Guessing Game of Patient Outcomes in the Renally Injured Critically Ill: Is There a Perfect Score?

Gautham M Raju

Dealing with a critically ill patient can be a challenge in terms of evaluation, assessment, categorization of disease pathology and severity, tailoring optimal treatment strategies and analysis of outcome (both individual and composite).

Acute kidney injury is a significant problem which adversely affects morbidity and mortality both independently or as part of multiorgan dysfunction. Over 30–60% of critically ill patients suffer from acute kidney injury due to various reasons common being sepsis, and associated mortality can range from 15 to 65%.[1]

Acute kidney injury is also associated with a higher risk of adverse cardiovascular events, other serious life-threatening complications, and associated negative outcomes like mortality and morbidity.[2,3]

There are various scoring systems to measure and assess to help indicate outcomes.

There are:

- **General risk prognostication scores** like Acute Physiology and Chronic Health Evaluation (APACHE II) score, Simplified Acute Physiology Score (SAPS II), mortality prediction model 3 score (MPM III), and
- **Organ-specific risk prognostication scores** like Glasgow Coma Score (GCS), Acute Kidney Injury Network (AKIN), Risk, Injury, Failure, Loss of function, End-stage renal disease (RIFLE) score, Kidney Disease: Improving Global Outcomes (KDIGO) score, organ dysfunction scores like `Sequential Organ Failure Assessment´ (SOFA) score, etc.

Among them APACHE II score has been widely used for prognostication among critically ill, specifically to predict mortality. APACHE II score was derived from datasets of patients in ICU's in North America led by Knaus et al. The score can range from 0 to 71 with weightage points assigned to age, acute physiological conditions and specific preexisting chronic disease.[4]

Scoring systems related to acute kidney injury *per se*, the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines 2012 proposed a modified version of the RIFLE and AKIN criteria.[5] The modified definition for acute kidney injury is based on serum creatinine and urine output. Despite potential limitations, the KDIGO classification has been accepted as the consensus standard for use in clinical practice to define acute kidney injury and associated outcomes. The KDIGO classification also gives an idea about the mortality corresponding to each of its class of acute kidney injury.

In this issue of IJCCM, an observational study by Patel et al. assessed whether APACHE II scores have sufficient sensitivity and specificity in predicting outcomes associated with acute kidney

Department of Critical Care, St John's Medical College and Hospital, Bengaluru, Karnataka, India

Corresponding Author: Gautham M Raju, Department of Critical Care, St John's Medical College and Hospital, Bengaluru, Karnataka, India, Phone: +91 9793004589, e-mail: drgauraju@gmail.com

injury when classified as per KDIGO guidelines. One-hundred patients with acute kidney injury admitted to intensive care unit (ICU) were studied in a tertiary care hospital in Haryana in North India over a period of 1 year. Multiple parameters (age, lab parameters, chronic) were studied and APACHE II score calculated and corresponding mortality noted. Similarly, AKI among patients were also classified as per KDIGO classification and corresponding mortality figures were noted. In this study, the mean APACHE II scores among patients who expired were significantly higher than those who survived, across the various stages of acute kidney injury. It was also noted sepsis (47%) was the most common cause of acute kidney injury in the critically ill. This is similar to that observed by other studies as well.[6,7] The authors have also noted higher mortality rates in patients who require ventilator and vasopressor support, and those with coexisting coronary artery disease.

In their study the authors have observed that in predicting mortality, APACHE II score has a sensitivity (95% CI) of 57.14% (39.4–73.7%), and specificity (95% CI) of 86.15% (75.3–93.5%), positive predictive value (95% CI) of 69% (49.2–84.7%), negative predictive value (95% CI) of 78.9% (67.6–87.7%) with area under ROC (AUROC) of 0.79.

While the overall the mortality observed closely corresponds to what is arrived at by correlating it with the AKI as per KDIGO, the APACHE II score does not directly correlate to the mortality figures as observed with various KDIGO classes.

Authors have observed good discrimination and calibration of the test; however, such an inference may not be as straightforward as it seems. It calls for a deeper understanding of discrimination and calibration. Discrimination and calibration are applied to clinical predictive models or scoring systems: where a number of factors (clinical, laboratory, age, acute or chronic health conditions) are considered and provide an estimate of patients absolute risk of an

outcome (ex: mortality risk in patients with acute kidney injury) or the likelihood of a particular diagnosis.

DISCRIMINATION AND CALIBRATION

"Discrimination" refers to how well the model or the prognostic scoring system can distinguish between those at higher risk of having a particular outcome (e.g. in this study mortality) from those with lower risk. For example, a prognostic score could discriminate well between mortality among patients with and without acute kidney injury, among heterogeneous population with diverse factors (clinical, laboratory, etc.) but may not do so in a more homogeneous population with limited number of factors.[8]

While using a prognosticating tool or model, discrimination alone is insufficient. The second most important property of any predictive model gives clues to clinicians how similar the predicted absolute risk estimates are, to the true (observed) risk estimates in patient groups classified according to diverse risk strata. "Calibration" refers to the accuracy of these absolute risk estimates. In any predictive score or model; the more accurate the estimate—the better the calibration.[8]

Measuring Discrimination

It can be measured by various methods, however, when it involves binary outcomes (dead or alive), the receiver operator characteristic (ROC) curve best characterizes discrimination. Greater the area under the ROC curve, better the prediction model/prognostic score. In general, area under the curve approaching 100% is better. On the contrary, ROC curves with areas, less than 50% the events are attributed to chance.[9]

Measuring Calibration

Calibration also known as goodness of fit is a crucial property of a model and indicates the extent to which a predictive model accurately shows the absolute risk (i.e., whether the values indicated as per the model match with the observed values). Poor calibration models will overestimate or underestimate the outcome.

Assessment of calibration involves comparing predicted and observed risk at various levels (a) the population under study; (b) various patient groups based on predicted risk; or (c) different patient groups based on combinations of predictive factors. A good model shows strong calibration for various patient groups with widely varying characteristics.[10]

Considering the above, the authors concluded that the APACHE II score has good discrimination and calibration is only partly true. Higher the APACHE II scores, worser the outcomes. But while applying the same logic in correspondence to estimating mortality from an individual component such as acute kidney injury measured only using only serum creatinine may not be accurate (APACHE II score includes serum creatinine alone). Attempting to correlate the mortality corresponding to various classes of acute kidney injury as per KDIGO may also lead to erroneous inferences as the KDIGO classification uses both changes in serum creatinine and urine output. Also, studies indicate serum creatinine or urine output alone cannot reliably predict acute kidney injury or related outcomes.[11]

Although the current study puts a lot of weight into the prognostic ability of APACHE II score, it should be noted that APACHE II scores are based on data sets of patient characteristics from 1980s to 1990s. Also calculating APACHE II score can vary widely depending on the clinicians knowledge, experience and training. Although web-based or computer-based calculators are available, there have been wide variations in APACHE calculations even among clinicians familiar with using APACHE score in its existing format. Also, APACHE II score does not give estimate of individual mortality.

Current technological advancements make it necessary to revamp and recalibrate the APACHE II with a database, which includes patient characteristics according to current times, specific patient population, demographics and geography for better accuracy of predictive models. Third-generation models of severity scoring systems like APACHE 4, MPM III, and SAPS III have been found to have very better discrimination and calibration.[12]

Among other prognostics scoring systems, SOFA score can be useful in assessment of daily disease progression and can give reliable clues in outcomes among critically ill patients. Although originally SOFA score was not designed to estimate mortality, a recent study by Wang et al. showed SOFA score has better ability to predict prognosis among critically ill patients with acute kidney injury.[13,14]

Trying to simplify a disease process and its outcome based on individual organ's impact on mortality and morbidity can give more insight into the degree a given organ dysfunction plays in the outcome of a patient. But human body works as a syncytium of multiple organ systems interconnected in ways so complex that, oversimplification and extrapolating the role of anyone organ system may not be possible without accounting for the interaction with others.

Bottomline

The purpose of a scoring system is to help comparative audit, assist clinical research and help in the clinical management and prognostication of patients. A good scoring system should have good internal and external validity, and good discrimination and calibration. While the APACHE II score has stood the test of time in terms of mortality estimation albeit its limitations mentioned above; mortality as the most important outcome measure may need a relook. Perhaps disability, chronic complications and need for sustained healthcare support, and quality of life post-acute kidney injury and critical illness may be a better outcome measure in terms of what constitutes a functional recovery. Comprehensive inclusion of population metrics with an emphasis on contemporary data and a practical outcome-based predictive model is needed.

ORCID

Gautham M Raju https://orcid.org/0000-0002-4747-3564

REFERENCES

1. Uchino S, Kellum JA, Bellomo R, Doig GS, Morimatsu H, Morgera S, et al. Acute renal failure in critically ill patients: a multinational, multicenter study. Journal of the American Medical Association 2005;294(7):813–818. DOI: 10.1001/jama.294.7.813.
2. Bouchard J, Acharya A, Cerda J, Maccariello ER, Madarasu RC, Tolwani AJ, et al. A prospective international multicenter study of AKI in the intensive care unit. Clin J Am Soc Nephrol 2015;10(8):1324–1331. DOI: 10.2215/CJN.04360514.
3. Odutayo A, Wong CX, Farkouh M, Altman DG, Hopewell S, Emdin CA, et al. AKI and long-term risk for cardiovascular events and mortality. J Am Soc Nephrol 2017;28(1):377–387. DOI: 10.1681/ASN.2016010105.
4. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985;13(10): 818–829. PMID: 3928249.

5. Kidney Disease Improving Global Outcomes (KDIGO) AKI Work Group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl 2012;2:1–138. DOI:10.1038/kisup.2012.7.

6. Kapadia MP, Kamdar KP, Jha PR. A study of clinical profile of patients with acute kidney injury in a tertiary care centre. IJASR [serial online] 2016;2(8):160–166. DOI: 10.7439/ijasr.v2i8.3511.

7. Chawla LS, Abell L, Mazhari R, Egan M, Kadambi N, Burke HB, et al. Identifying critically ill patients at high risk for developing acute renal failure: a pilot study. Kidney Int 2005;68(5):2274–2280. DOI: 10.1111/j.1523-1755.2005.00686.x.

8. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. Journal of the American Medical Association 2017;318(14):1377–1384. DOI: 10.1001/jama.2017.12126.

9. Pencina MJ, D'Agostino RB Sr. Evaluating discrimination of risk prediction models: the C statistic. Journal of the American Medical Association 2015;314(10):1063–1064. DOI: 10.1001/jama.2015.11082.

10. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167–176. DOI: 10.1016/j.jclinepi.2015.12.005.

11. Chu R, Li C, Wang S, Zou W, Liu G, Yang L, et al. Assessment of KDIGO definitions in patients with histopathologic evidence of acute renal disease. Clin J Am Soc Nephrol 2014;9(7):1175–1182. DOI: 10.2215/CJN.06150613.

12. Costa e Silva VT, de Castro I, Liaño F, Muriel A, Rodríguez-Palomares JR, Yu L. Performance of the third-generation models of severity scoring systems (APACHE IV, SAPS 3 and MPM-III) in acute kidney injury critically ill patients. Nephrol Dial Transplant 2011;26(12):3894–3901. DOI: 10.1093/ndt/gfr201.

13. Wang H, Kang X, Shi Y, Bai ZH, Lv JH, Sun JL, et al. SOFA score is superior to APACHE-II score in predicting the prognosis of critically ill patients with acute kidney injury undergoing continuous renal replacement therapy. Ren Fail 2020;42(1):638–645. DOI: 10.1080/0886022X.2020.1788581.

14. Patel P, Gupta S, Patel H, Bashar MD. A. Assessment of APACHE II Score to Predict ICU Outcomes of Patients with AKI: A Single Center Experience from Haryana, North India. Indian J Crit Care Med 2022;26(3):276–281.