivors. Calibration describes how the score performs over a wide range of predicted mortalities (i.e., the agreement between observed and expected numbers of survivors and non-survivors across all probabilities of death). APACHE systems (II and III) calibration can deteriorate over time largely due to the effects of case-mix and different patient interventions, which often results in an overestimation of mortality, and thus poor discrimination. APACHE systems were developed from, and validated in patients admitted to ICUs across many institutions and therefore may not accurately assess outcomes associated with specific diseases, such as sepsis. In addition, the degree of agreement likely varies substantially by population, therefore point estimates of agreement in terms of generalizability of our findings must be interpreted with caution especially because it is a retrospective single center study. Our study was performed in an academic referral hospital; therefore our results may not be applicable to institutions with different patient populations.

## Conclusion

APACHE III is more complex, has more variables, and data collection is more time consuming than APACHE II. This study shows that both APACHE II and APACHE III scores in septic patients were very strong predictors of hospital mortality. APACHE II was as good as APACHE III in predicting hospital mortality in septic patients. APACHE III is more burdensome, dependent upon proprietary software, and is more costly. On the other hand, APACHE II is easier to use, and is cheaper (free on line calculators).

## Disclosures

This study was not sponsored. None of the authors have anything to disclose.

## References

1. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, Kumar A, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Crit Care Med.* 2017;45(3):486-552.

2. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med*. 2003;348(16):1546-1554.
3. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med*. 2001;29(7):1303-1310.
4. Annane D, Aegerter P, Jars-Guincestre MC, Guidet B, Network CU-R. Current epidemiology of septic shock: the CUB-Rea Network. *Am J Respir Crit Care Med*. 2003;168(2):165-172.
5. Pittet D, Rangel-Frausto S, Li N, Tarara D, Costigan M, Rempe L, Jebson P, et al. Systemic inflammatory response syndrome, sepsis, severe sepsis and septic shock: incidence, morbidities and outcomes in surgical ICU patients. *Intensive Care Med*. 1995;21(4):302-309.
6. Quenot JP, Binquet C, Kara F, Martinet O, Ganster F, Navellou JC, Castelain V, et al. The epidemiology of septic shock in French intensive care units: the prospective multicenter cohort EPISS study. *Crit Care*. 2013;17(2):R65.
7. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):762-774.
8. 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(8):591-597.
9. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-829.
10. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100(6):1619-1636.
11. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, Mercier P, et al. A simplified acute physiology score for ICU patients. *Crit Care Med*. 1984;12(11):975-977.
12. 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(24):2957-2963.
13. Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS. A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med*. 1985;13(7):519-525.
14. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA*. 1993;270(20):2478-2486.
15. Zauner CA, Apsner RC, Kranz A, Kramer L, Madl C, Schneider B, Schneeweiss B, et al. Outcome prediction for patients with cirrhosis of the liver in a medical ICU: a comparison of the APACHE scores and liver-specific scoringsystems. *Intensive Care Med*. 1996;22(6):559-563.
16. Cho DY, Wang YC. Comparison of the APACHE III, APACHE II and Glasgow Coma Scale in acute head injury for prediction of mortality and functional outcome. *Intensive Care Med*. 1997;23(1):77-84.
17. Barie PS, Hydo LJ, Fischer E. Comparison of APACHE II and III scoring systems for mortality prediction in critical surgical illness. *Arch Surg*. 1995;130(1):77-82.
18. Hsu CW, Wann SR, Chiang HT, Lin CH, Kung MH, Lin SL. Comparison of the APACHE II and APACHE III scoring systems in patients with respiratory failure in a medical intensive care unit. *J Formos Med Assoc*. 2001;100(7):437-442.
19. Castella X, Artigas A, Bion J, Kari A. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. The European/North American Severity Study Group. *Crit Care Med*. 1995;23(8):1327-1335.
20. Suter P, Armaganidis A, Beaufil F. Consensus conference organized by the ESICM and the SRLE. Predicting outcome in ICU patients. *Intensive Care Med*. 1994;20:390-397.
21. Borlase BC, Baxter JK, Kenney PR, Forse RA, Benotti PN, Blackburn GL. Elective intrahospital admissions versus acute interhospital transfers to a surgical intensive care unit: cost and outcome prediction. *J Trauma*. 1991;31(7):915-918; discussion 918-919.
22. Dragsted L, Jorgensen J, Jensen NH, Bonsing E, Jacobsen E, Knaus WA, Qvist J. Interhospital comparisons of patient outcome from intensive care: importance of lead-time bias. *Crit Care Med*. 1989;17(5):418-422.