Rui P. Moreno[1], Antonio Paulo Nassar Jr.[2]

Is APACHE II a useful tool for clinical research?

O APACHE II é uma ferramenta útil para pesquisa clínica?

1. Hospital de São José, Centro Hospitalar de Lisboa Central - Lisboa, Portugal.
2. Intensive Care Unit, A.C. Camargo Cancer Center - São Paulo (SP), Brazil.

The population of patients admitted to the intensive care unit (ICU) is quite heterogeneous. Overall, the outcome of ICU treatment depends on the site, cause of admission, age, prior comorbidities, and acute physiological changes at admission and during the first several hours of treatment. Predictions of the in-hospital mortality of ICU patients play important roles with respect to inclusion/exclusion criteria in clinical trials, comparisons of observed mortality with predicted mortality using a score, and estimations of standardized mortality ratios in populations of critical patients. The need for such predictions has led many researchers to develop equations to calculate probabilities of associated mortality. Although prognostic scores have been used since the 1950s (such as the Apgar[1] score for newborns, which was developed by Virginia Apgar), their use for critically ill patients was established only in 1985, when Knaus et al. published the second version of the Acute Physiology and Chronic Health Evaluation (APACHE II),[2] which quickly became the most widely used prognostic index in ICUs and clinical trials worldwide. The original description of APACHE II is the most cited study in the intensive medicine literature to date.[3]

The ability of a prognostic index to predict an outcome (in this case, in-hospital mortality) is assessed based on its calibration and discrimination. Calibration refers to the correspondence between expected mortality predicted using the index and observed mortality in the examined population. Typically, calibration is evaluated by comparing observed and predicted mortality in given predicted risk groups (e.g., deciles, which are used in the Hosmer-Lemeshow test).[4] The calibration of a prognostic model generally deteriorates over time due to changes in ICU admission and discharge criteria, the evolution of support, and variations in the availability and effectiveness of different treatments for particular conditions. Thus, technological and scientific developments in intensive medicine over the last 30 years have rendered APACHE II obsolete. At present, this model generally overestimates mortality in many scenarios in which it is applied. Subsequent versions of this model, such as the most recent variant, APACHE IV,[5] correct this problem, at least in part. As described by Soares et al.,[6] APACHE II should not be used as a benchmarking tool in the ICU because almost any ICU today would be considered "high performance" based on having hospital mortality much lower than that expected in 1985.

Conflicts of interest: None.

Submitted on February 21, 2017
Accepted on February 22, 2017

Corresponding author:
Antonio Paulo Nassar Jr.
A.C. Camargo Cancer Center
Rua Prof. Antonio Prudente, 211 - Liberdade
Zip code: 01509-010 - São Paulo (SP), Brazil
E-mail: paulo_nassar@yahoo.com.br

Responsible editor: Jorge Ibrain Figueira Salluh

DOI: 10.5935/0103-507X.20170046

In contrast, discrimination refers to the ability of a prognostic index to differentiate between patients who survive and patients who die. This metric is evaluated based on the area under a receiver operating characteristic (ROC) curve,[7] with a larger area indicative of greater accuracy (provided, of course, that the area is greater than 0.5, the value at which discrimination is no better than chance) (Table 1).

Table 1 - Discrimination capacity of a prognostic index based on the area under the corresponding receiver operating characteristic curve

Discrimination	Area under the curve
Excellent	0.90 - 0.99
Very good	0.80 - 0.89
Good	0.70 - 0.79
Moderate	0.60 - 0.69
Poor	< 0.60

Although the calibration of APACHE II has deteriorated over time, a MEDLINE search of studies from the prior 2 years that assessed the performance of this index shows that, overall, it continues to exhibit good or very good discrimination in the various populations in which it has been evaluated (Table 2).[8-15] That is, higher APACHE II scores were associated with greater hospital mortality in the examined groups of subjects.

In addition to the heterogeneity of the patient population admitted to the ICU, another consideration is that intensive medicine encompasses syndromes with equally broad spectra of presentation, such as sepsis, acute respiratory distress syndrome, delirium, and postoperative care for major surgeries. Thus, a method to measure the severity of all of these patients is required. This need is especially apparent in clinical studies, which must include a representative population sample to ensure that their findings can be extrapolated to clinical practice. APACHE II was the first index to indicate or contraindicate the use of a certain therapy (in particular, activated protein C in sepsis);[16] the treatment in question was eventually determined to be inappropriate and detrimental.[17] Another therapeutic intervention, the use of low doses of corticosteroids in sepsis, proved beneficial in a study that included patients with greater severity (in this case, another index was used: the Simplified Acute Physiology Score - SAPS- II)[18] but not in another investigation that involved less severely ill patients.[19] This difference in findings led the Sepsis Surviving Campaign to recommend the use of hydrocortisone as an option for septic shock patients who remain unstable after volume expansion and vasopressor use.[20]

Because it continues to exhibit good discrimination capacity, APACHE II remains a widely used index to describe severity in populations of critically ill participants in clinical trials. In 2016, 12 clinical trials involving critically ill patients were published in the 3 highest-impact medical journals.[21-32] APACHE II was the index that was most frequently utilized to describe the severity of the patients included in these studies; this index appeared in 9 of these 12 studies (Figure 1).

One recurring criticism of APACHE II and its subsequent versions is that these indices have been developed from an exclusively North American database. This fact introduces a large bias due to region-specific differences in the availability of different technologies[33] and in patient characteristics;[34] modifications to the

Table 2 - Studies from the prior 3 years that evaluated the performance of APACHE II

Study	Country	Condition	Number of patients	AUC (95% CI)
Pérez Campos et al.[8]	Peru	Acute pancreatitis	334	0.85 (0.77 - 0.94)
Serpa Neto et al.[9]	Brazil	General ICU	3,333	0.80 (0.77 - 0.83)
Que et al.[10]	Switzerland	Severe sepsis/septic shock	Development (Switzerland): 158	0.64 (0.54 - 0.73)
	Brazil		Validation (Brazil): 91	0.64 (0.52 - 0.75)
Ariyaratnam et al.[11]	United Kingdom	Post-operative care for cardiac surgery	1,646	0.65 (0.56 - 0.74)
Williams et al.[12]	Australia	Admission from the emergency room for suspected infection	8,871	0.90 (0.88 - 0.91)
Hashmi et al.[13]	Pakistan	General ICU	213	0.83 (0.77 - 0.88)
Khwannimit et al.[14]	Thailand	Sepsis	913	0.91 (0.89 - 0.93)
Huang et al.[15]	China	Severe ARDS on ECMO	23	0.76 (0.56 - 0.96)

AUC - area under the curve; 95% CI - 95% confidence interval; ICU - intensive care unit; ARDS - acute respiratory distress syndrome; ECMO - extracorporeal membrane oxygenation.

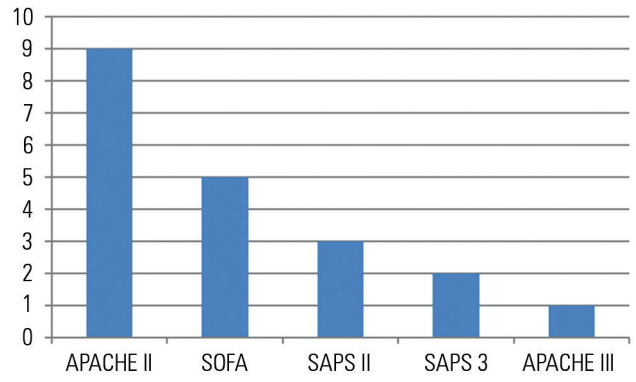

Figure 1 - Numbers of trials in which various prognostic indices were used to describe patient severity, out of 12 clinical trials performed in intensive care units and published in the New England Journal of Medicine, The Lancet, or JAMA in 2016.[21-32] These numbers sum to more than 12 because certain studies involved the use of multiple prognostic indices. APACHE - Acute Physiology and Chronic Health Evaluation; SOFA - Sequential Organ Failure Assessment; SAPS - Simplified Acute Physiology Score.

equations used for these indices cannot fully correct for this bias.[35] Today, other scores are better calibrated and should be used to assess predicted mortality[36] to provide ways to express the severity of patients included in clinical trials.

However, because APACHE II continues to perform well in determining severity for a group of patients (although it cannot and should not be used to assess individual patients), its use in clinical research may be justified, in contrast to its use in the assessment of ICU performance or the prognostic evaluation of patient groups. In the latter contexts, APACHE II should return to libraries and merits respect only for having pioneered the field of prognostic evaluation in the ICU.

REFERENCES

1. Apgar V. A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg. 1953;32(4):260-7.
2. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818-29.
3. Baltussen A, Kindler CH. Citation classics in critical care medicine. Intensive Care Med. 2004;30(5):902-10.
4. Hosmer DW, Lemeshow S. A goodness-of-fit test for the multiple logistic regression model. Commun Stat. 1980;9(10):1043-69.
5. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006;34(5):1297-310.
6. Soares M, Dongelmans DA. Why should we not use APACHE II for performance measurement and benchmarking? Rev Bras Ter Intensiva. 2017;29(3):268-70.
7. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
8. Pérez Campos A, Bravo Paredes E, Prochazka Zarate R, Bussalleu A, Pinto Valdivia J, Valenzuela Granados V. [BISAP-O y APACHE-O: utility in predicting severity in acute pancreatitis in modified Atlanta classification]. Rev Gastroenterol Peru. 2015;35(1):15-24. Spanish.
9. Serpa Neto A, Assunção MS, Pardini A, Silva E. Feasibility of transitioning from APACHE II to SAPS III as prognostic model in a Brazilian general intensive care unit. A retrospective study. Sao Paulo Med J. 2015;133(3):199-205.
10. Que YA, Guessous I, Dupuis-Lozeron E, Alves de Oliveira CR, Ferreira Oliveir C, Graf R, et al. Prognostication of mortality in critically ill patients with severe infections. Chest. 2015;148(3):674-82.
11. Ariyaratnam P, Loubani M, Biddulph J, Moore J, Richards N, Chaudhry M, et al. Validation of the Intensive Care National Audit and Research Centre Scoring System in a UK Adult Cardiac Surgery Population. J Cardiothorac Vasc Anesth. 2015;29(3):565-9.
12. Williams JM, Greenslade JH, Chu K, Brown AF, Lipman J. Severity scores in emergency department patients with presumed infection: a prospective validation study. Crit Care Med. 2016;44(3):539-47.
13. Hashmi M, Asghar A, Shamim F, Khan FH. Validation of acute physiologic and chronic health evaluation II scoring system software developed at The Aga Khan University, Pakistan. Saudi J Anaesth. 2016;10(1):45-9.
14. Khwannimit B, Bhurayanontachai R, Vattanavanit V. Validation of the sepsis severity score compared with updated severity scores in predicting hospital mortality in sepsis patients. Shock. 2017;47(6):720-5.
15. Huang L, Li T, Xu L, Hu XM, Duan DW, Li ZB, et al. Performance of Multiple Risk Assessment Tools to Predict Mortality for Adult Respiratory Distress Syndrome with Extracorporeal Membrane Oxygenation Therapy: An External Validation Study Based on Chinese Single-center Data. Chin Med J (Engl). 2016;129(14):1688-95.
16. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, Steingrub JS, Garber GE, Helterbrand JD, Ely EW, Fisher CJ Jr; Recombinant human protein C Worldwide Evaluation in Severe Sepsis (PROWESS) study group. Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med. 2001;344(10):699-709.
17. Ranieri VM, Thompson BT, Barie PS, Dhainaut JF, Douglas IS, Finfer S, Gårdlund B, Marshall JC, Rhodes A, Artigas A, Payen D, Tenhunen J, Al-Khalidi HR, Thompson V, Janes J, Macias WL, Vangerow B, Williams MD; PROWESS-SHOCK Study Group. Drotrecogin alfa (activated) in adults with septic shock. N Engl J Med. 2012;366(22):2055-64.
18. Annane D, Sébille V, Charpentier C, Bollaert PE, Francois B, Korach JM, et al. Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. JAMA. 2002;288(7):862-71.
19. Sprung CL, Annane D, Keh D, Moreno R, Singer M, Freivogel K, Weiss YG, Benbenishty J, Kalenka A, Forst H, Laterre PF, Reinhart K, Cuthbertson BH, Payen D, Briegel J; CORTICUS Study Group. Hydrocortisone therapy for patients with septic shock. N Engl J Med. 2008;358(2):111-24.
20. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med. 2017;43(3):304-77.
21. Gaudry S, Hajage D, Schortgen F, Martin-Lefevre L, Pons B, Boulet E, Boyer A, Chevrel G, Lerolle N, Carpentier D, de Prost N, Lautrette A, Bretagnol A, Mayaux J, Nseir S, Megarbane B, Thirion M, Forel JM, Maizel J, Yonis H, Markowicz P, Thiery G, Tubach F, Ricard JD, Dreyfuss D; AKIKI Study Group. Initiation strategies for renal-replacement therapy in the intensive care unit. N Engl J Med. 2016;375(2):122-33.

22. Gordon AC, Perkins GD, Singer M, McAuley DF, Orme RM, Santhakumaran S, et al. Levosimendan for the prevention of acute organ dysfunction in sepsis. N Engl J Med. 2016;375(17):1638-48.
23. Legriel S, Lemiale V, Schenck M, Chelly J, Laurent V, Daviaud F, Srairi M, Hamdi A, Geri G, Rossignol T, Hilly-Ginoux J, Boisramé-Helms J, Louart B, Malissin I, Mongardon N, Planquette B, Thirion M, Merceron S, Canet E, Pico F, Tran-Dinh YR, Bedos JP, Azoulay E, Resche-Rigon M, Cariou A; HYBERNATUS Study Group. Hypothermia for neuroprotection in convulsive status epilepticus. N Engl J Med. 2016;375(25):2457-67.
24. Schaller SJ, Anstey M, Blobner M, Edrich T, Grabitz SD, Gradwohl-Matis Heim M, Houle T, Kurth T, Latronico N, Lee J, Meyer MJ, Peponis T, Talmor D, Velmahos GC, Waak K, Walz JM, Zafonte R, Eikermann M; International Early SOMS-guided Mobilization Research Initiative. Early, goal-directed mobilisation in the surgical intensive care unit: a randomised controlled trial. Lancet. 2016;388(10052):1377-88.
25. Su X, Meng ZT, Wu XH, Cui F, Li HL, Wang DX, et al. Dexmedetomidine for prevention of delirium in elderly patients after non-cardiac surgery: a randomised, double-blind, placebo-controlled trial. Lancet. 2016;388(10054):1893-902.
26. Walsh TS, Kydonaki K, Antonelli J, Stephen J, Lee RJ, Everingham K, Hanley J, Phillips EC, Uutela K, Peltola P, Cole S, Quasim T, Ruddy J, McDougall M, Davidson A, Rutherford J, Richards J, Weir CJ; Development and Evaluation of Strategies to Improve Sedation Practice in Intensive Care (DESIST) study investigators. Staff education, regular sedation and analgesia quality feedback, and a sedation monitoring technology for improving sedation and analgesia quality for critically ill, mechanically ventilated patients: a cluster randomised trial. Lancet Respir Med. 2016;4(10):807-17.
27. Girardis M, Busani S, Damiani E, Donati A, Rinaldi L, Marudi A, et al. Effect of conservative vs conventional oxygen therapy on mortality among patients in an intensive care unit: the oxygen-ICU randomized clinical trial. JAMA. 2016;316(15):1583-9.
28. Gordon AC, Mason AJ, Thirunavukkarasu N, Perkins GD, Cecconi M, Cepkova M, Pogson DG, Aya HD, Anjum A, Frazier GJ, Santhakumaran S, Ashby D, Brett SJ; VANISH Investigators. Effect of early vasopressin vs norepinephrine on kidney failure in patients with septic shock: the VANISH randomized clinical trial. JAMA. 2016;316(5):509-18.
29. Hernández G, Vaquero C, Colinas L, Cuena R, González P, Canabal A, et al. Effect of postextubation high-flow nasal cannula vs noninvasive ventilation on reintubation and postextubation respiratory failure in high-risk patients: a randomized clinical trial. JAMA. 2016;316(15):1565-74.
30. Keh D, Trips E, Marx G, Wirtz SP, Abduljawwad E, Bercker S, Bogatsch H, Briegel J, Engel C, Gerlach H, Goldmann A, Kuhn SO, Hüter L, Meier-Hellmann A, Nierhaus A, Kluge S, Lehmke J, Loeffler M, Oppert M, Resener K, Schädler D, Schuerholz T, Simon P, Weiler N, Weyland A, Reinhart K, Brunkhorst FM; SepNet-Critical Care Trials Group. Effect of Hydrocortisone on development of shock among patients with severe sepsis: the HYPRESS randomized clinical trial. JAMA. 2016;316(17):1775-85.
31. Patel BK, Wolfe KS, Pohlman AS, Hall JB, Kress JP. Effect of noninvasive ventilation delivered by helmet vs face mask on the rate of endotracheal intubation in patients with acute respiratory distress syndrome: a randomized clinical trial. JAMA. 2016;315(22):2435-41.
32. Reade MC, Eastwood GM, Bellomo R, Bailey M, Bersten A, Cheung B, Davies A, Delaney A, Ghosh A, van Haren F, Harley N, Knight D, McGuiness S, Mulder J, O'Donoghue S, Simpson N, Young P; DahLIA Investigators; Australian and New Zealand Intensive Care Society ClinicalTrials Group. Effect of dexmedetomidine added to standard care on ventilator-free time in patients with agitated delirium: a randomized clinical trial. JAMA. 2016;315(14):1460-8.
33. Bastos PG, Knaus WA, Zimmerman JE, Magalhães A Jr, Sun X, Wagner DP. The importance of technology for achieving superior outcomes from intensive care. Brazil APACHE III Study Group. Intensive Care Med. 1996;22(7):664-9.
34. Moreno R, Apolone G, Miranda DR. Evaluation of the uniformity of fit of general outcome prediction models. Intensive Care Med. 1998;24(1):40-7.
35. Moreno R, Apolone G, Fidler V, Reis Miranda D. The impact of first level customization on the uniformity of fit of SAPS II [abstract]. Intensive Care Med. 1996;22(Suppl 3):S267.
36. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR; SAPS 3 Investigators. 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(10):1345-55. Erratum in: Intensive Care Med. 2006;32(5):796.