

Acute physiology and chronic health evaluation II for critically ill children?

Jhuma Sankar

Mortality risk prediction models are used to evaluate the risk of mortality in Intensive Care Units (ICUs) at admission or over the course of stay. They allow for inter- and intra-unit comparisons with time, and also provide useful information for comparing the severity of illness of patients enrolled into clinical trials.^[1] The acute physiology and chronic health evaluation (APACHE) was the first scoring system introduced in the year 1981 to analyze disease severity in critically ill adults.^[2] The APACHE score comprised of acute physiological parameters and other clinical information. The basis for including variables in this system was based mostly on expert opinion. However, this system was not validated as it was cumbersome with a large number of variables. Subsequently, in 1985 a new abbreviated version of APACHE known as APACHE II was published. Apart from the physiologic variables, APACHE II also included diagnosis and used a logistic regression equation to compute the probability of death.^[3]

APACHE II is a composite score comprising of 12 physiologic variables, age points, and chronic health points collected within 24 h of admission. Based on the cumulative of these scores mortality is predicted. Since, the introduction of the score, it has been extensively validated worldwide in a variety of ICU setting and patient population such as cardiac, neurosurgical, postoperative, and medical.^[4]

In the pediatric age group, the commonly used mortality risk prediction models that have been extensively used and validated are the Pediatric Risk of Mortality (PRISM) and the Pediatric Index of Mortality (PIM) scores.^[1,5] The choice of using

From:
Department of Pediatrics, All India Institute of Medical Sciences, New Delhi, India

Correspondence:
Dr. Jhuma Sankar, All India Institute of Medical Sciences, New Delhi, India.
E-mail: jhumaji@gmail.com

Access this article online

Website: www.ijccm.org

DOI: 10.4103/0972-5229.162458

Quick Response Code:



one or the other of these scoring systems is dictated by their performance and feasibility. Apart from these two models, there are other models based on organ dysfunction such as Pediatric Multiple Organ Dysfunction Score and Pediatric Logistic Organ Dysfunction scores which predict mortality based on the organ dysfunction occurring during the ICU course.^[6] APACHE II or its newer versions have not been used widely and reported in children so far.

In this issue of the journal, Choudhary *et al.* have published their findings of how well the APACHE II discriminated between survivors and nonsurvivors accurately in 100 critically ill children in their unit.^[7] Previously, the APACHE II score has been validated only in children with severe trauma and therefore, Choudhary *et al.*'s study is a value addition to the use of this model in critically ill children in general.^[7]

The authors found higher scores among nonsurvivors with mortality reaching up to 100% with a score

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

How to cite this article: Sankar J. Acute physiology and chronic health evaluation II for critically ill children?. Indian J Crit Care Med 2015;19:446-8.

of >34. The mean score was 26.11 in nonsurvivors as compared to 16.60 among the survivors. The area under the receiver operating characteristic curve (AU-ROC) was 0.889 suggesting excellent discrimination. The calibration was also found to be good with no difference between expected outcome and observed outcome by both the Hosmer-Lemeshow goodness-of-fit (GOF) Chi-square test and the standardized mortality ratio (SMR). The *P* value for GOF was 0.72 and the SMR was 1.

The performance of any scoring system in a particular unit is assessed by two important statistical methods/ tests. One is the discrimination, which is the ability of a model to distinguish accurately between survivors and nonsurvivors and is usually assessed using AUC of the corresponding ROC curves.^[1,8] An acceptable discrimination is defined as an AUC between 0.70 and 0.79, and good discrimination as ≥ 0.80 . One of the major problems with the AU-ROC plot is that, it does not tell us whether the model predicts mortality well for both the ill and the not-so-ill children. The Hosmer-Lemeshow GOF test was developed to overcome this problem and is a better indicator of how well the score performs across different probabilities of death. This is important than predicting death for an individual patient which is not the primary goal or purpose for which these scoring systems were developed.^[8] Therefore, it is not surprising to find that the confidence intervals for individual patients are always wider than for patient populations whenever these models have been validated. Therefore, these models are best applied to patient populations rather than individuals.

The SMR is another test for calibration like the GOF. Both the GOF *P* values and the SMR are calculated from the same tables. A good calibration is represented by a *P* value of ≥ 0.05 in the Hosmer-Lemeshow test and an SMR close to or equal to 1. Thus, calibration is an actual measure of the performance of the scoring system in the unit in which the score is being validated with that of the unit where it was developed. In other words, it helps in interunit and intraunit comparisons, which are important for quality control and benchmarking of ICU, care worldwide.^[1,8]

Although Choudhary *et al.* found the model to have good discrimination and calibration, it should be borne in mind that the sample size in their study was too small (only 100 patients) and the mortality rates much higher than that of the units where the scores were developed (55%). The *P* value of Hosmer-Lemeshow

GOF test is unreliable with sample sizes of <400.^[8] The excellent calibration found could be simply due to the small sample size. Therefore, their findings need further validation in a larger cohort of patients and preferably from more than one unit.

Both PIM and PRISM have been validated through infancy to adolescents and infants are a substantial proportion of patients admitted to the ICU. In the present study, infants were excluded. Including infants could have affected the AUC as the models have shown to perform less reliably in this sub group. In a previously published study by our group^[9] on PIM and PIM-2 scores, we observed that the discrimination was not as good in infants as it was in other age groups (AUC was 0.64 and 0.67 for infants for PIM and PIM-2 respectively, while it was >0.70 for all other age groups). Therefore, it remains to be seen how reliably APACHE II would discriminate survivors from nonsurvivors in this group of children.

Like any technology that requires upgradation as per user requirements in terms of ease of use as well as applicability, the mortality prediction models also have been updated over periods of time to adjust for the differences in case mix and improvement in quality of care with time. As a result, we have updated models of PIM, PRISM, and APACHE available now and currently the APACHE IV is in use. With the newer versions of this model available, it would only be prudent to use the latest version rather than a previous version to validate the score across units. And it remains to be seen if the latest version of the APACHE, that is, APACHE IV discriminates as well in an adequate sample of critically ill children across all age groups.

References

1. Pollack MM. Clinical scoring systems in pediatric intensive care. In: Fuhrman BP, Zimmerman JJ, editors. *Pediatric Critical Care*. St. Louis: Mosby Year Book; 1992. p. 153-62.
2. 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.
3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. *Crit Care Med* 1985;13:818-29.
4. Hargrove J, Nguyen HB. Bench-to-bedside review: Outcome predictions for critically ill patients in the emergency department. *Crit Care* 2005;9:376-83.
5. Marein JP, Pollack MM. Review of the acuity scoring systems for the Pediatric Intensive Care Unit and their use in quality improvement. *J Intensive Care Med* 2007;22:131-40.
6. Graciano AL, Balko JA, Rahn DS, Ahmad N, Giroir BP. The Pediatric Multiple Organ Dysfunction Score (P-MODS): Development and validation of an objective scale to measure the severity of

- multiple organ dysfunction in critically ill children. *Crit Care Med* 2005;33:1484-91.
7. Choudhary S, Chhangani NP, Amandeep M, Gupta V, Goyal V. Role of APACHE II: Scoring system in determining the severity and prognosis of critically ill patients in PICU. *Indian J Crit Care Med* 2015;19:462-5.
 8. Shann F. Are we doing a good job: PRISM, PIM and all that. *Intensive Care Med* 2002;28:105-7.
 9. Sankar J, Singh A, Sankar MJ, Joghee S, Dewangan S, Dubey N. Pediatric Index of Mortality and PIM-2 scores have good calibration in a large cohort of children from a developing country. *Biomed Res Int* 2014;2014:907871.