Assessment and Optimization of Prognostic Scores in Portuguese Pediatric Intensive Care Units (PICUs)

"Are they doing a good job?"

MARTINS, D. A. A.[1]; SANTOS, L. R.[1]; FIGUEIREDO, M. A. R. M.[1]; GUERRA, M. D. L.[1]; MAGALHÃES, R. S.[1]; FRANCISCO, R. M. H.[1]; REBELO, C. A. B. S. T.[1]; ABREU, C. A. M. L. M.[1]; LAIGINHAS, A. R. A.[1]; DIAS, M. O.[1]; CHAVES, J. G. A.[1]; OLIVEIRA, R. C. S.[2];

[1]Class 4, Introdução à Medicina, Faculdade de Medicina da Universidade do Porto (turma4fmup09@gmail.com);

[2]Adviser, Introdução à Medicina, Faculdade de Medicina da Universidade do Porto (rcoliveira@med.up.pt)

Abstract

Introduction: Rapid advances in critical care technology and rising cost of medical care have spurred the development of outcome analysis including mortality risk prediction. Severity scoring systems integrate clinical data to estimate the probability of mortality, which can be used to facilitate resource utilization or continuing quality improvement and to stratify patients for clinical research. In spite of the development of specifics scores for pediatric populations in intensive care context and their effective implement at located realities no validation evidences, in order to its application in Portuguese PICUs, have been previously referenced.

Aims: To assess and optimize the Pediatric Risk of Mortality (PRISM and PRISM III) and the Pediatric Index of Mortality (PIM and PIM2) scoring systems, in comparing the risk-adjusted mortality of children after admission in Portuguese Pediatric Intensive Care Units (PICUs).

Design: Prospective, observational, analytical and multicenter study.

Methods: Data was acquired from a database previously created in the context of a precursor project developed in the institution of filiation. The PRISM, PRISM III, PIM and PIM 2 scores of all patients included in the study were computed according to the published algorithms, and the outcome was noted in terms of survival or non-survival and compared with observed mortality by Standardized Mortality Ratio (SMR). Mortality discrimination was quantified calculating the area under the receiver operating characteristic (ROC) curve. Hosmer and Lemeshow goodness-of-fit test was used to assess scores calibration. To improve calibration of PIM2 prognostic model, a first-level customization was performed, using logistic regression on the original score, with base on

Portuguese patients data, and the corresponding probability of PICU death was calculated for the customized score (C-PIM2).

Results: One thousand and eight hundred and nine patients, with a mean age of 4.6 years and male to female ratio of: 1.2:1, admitted at three volunteers Portuguese PICUs (Oporto, Coimbra, Lisbon) were enrolled. Hosmer-Lemeshow statistics showed good calibration for all original models, except for PIM2, which displayed significant lack of fit and therefore poor calibration (p=0,027). Discrimination was generally good for all models, with areas under the receiver operating characteristic curves (AUC) ranged from 0,84 (PIM) to 0,91 (PRISM III).

Conclusion: Excluding PIM2**,** the predicted mortality using all prediction models correlated well with the observed mortality. All scores present good capacity of discrimination between survivors and non-survivors patients. With the exception of PIM2, all scores are tools with comparable performance at the prognostic evaluation of the pediatric patients admitted at a general Portuguese PICU. It is now important that these tools be used to monitor outcome and to improve the quality of pediatric intensive care within Portugal.

Key words: *Pediatric Intensive Care (MESH); Health quality (MESH); Clinical score (MESH); Mortality prediction (MESH); Validation (MESH);*

Introduction

Following the rapid advances in medical therapy and critical care technology over the past 30 years, outcome analysis, including **mortality risk prediction**, has become a challenge for the modern day intensivists.

Actually, Pediatric intensive care units (PICU), aiming at promoting **qualified care** and so, as points of **major technology transfer**, constitute, effectively, one of the main consumers of hospital budgets. One way to assess performance considers PICU outcomes measures, such as patient risk-adjusted mortality, often provided by clinical scoring systems. When patients with various prognoses and degrees of clinical severity are being treated, the final results of employing the resources available at such units are often uncertain[1]. In this context, the incorporation of technology does not always follow strict analytical rules, with respect to support scientific evidence or, even less frequently, cost-efficiency relationships.[2] Routinely assessment of the quality of the services provided, thus, emerges as a central point in the way of cost-benefit analysis' increment.

One means of comparing the quality and efficacy of care provided at a given unit is to compare it with others in similar situations. Pediatric Intensive Care Units (PICUs) compare components that are related with **disease severity** and the resources available with the outcomes of specific types of patients. Mortality decrease is unquestionably a primary aim of a PICU and given the relatively high mortality among intensive care patients, **death** is, veritably, a sensitive, appropriate, and meaningful measure of outcome.[2]

The assessment of disease severity is essential for a wide range of analysis in Intensive Care Units (ICU), including a quality assessment stratification of severity in clinical trials and studies on the management and use of resources at ICU. A physician's estimation accuracy of mortality risk for patients admitted in PICUs may be skewed and subjective. A rational and objective way to define and quantify severity of illness is, so, through the development of **probability models predicting mortality risk**. In order to measure severity risk of mortality, **scores** are employed, making possible a reproducible comparison of the **estimated mortality** (in percent) with the **really observed mortality**. Known as **prognostic scores**, they integrate clinical data to estimate the probability of mortality, which can be used to evaluate the quality of medical care, including decision-making for individuals patients, and to optimize the employment of resources, aiming at final an improving of the cost-benefit relationship.[3]

In fact, although rigorous experiments or large randomised controlled trials are the gold standard for evaluating existing or new interventions, these are not always possible in intensive care. The alternative is to use observational methods that study the outcome of care patients receive as part of their natural treatment. However, before inferences can be drawn about outcomes of treatment in such studies, the characteristics of the patients admitted to intensive care have to be taken into account, being this process known as adjusting for case mix. Scoring systems are aimed at quantifying case mix and designed to estimate outcome (death before discharge from hospital after intensive care), covering an intrinsic description of the health care system, PICU organization and case-mix of the population used to create it.[3]

Nowadays various are the scores used for the stratification of disease severity in patients hospitalized in ICU, for example: *Acute Physiology, Age, Chronic Health Evaluation* - APACHE II, *Simplified Acute Physiology Score* - SAPS II, *Clinical Risk Index for Babies* - CRIB, *Pediatric Risk of Mortality* – PRISM and *Paediatric Index of Mortality* - PIM). This scores use probabilistic prediction models of individual risk of death and are developed by identifying variables relevant to mortality risk (namely, recovery post procedure, systolic blood pressure, mechanical ventilation, pupillary reactions, potassium, respiratory rate, heart rate, etc.) and scoring them after a multivariate statistical analysis by logistic regression, the main scores used for the pediatric population are PRISM[3] and PIM[4]. Intending to respond to the necessity of regular reevaluation of the relationship between

physiologic status and mortality risk as new treatment protocols, therapeutic interventions or monitoring strategies are introduced, recent versions of PRISM and PIM scoring systems, PRISM III[5] and PIM-2[6], have been introducing in pediatric intensive care context. Advantages and disadvantages of the most frequently used scores for pediatric intensive care population are presented in *Table 1*.

	PRISM	**PRISM III**	**PIM**	**PIM 2**
	1988 (USA)	1996 (USA)	1997 (Australia)	2003 (Australia)
Advantages	• Lesser number of physiologic variables required for PICUs mortality risk assessment relatively Physiologic Stability Index (PSI); • Excellent discriminatory and predictive performance • Most widely used in clinical routine;	• Better predictive capability; • Appropriated age-adjusted physiologic variable;	• Analyzes the condition of the patient directly upon arrival in the PICU; • Good predictions and classifications of mortality in groups of children hospitalized in intensive care units; • Reduced number of variables; • Data collection facilitation; • Well performing at predicting death;	• Better calibrated, safer and better adjusted for varying diagnostic groups;
Disadvantages	• Laborious collection of information; • Fee required for the use of its regression equations; • Long period necessary to collect variables; • Obscured poor quality of care;	• Obscured poor quality of care; • Long period necessary to collect variables; • Laborious collection of information;		
Number of variables	14	17	8	10
Period of observation	24 hours	12/24 hours	Directly upon arrival	Directly upon arrival

***Table 1** – General scores description and comparison*

Reliability and **validity** are important parameters that allow confident use of a scoring system in intensive care patients with different case-mixes and baseline characteristics. In fact, **validation** of a mortality-scoring model is really vital when assessing its ability to predict one of the many important outcome measures, namely death. In fact, the validation needs both discrimination and calibration in order to achieve a global evaluation of the score.[7]

In spite of the development of specifics scores for **pediatric populations** and more, their effective validation at located realities, no validation evidences, in order to its application in Portuguese PICUs, have previously been referenced. A project, however, even though with no published evidences, *Development and Assessment of Optimal Risk Scores for Outcomes in Paediatric Intensive Care (DAIP-CIP), was* conducted in the institution of filiation by Pediatric Intensive Care Unit – Department of Pediatrics (H. S. João – Oporto), providing preliminary results in what concerns to assess the performance of PRISM, PRISM III and PIM in predicting patients' mortality risk in Portuguese PICUs.

In point of fact, the development of referred prognostic systems occurred in a **specific geographic context** and several **contradictory data** on its application in other populations have already been referenced[8,9,10], emphasizing the vital necessity of its **accuracy evaluation** in large cohort of Portuguese context. Special caution is effectively needed in adopting a severity of illness scoring system to assess performance of care, particularly in contexts different from the ones in which the instrument was originally developed.

Remaining the doubt about its **accuracy in mortality prediction** in Portuguese PICUs, the present study intend, so, to proceed in order to assess and optimize Pediatric Risk of Mortality (PRISM and PRISM III) and Pediatric Index of Mortality (PIM and PIM2) scoring systems for **use in comparing the risk-adjusted mortality** of children after admission for pediatric intensive care, in a large, prospective and representative sample of admissions to Portuguese PICUs.

Methods:

Methods scheme was designed with base on three major tasks, above extensively described: Data acquisition, Algorithms Calculation, Validity Statistical Assessment.

Strategy was mainly defined attending a parallel study, with the same thematic and analytical bases, conducted in the United Kingdom[11] and other prognostic scores validation studies conducted in Portugal[12] [e.g. APPACHE (*Acute Physiology, Age, Chronic Health Evaluation*), SABS (*Clinical Risk Index for Babies*)].

1) Data acquisition

Data was acquired from a database previously created in the context of the precursor project *Development and assessment of optimal risk scores for outcomes in pediatric intensive care (DAIP-CIP),* extensively described before.

Data was already gathered previously and history of database creation describes a prospective collection procedure, elapsed during a period of 30 months in 3 volunteers Portuguese PICUs (Hospital Pediátrico de Coimbra - Coimbra, Hospital D. Estefânia - Lisbon, Hospital São João - Oporto). All necessary variables used for the calculation of PRISM, PRISM III, PIM and PIM 2 (through routinely collection performed by health professionals and added specific pro-form, not routinely preconized) were gathered. All admissions between 29 days and 16 years old were included in a total of 1809 admissions, without any more inclusion/exclusion criteria known. Data analysis of inter-observer, conducted in the second quarter of data collection, was performed in the way of data collection method's quality assessment.

2) Algorithms calculation

Predicted probability of PICU mortality was calculated using the published algorithms for PIM, PIM2, PRISM and PRISM III.

3) Statistical Analysis

PASW Statistics18.0 was used for the statistical analysis, and α=0.05 was set as significance level. Performance of mortality risk scores were evaluated by assessing algorithms discrimination and calibration, and by comparison of observed and expected number of deaths through Standardized Mortality Ratio (SMR) analysis. Measuring the area under the Receiver Operating Characteristic (ROC) curve assessed discrimination. Hosmer and Lemeshow goodness-of-fit chi-square test assessed calibration. In descriptive Statistical analysis data are presented as mean±standard deviation.

Standardized Mortality Ratio (SMR)

An indirect mean of adjusting a rate, **Standardized Mortality Ratio (SMR)** is, commonly, defined as a ratio of **observed deaths** to **expected deaths** according to a specific health outcome. It is often used for comparing the observed mortality with the expected mortality would occur if the standard rates were applied.

The SMR may well be quoted with an indication of the uncertainty associated with its estimation, such as a confidence interval (CI) or p-value, which allows it to be interpreted in terms of statistical significance.

In clinical context, SMR is frequently used serving the **comparative audit purpose** of prediction systems, such as prognostic scores, in the way of evaluation of the services quality in clinical institutions. A SMR >1 usually reflects poor care.[12]

Discrimination

Discrimination is often defined as the ability of distinction between **survivor and non-survivor**. Predictions from each model are assessed using the **c index** (area under the receiver operating characteristic curve) for discrimination, which indicates, in fact, the probability of concordance between outcomes and predictions. In this study, it represents the probability that a randomly chosen patient who died will have **a higher predicted probability of mortality than a randomly chosen patient who survived**.

Published c-index criteria suggest that an area under the curve of 0.70-0.79 represents acceptable discrimination, being good discrimination represented by an area higher than 0.80. Excellent discriminatory power is represented by an area under the curve higher than 0.9[13].

Calibration

Calibration measures the correlation between the predicted outcomes and actual outcome over the entire range of risk prediction, this is, how well the predicted **probabilities of mortality** that were generated by the risk-adjustment models compared with **the observed mortality** will be assessed using the **Hosmer-Lemeshow test**. For each risk-adjustment model, for the Hosmer-Lemeshow goodness-of-fit chi square test, patients were categorized into 10 groups (eventually less)

according to quintiles of their associated predicted probability of mortality, and the observed and expected outcomes were compared using a **chi-square statistic**.

Interpretation of Hosmer-Lemeshow goodness of fit test shows that if the difference between the observed and expected mortality is not significant, then they are comparable and the model has a significantly good calibration. Perfect calibration would be indicated by a model with a constant term of 0 and a slope term of 1. Significant differences from these values give a quantifiable indication of where the calibration of the models has failed. If **the model predicts well**, the events will be **concentrated in the highest risk groups**.[13]

In present study, Hosmer-Lemeshow goodness of fit test was applied in the way of score's calibration evaluation in five categories of expected mortality probability (namely <1%; [1,5[%; [5,15[%; [15,30[%; >30%).[15]

3) Optimization of PIM2 current model

Attending to the poor calibration revealed by PIM2 in Portuguese data, a first-level customization, in the way of its optimization for a better fitting in Portuguese reality, was performed. Logistic regression on the original score, with base on Portuguese patients data, was made and the corresponding probability of PICU death was calculated for the customized score (C-PIM2). Calibration and discrimination were assessed, on the development sample, in the customized model, as previously described for the original models.

Results

1) Data characterization

Study group comprised a total of 1809 patients admitted at three volunteers Portuguese PICUs (Oporto, Coimbra, Lisbon), 977 male, with male:female ratio of 1.2:1 and a mean age of 4.6 years (range: 0.8 years). The average duration of internment was 7.7 days, during which 57.4% of patients were submitted to mechanical ventilation. Medical (59.9%) and surgical (37.3%) were the most common reasons for admission, being a minority of them (2.9%) justified by monitoring and prevention intentions. A majority of total admissions (68.6%) was unplanned. On the total 1809 admissions included in the analysis, 8.6% died in the admitted PICU. An additional 1.4% died before discharge from the hospital (*Table 2*).

	Value
Number of patients	1809
Observed mortality (%)	(8.6)
Age – mean (standard deviation)	4.6 (4,83) years
Internament duration – mean (standard deviation)	7.7 (23,0) days
Gender: male (N (%)	977 (54)
Admission reason (%) •Surgery •Medical •Monitorization and prevention procedures	 (37.3) (59.9) (2.9)
Mechanical ventilation during internament **(%)**	(57.4)

Table 2 – Study group characteristics (Patients' demographic, type of admission and general diagnostic categories).

Score	Observed Mortality (%)	Expected Mortality(%)	SMR (CI 95%)
PIM	8.6	6.0±0,4	1.43 [1.21;1.66]
PIM2	8.6	5.3±0,3	1.62 [1.37;1.87]
PRISM	8.6	9.9±0,6	0.87 [0.74;1.00]
PRISMIII	8.6	7.4±0,5	1.16 [0.98;1.34]

Table 3 – Expected mortality and Standardized Mortality Ratio according to original score prediction

PIM, PIM2, PRISM and PRISM III quantities for original scores predicted mortality percentage means were respectively 6, 5.3, 9.9 and 7.4, compared to 8.6 (%) of observed mortality. Concerning the Standardized Mortality Ratios, 1.43, 1.62, 0.87 and 1.16 were the obtained values, respectively (*Table 3*).

2) Discrimination and calibration assessment

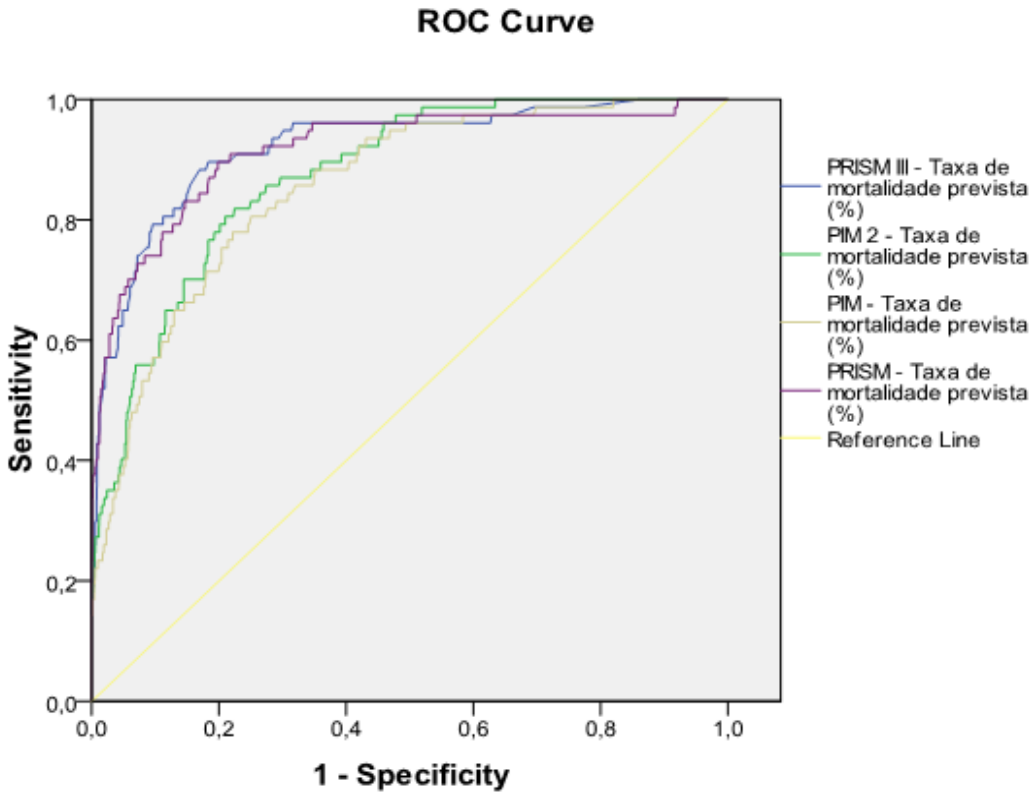

Figure 1 – Receiver operating characteristics (ROC) curves.

Score	Discrimination	Calibration	
	AUC	Chi-square	p-value
PIM (n = 1809)	0,84	4,05	0,132
PIM2 (n = 1809)	0,89	7,23	0,027
PRISM (n = 1809)	0,90	3,62	0,305
PRISM III (n = 1809)	0,91	1,96	0,375

Table 4 – Discrimination and calibration according to original scores prediction.

Hosmer-Lemeshow statistics showed good calibration for PRISM (p=0,305), PRISM III (p=0,375) and PIM (0,132), but poor calibration for PIM2 (p=0,027). Discrimination was generally good for all models, with areas under the receiver operating characteristic curves (AUC) ranged from 0,84 (PIM) to 0,91 (PRISM III). No one of them revealed a significantly better discrimination power than the others, at a level of significance of 5% *(Table 4)*.

3) Optimized version of PIM2 (C-PIM2) - Discrimination and calibration assessment

Concerning optimized version of PIM2's (C-PIM2), discrimination and calibration tests information showed good discriminatory power (AUC = 0,87) and good calibration (p=0,477).

Discussion

Modern paediatric intensive care is characterized by increased sophistication, resulting in spiralling costs. Auditing the PICU is thus an integral component in health care planning and management. There is a need to accurately define prognosis, so that the physician can be guided in clinical decision-making, including the appropriateness of therapy. Moreover, the impact of new technologies and medical intervention can be assessed in a more objective fashion[16].

Wells et al., attributes the difficulties in achieving exactly the same progress for two patients with the same level of clinical instability, i.e. the same prognostic score results, to two basic causes. The first cause is the differences in individual clinical conditions that are not evaluated by the score, such as, for example, the nutritional status or physical reserves of each individual. The second cause is the differences in working conditions and infrastructure at each PICU. Units with greater availability of machines and medication can offer their patients treatment more quickly and thus impact on their progress.[17]

When assessing the performance of prognostic scores in clinical context, there is no consensus on which function is more important for a prognostic score: calibration or discrimination. Both are important for determining the adjustment capacity of a model. Which function is most important will depend on the objective for which the prognostic score is being used. If, for example to distinguish between those who are more likely to die from those who are more likely to survive, then the capacity to discriminate is most important, but, if, however, the reason for using a score is to compare observed and expected mortality at different intervals of severity, then calibration capacity is

more important. However, in order to achieve a global evaluation of the score, both discrimination and calibration should be considered.

Attending to the results obtained about the assessment of scores performance in Portuguese clinical context, and with the exception of PIM2, the good discriminatory power and the so acceptable calibration are appointing in the way of its use with clinical significance. It is clear that there are many variables unmeasured by the prognostic scores studied, which make it difficult to classify severity levels of different patients in different intensive care units and, therefore, to find a prognostic index model with an extraordinary calibration capacity. The real challenge is to identify which variables do not have a similar predictive power for the population being studied. The interpretation of the mortality index of a PICU is dependent of statistical factors such as sample size, mortality rate at each severity level and random variations in the study population. The most powerful variable will be the one that when added to the model is observed often, i.e., will be found with higher frequency among the patients in the population. We should, therefore, seek the power of the variables that a most similar to the reality of our population. Actually, for a tool to calibrate well in a validation sample, all factors that influence outcome must either be included in the model or have the same distribution in the validation sample as the sample used to develop the model. Differences between countries and over time make this second condition less likely. PICUs have been found to vary considerably in their structure, organization, and staffing across European countries and still can be anticipated to do so between Portugal and USA or Australia. A first level process of scores customization, with an eventual re-estimation of score coefficients for national use purposes, emerges so as a central point before the appliance of its scores to particular clinical contexts, such as made for PIM2. A process of external validation of score is, however, required, once the use of same sample to fit the model and to evaluate the customized version performance can introduce bias.

The amount of data to be collected is an important consideration in the design of any study; therefore, PIM may be valuable for most quality improvement and audit studies that require the minimum of data collection necessary to undertake the study. Larger scale research projects with funding to collect additional data could reasonably choose to use PRISM III, or ever PRISM (attending to economical aspects, once the last one is free). Whichever risk-adjustment method is chosen, units that wish to participate in comparative research studies or collaborative quality improvement efforts must ensure that participating centers not only collect the data for the risk-adjustment tool in a standardized manner but also that standardized reporting of outcomes are fed back to the units[18].

Admissions to PICUs in Portugal have an acceptable variance in their characteristics, making adjustment for case mix imperative when comparing outcomes between different units. However, the

high percentage of absent information in this study relatively to the diagnostic group derails any possible approach to it. An analysis of scores performance in specific patients diagnostic groups is effectively needed.

This study reinforces the importance of the use of a representative and large number of units when assessing such risk adjustment methods in another country or health cares system. A small or unrepresentative selection of units could lead to confusion between unit or overall system performance and the need to recalibrate the score. Even including all units may result in poor calibration, but this does not necessarily invalidate the scores as a simple recalibration of these scores can lead to their being useful and valid tools for that country or health care system.

Any variation in risk-adjusted PICU mortality needs to be tempered by an assessment of variation in longer-term outcome. A reduction in mortality is of dubious benefit if it is at the expense of increased severe morbidity[19]. Besides the uncertainty about the accuracy of different death prediction indexes, which is a rare event after admission to PICUs, there is still the concern of keeping alive patients with severe disability from the point of physiological status and for prolonged periods of time. Thus, the morbidity can be more relevant than the mortality for the comparison of treatment efficacy between groups patients undergoing PICUS and life expectancy of the survivors can minimize the impact morbidity children subject to the PICU.

Standardized and reliable methods of measuring health status (or at least clinical status) are required for use in Portugal. Currently, only the PRISM family of tools has been shown to have any relationship to longer-term outcomes in the United States[19].

Conclusion

This study is the first, of our knowledge, to systematically validate mortality prognostic scores for using in Portuguese pediatric intensive care context. It reinforces the idea that risk-adjustment methods that are developed primarily in other countries require validation before being used to provide risk-adjusted outcomes of PICU mortality for units within a new health care setting. It is also important that the calibration of these tools be reassessed periodically to ensure their continued validity.

The identification of the most suitable score, for use in the context of Portuguese pediatric intensive care, could be, in this moment, something imprudent. More studies, evaluating their appliance in a more detailed way (namely economical and logistic questions), are needed. However, with the exception of PIM2, all scores seems to be tools with comparable performance at the prognostic evaluation of the pediatric patients admitted at a general Portuguese PICU.

The identification of suitable risk-adjustment tools is only a first step. It now is important that they be applied effectively to monitor outcome and to improve the quality of pediatric intensive care within Portugal.

Acknowledgements

The authors gratefully acknowledge: the guidance provided by professors Armando Teixeira Pinto e Rosa Oliveira; the criticism and suggestions made by professor Altamiro da Costa Pereira; the providing of data to the implementation of this project by Dr. Francisco Cunha.

References

1. Gemke RJ, Bonsel GJ, Bught AJ. Outcome assessment and quality assurance in pediatric intensive care. In: Tibboel D, van der Voort E, editors. Intensive care in childhood – a challenge to future. 2nd ed. Berlin: Springer; 1996. p. 117-32.
2. Mitchell I. Nature and nurture: the future of predictor variables. Curr Opin Crit Care. 2000;6:166-70.
3. Pollack MM, Ruttimann UE, Getson PR. The Pediatric Risk of Mortality (PRISM) score. Crit Care Med. 1988;16:1110-6.
4. PRISM III: an updated Pediatric Risk of Mortality score. PRISM III: An updated Pediatric Risk of Mortality score. Crit Care Med. 1996;24:743-52.
5. 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.
6. Slater A, Shann F, Pearson G. PIM2: a revised version of the Paediatric Index of Mortality. Intensive Care Med. 2003;29:278-85.
7. Gunning K, Rowan K. ABC of intensive care outcome data and scoring systems. BMJ. 1999;319:241-4.
8. 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;
9. Leteurtre S, Leclerc F, Martinot A, Cremer R, Fourier C, Sadik A, et al. Can generic scores (Pediatric Risk of Mortality and Pediatric Index of Mortality) replace specific scores in

predicting the outcome of presumed meningococcal septic shock in children? Crit Care Med. 2001;29:1239-46.

10. Shann F. Are we doing a good job: PRISM, PIM and all that. Intensive Care Med. 2002;28:105-7.
11. *Brady, R. A. (2006). Assessment and Optimization of Mortality Prediction Tools for Admissions to Pediatric Intensive Care in the United Kingdom. Pediatrics;*
12. *Francisco, G. A. (2003). Performance of SAPS3, compared with APACHE II and SOFA, to predict hospital mortality in a general ICU in Portugal;* European Journal of Anaesthesiology;
13. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA. 1982;247:2543–2546;*
14. Deeks, J. J.; Altman, D. G. (2004); *Diagnostic tests 4: likelihood ratios ; Intensive Care Med;*
15. Martha, V.; Garcia, P.; (2005); *Comparison of two prognostic scores (PIM and PRISM) at a pediatric intensive care unit;* Brazilian Journal of Pediatry; 0021-7557/05/81-03/259;
16. Ruttimann UE, Pollack MM, Fiser DH. Prediction of three outcome states from pediatric intensive care. *Crit Care Med.* 1996;24:78–85;
17. Tarnow-Mordi WO, Tucker J, Parry GJ. Should paediatric intensivecare be centralised? *Lancet.* 1997;350:66–67;
18. Kuhlthau K, Ferris TGG, Iezzoni LI. Risk adjustment for pediatric quality indicators. *Pediatrics.* 2004;113:210–216;
19. Keizer NF, Bonsel GJ, Gemke RJBJ. Health status prediction in critically ill children: a pilot study introducing standardized health ratios. *Qual Life Res.* 1997;6:192–199;