Use of Pediatric Index of Mortality (PIM) and the Pediatric Risk of Mortality (PRISM) III-24 for prediction of mortality

To the Editor—I read with interest the article by Choi et al[1] regarding the use of the Pediatric Index of Mortality (PIM) and the Pediatric Risk of Mortality (PRISM) III-24 score for predicting mortality in a paediatric intensive care unit (PICU) in Hong Kong. I agree with the authors that mortality prediction models are useful for audit purposes and allow comparison of disease severity and clinical outcomes in different PICUs to be made. Nonetheless proper validation of such systems is essential.

I have two comments. First, readers should be given additional information about those patients who were transferred from their PICU for specialist care elsewhere, including their scores, diagnoses, and outcomes. Patients who are transferred because of the need for advanced care often have complicated clinical problems outside the management expertise of the transferring PICU. Omitting such cases may affect the validity of any conclusions drawn in the study. Second, probable statistical errors in the validation of the models in the study may render the conclusion invalid.

Validation of a mortality-scoring model is vital when assessing its ability to predict one of the many important outcome measures, namely death. The validation needs both discrimination and calibration: a poorly fitting model may have good discrimination but poor calibration. Discrimination is assessed by the receiver operating characteristics (ROC) plot that estimates how well the model distinguishes between patients who live and those who die.[2-4] Usually, an area under the ROC curve (AUC) of 0.75 or above is considered clinically useful. In the study, the AUC was 0.910 and 0.912 for PRISM III-24 and PIM, respectively, indicating that both systems had good discriminating power. Nonetheless in terms of calibration, the article statistics were flawed. Both models are logistic regression models and Hosmer-Lemeshow goodness-of-fit test is recommended for their calibration.[2,3,5] In the Hosmer-Lemeshow goodness-of-fit test,[6] study subjects are divided into 10 groups. There are two grouping strategies: (1) based on percentiles of the estimated probabilities, and (2) based on fixed values of the estimated probability. The first method is preferable to the second as it adheres better to the Chi squared distribution, especially when many of the estimated probabilities are small (ie <20%)—this was not the case in this study (291 of 303 patients belonged to the group 0-25%). For method 1, subjects are first arranged in ascending order of expected mortality and evenly divided into 10 groups from low to high mortality. These groups are often referred to as the "deciles of risk", and Chi squared statistical analysis is then applied. The appropriateness of the P value depends on the validity of the assumption that the estimated expected frequencies of both survival and death are large, preferably not less than five. If not, selected adjacent rows of the table may be combined to increase the size of the expected frequencies but not to the extent of fewer than six groups. Hence to fulfill this statistical assumption, the number of expected deaths must be at least 30, much greater than that in the study. In the study, the authors used the second method and divided patients according to fixed endpoints into only four groups that were inadequate for meaningful analysis. Even if these groupings are accepted, the expected death frequencies were far below five in three of the four cells of each model. It appears that the statistical assumption above was violated. Quoting the P values in the table would then seem inappropriate. It is apparent that calibration was poor or could not be performed because of the small sample size and small number of deaths. In summary, both PRISM III-24 and PIM could not be validated in this study because of the small study population and low death rate. Any conclusion drawn from a study with inherent methodological error is risky.

YW Hui, MSc, FHKAM (Paediatrics)
(e-mail: ywhui@netvigator.com)
Department of Paediatrics and Adolescent Medicine
Princess Margaret Hospital
Laichikok
Hong Kong

References

1. Choi KM, Ng DK, Wong SF, et al. Assessment of the Pediatric Index of Mortality (PIM) and the Pediatric Risk of

Mortality (PRISM) III score for prediction of mortality in a paediatric intensive care unit in Hong Kong. Hong Kong Med J 2005;11:97-103.
2. Shann F, Pearson G, Slater A, Wilkinson K. Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. Intensive care Med 1997;23: 201-7.
3. Pearson GA, Stickley J, Shann F. Calibration of the paediatric index of mortality in UK paediatric intensive care units. Arch Dis Child 2001;84:125-8.
4. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993;39:561-77.
5. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982;115:92-106.
6. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd edition. New York: John Wiley & Sons; 2000.

Authors' reply

To the Editor—We would like to thank Dr Hui for his comments. In our study, we only included patients who were discharged from our paediatric intensive care unit (PICU) or who died. Those patients transferred out of this hospital were not included.[1] We agree in part with Dr Hui that "omitting such cases may affect the validity of any conclusion drawn in the study"; or it may not. It would depend on the number of transferred patients and the outcome which, in turn would depend on the quality of health care service provided by the receiving unit as well as the referring unit. During the study period, only 13 patients (most had cancer or congenital heart disease) were transferred to another hospital. One death occurred in a child with hydrocephalus and multi-organ failure. In view of the small number of patients transferred, including them would not have altered the study conclusion.

We elected not to use the Hosmer-Lemeshow goodness-of-fit test on the advice of the author of the Paediatric Index of Mortality (PIM).[2] He stated that "the p value (from Hosmer-Lemeshow test) is particularly unreliable with sample sizes less than 400, or when there are few deaths, or when more than four of the 20 values in the expected columns in the table are less than 5.0". We also acknowledged that categorisation of risk by "deciles of risk" should be used in a large-scale study such as those quoted by Dr Hui. Nonetheless other validation studies of PIM or PRISM III with smaller sample sizes rarely classified risk by "deciles of risk".[3,4] We agree with Dr Hui that P values from the Chi squared tests should not be used to interpret our results. We actually did not use it to reach our conclusion in the study. Instead, we used standardised mortality ratio and 95% confidence intervals (CIs).

We fully agree with Dr Hui that our study did not properly assess the calibration of both PIM and PRISM models because of the limited sample size. We stated that "no conclusions can be drawn because of wide 95% CIs with insufficient sample size" and recommended "further study with sufficient mortality should be carried out to confirm these findings". We therefore stand by our statement that "both PRISM III and PIM scoring systems accurately predicted mortality" in our PICU and would like to reiterate that "a multicentre study is required to validate the models in Hong Kong." It is hoped that our study, which showed favourable preliminary results, will be the first step in the long process of validating PIM in Hong Kong PICUs.

DKK Ng, FRCP, FHKAM (Paediatrics)
(e-mail: dkkng@ha.org.hk)
SF Wong, BSc
CH Chan, BSc
KL Kwok, MRCP, FHKAM (Paediatrics)
PY Chow, MRCP, FHKAM (Paediatrics)
Department of Paediatrics
Kwong Wah Hospital
Hong Kong

References

1. Choi KM, Ng DK, Wong SF, et al. Assessment of the Pediatric Index of Mortality (PIM) and the Pediatric Risk of Mortality (PRISM) III score for prediction of mortality in a paediatric intensive care unit in Hong Kong. Hong Kong Med J 2005;11: 97-103.
2. Shann F. Are we doing a good job: PRISM, PIM and all that. Intensive Care Med 2002;28:105-7.
3. Gemke RJ, van Vught J. Scoring systems in pediatric intensive care: PRISM III versus PIM. Intensive Care Med 2002;28: 204-7.
4. Ozer EA, Kizilgunesler A, Sarioglu B, Halicioglu O, Sutcuoglu S, Yaprak I. The comparison of PRISM and PIM scoring systems for morality risk in infantile intensive care. J Trop Pediatr 2004;50:334-8.