



Comparison of two prognostic scores (PRISM and PIM) at a pediatric intensive care unit

Vanessa Feller Martha,¹ Pedro Celiny Ramos Garcia,² Jefferson Pedro Piva,³
Paulo Roberto Einloft,⁴ Francisco Bruno,⁴ Viviane Rampon⁵

Abstract

Objective: To compare the performance of the PRISM (Pediatric Risk of Mortality) and the PIM (Pediatric Index of Mortality) scores at a general pediatric intensive care unit, investigating the relation between observed mortality and survival and predicted mortality and survival.

Methods: A contemporary cohort study undertaken between 1 June 1999 and 31 May 2000 at the *Pontifícia Universidade Católica do Rio Grande do Sul, Hospital São Lucas* pediatric intensive care unit. The inclusion criteria and the PRISM and PIM calculations were performed as set out in the original articles and using the formulae as published. Statistical analysis for model evaluation employed the Flora z test, Hosmer-Lemeshow goodness-of-fit test, ROC curve (receiver operating characteristic) and Spearman's correlation tests. The study was approved by the institution's Ethics Committee

Results: Four hundred and ninety-eight patients were admitted to the pediatric intensive care unit, 77 of whom presented exclusion criteria. Thirty-three (7.83%) of the 421 patients studied died and 388 patients were discharged. Estimated mortality by PRISM was 30.84 (7.22%) with a standardized mortality rate of 1.07 (0.74-1.50), $z = -0.45$ and by PIM this was 26.13 (6.21%) with a standardized mortality rate of 1.26 (0.87-1.77), $z = -1.14$. The Hosmer-Lemeshow test gave a chi-square of 9.23 ($p = 0.100$) for PRISM and 27.986 ($p < 0.001$) for PIM. The area under the ROC curve was 0.870 (0.810-0.930) for PRISM and 0.845 (0.769-0.920) for PIM. The Spearman test returned $r = 0.65$ ($p < 0.001$).

Conclusion: Analyzing the tests we can observe that, although the PIM test was less well calibrated overall, both PRISM and PIM offer a good capacity for discriminating between survivors and moribund patients. They are tools with comparable performance at the prognostic evaluation of the pediatric patients admitted to our unit.

J Pediatr (Rio J). 2005;81(3):259-64: Prognostic scores, PRISM, PIM, mortality.

1. MSc. Physician, Pediatric Intensive Care Unit, Hospital Moinhos de Vento and Emergency Service, Hospital da Criança Santo Antônio, Porto Alegre, RS, Brazil.
2. PhD. Professor, School of Medicine, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Chief of the Intensive Care Unit and Pediatric Emergency, Hospital São Lucas, PUCRS, Porto Alegre, RS, Brazil.
3. PhD. Professor, School of Medicine, PUCRS. Associate physician, Intensive Care Unit and Pediatric Emergency, Hospital São Lucas, PUCRS, Porto Alegre, RS, Brazil.
4. Professor, School of Medicine, PUCRS. Associate physician, Intensive Care Unit and Pediatric Emergency, Hospital São Lucas, PUCRS, Porto Alegre, RS, Brazil.
5. Physician, Intensive Care Unit and Pediatric Emergency, Hospital São Lucas, PUCRS, Porto Alegre, RS, Brazil.

Manuscript received May 10 2004, accepted for publication May 16 2005.

Suggested citation: Martha VF, Garcia PC, Piva JP, Einloft PR, Bruno F, Rampon V. Comparison of two prognostic scores (PRISM and PIM) at a pediatric intensive care unit. *J Pediatr (Rio J)*. 2005;81:259-64.

Introduction

Pediatric intensive care units (PICU) aim at promoting qualified care with the objective of achieving the best results and better progress for critically ill children. These units are points of major technology transfer and constitute one of the main consumers of hospital budgets. However, when patients with varying prognoses and degrees of clinical severity are being treated the final results of employing the resources available at such units is often uncertain. In this context the incorporation of technology does not always follow strict analytical rules with respect of supporting scientific evidence or, even less frequently, cost-efficiency relationships.¹

One means of comparing the quality and efficacy of care provided at a given unit is to compare it with others

in similar situations.² Pediatric ICUs compare components that are related with disease severity and the resources available with the outcomes of specific types of patients. Mortality and length of hospital stay are examples of the most used outcomes. In order to measure severity risk of mortality scores are employed that establish a numerical scale and in this way they compare estimated mortality in percent with the observed mortality.³ Known as prognostic scores, these can be used to evaluate the quality of medical care and to optimize the employment of resources, aiming at improving the cost-benefit relationship. Since they compare mortality adjusted by disease severity, these scores can also be used for comparisons between clinical trials and for planning technological resources in this area.⁴

The principal scores that have been developed for the pediatric population are the PRISM (Pediatric Risk of Mortality)⁵ and PIM (Paediatric Index of Mortality),⁶ with their most recent versions being PRISM III⁷ and PIM-2.⁸ These scores were developed by identifying variables relevant to mortality risk and scoring them after a multivariate statistical analysis by logistic regression.⁹

The PRISM score was published in 1988 by Pollack *et al.* and exhibited an excellent discriminatory and predictive performance.⁵ It is still the most widely known and used at PICU and is used in clinical trials as a standard prognostic score for evaluation of disease severity in pediatric patients. A revised version of the PRISM score, PRISM III, has been available since 1996,⁶ which, according to its authors, offers better predictive capability.¹⁰ However, a considerable fee is charged for using it routinely, which has limited its use, even in developed countries¹¹⁻¹³ and for this reason it was not evaluated in this study.

The results of the original PIM article, published in 1997 by Shann *et al.*, provided evidence that the model was capable of good predictions and classifications of mortality in groups of children hospitalized in intensive care units.⁷ The authors suggest that one advantage of the PIM over the PRISM is the fact that the PIM is based on just 8 variables, all of which are collected at the point of admission, which facilitates data collection and avoids any impact on the results from 24 hours of intensive management strategies.¹⁴ Several articles that have evaluated the PIM have shown that it performs well at predicting death.¹¹⁻¹⁶ In 2003 the PIM Study Group published a revised version of the PIM, the PIM-2^{16,17} which, compared with the original version,¹⁸ is supposed to be better calibrated, safer and better adjusted for varying diagnostic groups. This new version has not yet been evaluated independently and more information is necessary with respect of its performance in other regions and, because it was published after this study began, it was not investigated.

The performance of the PRISM and PIM systems have been compared a number of times by the authors who developed the scores themselves,^{15,18-20} but have rarely been compared independently. To date, those studies that have been performed independently have not used

heterogeneous groups of patients from PICUs, but have investigated certain specific disease categories,^{11,16,21} new versions of the methods¹⁰ or homogeneous groups of high mortality patients.²² No studies of this type have been published in Latin America.

In this, independent, study our objective was to compare the performance of the PRISM and the PIM at a general PICU, investigating the relationship between observed mortality and survival and the mortality and survival rates estimated by the two scores.

Methods

A contemporary cohort study performed between 1st June 1999 and 31st May 2000 at the PICU of the PUCRS Hospital São Lucas. Data for calculating scores and predictions was collected prospectively over the period and with the techniques set out for each score (PRISM - first 24 hours after admission, PIM - up to one hour after admission).^{5,6} Patients were excluded from the study if they died within the first 8 hours or were discharged within the first 24 hours after admission.

A minimum of 253 patients were estimated to be necessary for the present study. The sample size calculation was based on a mean population of 500 patients in the PICU, setting mortality at 15% and tolerating a mortality range of 10 to 20% with a 99% confidence limit.

The PRISM and PIM scores were calculated using the formulae available in their original articles.^{5,6} No tests were performed to meet the needs of this research since it considered non-collection as normal. Demographic data was collected in order to characterize the sample, including age at admission, sex, origin. The outcomes assessed were length of hospital stay at the unit and patient progress (discharge or death).

Simple descriptive analysis was utilized for the groups and subgroups under study (mean, median, standard deviation). The "z" statistic, as described by Flora,²³ was utilized to compare the general similarity between observed mortality and that estimated by the standardized mortality rate (SMR). In order to calibrate the scores, the Hosmer-Lemeshow goodness-of-fit test was employed to test the agreement between observed and expected mortality, at five different risk intervals.²⁴ The capacity for discrimination between survivors and moribund patients was made using the typical area under a receiver operating characteristic curve (ROC curve)^{25,26} and quantitative correlation between the results of the scores was analyzed using the Spearman test.

The study was approved by the Committee for Ethics in Research at the *Pontifícia Universidade Católica do Rio Grande do Sul* and, since the study incurred no additional risk to patients, informed consent was waived with commitments made to maintain patient's identities confidential.

Results

During the study period, 498 patients were admitted to the PICU. However, 77 patients presented exclusion criteria; eight died during the first 8 hours after admission and the remainder because they were discharged before they had spent 24 hours in the PICU. No patients were excluded because of lack of data. The general sample characteristics are given in Table 1.

Table 1 - Characteristics of the general sample

Characteristics	Value
Total sample	421 patients
Death	33 (7.8%)
Age (months); mean (median)	44.3 (26%)
Time of hospital stay (days); mean (median)	11.6 (5.5%)
Weigth (kilos); mean (median)	14.9 (11.4%)
Caucasian	345 (82%)
Male gender	212(50%)
Admissions	
Emergency	107 (25%)
Hospital	314 (75%)
Type of hospital stay	
Clinic	234 (56%)
Surgical	187 (44%)
Organic dysfunction diagnosis	
Cardiocirculatory	126 (30%)
Respiratory	113 (27%)
Neurological	92 (22%)
Other dysfunction	90 (21%)

Thirty-three (7.83%) of the 421 patients studied died. Estimated mortality according to the PRISM was 30.84 (7.22%) and by the PIM this figure was 26.13 (6.21%) patients. This corresponds to an SMR (CI= 95%) of 1.07 (0.74-1.50) ($z = -0.45$) for the PRISM and 1.26 (0.87-1.77) ($z = -1.14$) for the PIM. When tested by Flora's z test, these were within the limits for not rejecting the null hypothesis (< 1.96 and > -1.96). Table 2 synthesizes the performance of the models.

Table 3 evaluates similarities in observed and expected mortality, at five mortality risk intervals according to the Hosmer-Lemeshow goodness-of-fit test for the PRISM and the PIM (Table 3).

The discriminatory performance of the models, measured by area under the ROC curve, resulted in an area of 0.870 (0.810-0.930) for the PRISM and 0.845 (0.769-0.920) for the PIM (Figure 1).

The estimated probabilities of death reveal a positive and significant correlation between the PRISM and the PIM, with Spearman's correlation coefficient being $r = 0.65$ ($p < 0.001$).

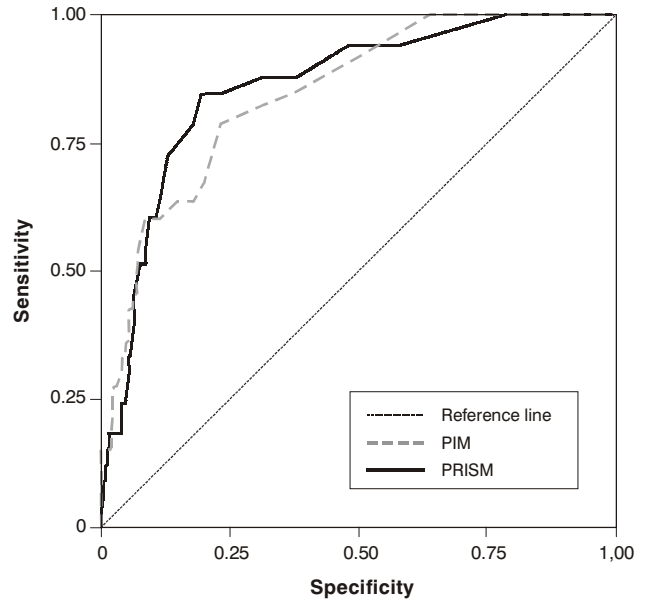


Figure 1 - Superposition of two receiver operating characteristic curves (ROC). The area under the ROC curve was 0.87 for PRISM (CI 95% 0.78-9.2) and 0.85 (CI95% 0.78-0.91) for PIM. Findings were shown to have a good discriminatory performance between survivals and non-survivals

Discussion

Individual analysis of the scores' results by SMR shows us that the PRISM and the PIM offer good performance in predicting the general mortality of our population. Although both models underestimated mortality (PRISM predicted 93.45% of deaths and PIM 79.18%) the two results did not exhibit any significant differences between each other or from observed mortality when tested.

In evaluating the power of calibration by the Hosmer-Lemeshow goodness-of-fit test, it was observed that, in the case of the PRISM, the predicted results were similar to those observed, whereas for the PIM they were different, indicating that the PRISM was well calibrated and that the PIM calibration performed poorly.

The results found when the discriminatory performance of the models was evaluated using the ROC curve showed that both the PRISM and the PIM have good power to discriminate between survivors and moribund patients and that they had similar power.

The present study attempted to validate the PRISM and PIM scores and to compare their results. In certain aspects, the results produced by the PRISM were slightly better than those returned by the PIM. Their performances were similar in terms of the capacity to discriminate between survivors and moribund patients and they exhibited values that correlated directly and positively. Nevertheless, the PIM exhibited poor calibration capacity, which is a problem commonly encountered by studies evaluating prognostic scores.²⁷

Table 2 - Performance of the models

	PRISM	PIM
Mean of mortality risk; % (SD)	7.3±5.5	6.2±6.1
Median of mortality risk; % (IQ)	4.0 (1.5-8.0)	2.0 (1.0-5.0)
Estimated mortality; n	30.84 (7.22%)	26.13 (6.21%)
Standardized mortality rate (SMR) (CI 95%)	1.07 (0.74-1.50)	1.26 (0.87-1.77)
Flora's z test *	-0.45	-1.14
Hosmer-Lemeshow goodness-of-fit test	$\chi^2 = 9.23, p = 0.100$	$\chi^2 = 27.98, p < 0.001$
Area under a ROC (IC 95%)	0.87 (0.81-0.93)	0.85 (0.77-0.92)

SD = standard deviation; IQ = interquartile interval; SMR = standard mortality ratio; CI = confidence interval;

ROC = receiver operating characteristic curve

* When tested with the Flora's z test, PRISM and PIM are within the limits for not rejecting the null hypothesis (< 1.96 and > -1.96).

Table 3 - Calibration of models in five intervals of mortality risk

	Risk %	Number of patients	Observed survival	Expected survival	Observed death	Expected death
PRISM	0 - 1	83	83	82.95	0	0.05
	> 1-5	200	190	196.78	10	3.22
	> 5-15	84	73	77.6	11	6.4
	> 15-30	24	17	19.88	7	4.12
	> 30-100	30	25	17.66	5	12.34
	Total	421	388	394.87	33	26.13
PIM	0-1	33	33	32.95	0	0.05
	> 1-5	269	261	261.15	8	7.85
	> 5-15	67	56	60.26	11	6.74
	> 15-30	27	19	21.65	8	5.35
	> 30-100	25	19	14.15	6	10.85
	Total	421	388	390.16	33	30.84

Expected findings were similar to those found for PRISM ($\chi^2 = 9.23, p = 0.100$) and different from those found for PIM ($\chi^2 = 27.986, p < 0.001$).

There is no consensus on which function is more important for a prognostic score: calibrate or discriminate. 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.²⁸ If, for example, the objective is 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 with 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.

The PIM did not demonstrate good calibration. One possible reason for this could be the small number of deaths at each level. In an article evaluating the PIM in English intensive care units, the author suggests that special care should be taken when differences are small in small series (for example, less than 20 deaths per unit).¹⁵ Slater & Shann recently published²⁰ a comparative study of the performance of the PIM, PIM-2, PRISM and PRISM III scores in units in Australia and New Zealand and found that the PIM-2 was the safest and had the best

adjustment for different diagnostic groups. This allows us to speculate that the new version may also present better results in our milieu.

Wells *et al.*,²⁹ 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.

In our study almost 50% of the patients came from surgery and arrived at the unit in need of intensive care. However, in the majority of cases they were already stable, both hemodynamically and in terms of ventilation. Such patients, while given a low severity score at admission (and consequently a low PIM score), were patients at risk of death and whose condition could deteriorate during the first 24 hours (and consequently have a higher PRISM score) because of postoperative complications.

Patients with respiratory dysfunction arriving at the PICU from emergency or other hospitals may have had blood gas analysis results within normal limits at the cost of elevated ventilator parameters and normal respiratory frequency (set by the respirator or the physician ventilating with a self-inflating bag) and the remainder of clinical conditions stable. These patients would have low PIM scores despite being considered critical patients at elevated risk of death, since their base conditions would not yet have been resolved and could deteriorate to past the limits of assisted ventilation and would be better detected by the PRISM score later.

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 a good calibration capacity.²⁵ The great 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 that which, in addition to changing the score, is observed often, i.e. is to be found in many patients in the population. We should, therefore, seek the power of the variables that a most similar to the reality of our population.³⁰

Until such questions are established, evaluating the performance of the PIM and of the PRISM at a Brazilian PICU, we can state that although the PIM offers poorer calibration, when the results are taken as a whole both scores exhibit good capacity to discriminate between survivors and moribund patients and are tools with comparable performance for the prognostic evaluation of pediatric patients admitted to our unit.

References

- 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.
- Mitchell I. Nature and nurture: the future of predictor variables. *Curr Opin Crit Care*. 2000;6:166-70.
- Pollack MM, Cuerdon TT, Patel KM, Ruttimann UE, Getson PR, Levetown M. Impact of quality-of-care factors on pediatric intensive care unit mortality. *JAMA*. 1994;272:941-6.
- Seneff M, Knaus WA. Predicting patient outcome from intensive care: a guide to APACHE, MPM, SAPS, PRISM, and other prognostic scoring systems. *J Intensive Care Med*. 1990;5:33-52.
- Pollack MM, Ruttimann UE, Getson PR. The Pediatric Risk of Mortality (PRISM) score. *Crit Care Med*. 1988;16:1110-6.
- 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.
- 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.
- Slater A, Shann F, Pearson G. PIM2: a revised version of the Paediatric Index of Mortality. *Intensive Care Med*. 2003;29:278-85.
- Gunning K, Rowan K. ABC of intensive care outcome data and scoring systems. *BMJ*. 1999;319:241-4.
- Marcin JP, Pollack MM, Patel KM, Ruttimann UE. Combining physician's subjective and physiology-based objective mortality risk predictions. *Crit Care Med*. 2000;28:2984-90.
- Gemke RJ, van Vught J. Scoring systems in pediatric intensive care: PRISM III versus PIM. *Intensive Care Med*. 2002;28:204-7.
- Tibby SM, Taylor D, Festa M, Hanna S, Hatherill M, Jones G, et al. A comparison of three scoring systems for mortality risk among retrieved intensive care patients. *Arch Dis Child*. 2002;87:421-5.
- Slater A. Monitoring outcome in paediatric intensive care. *Paediatr Anaesth*. 2004;14:113-6.
- Jones GD, Thorburn K, Tigg A, Murdoch IA. Preliminary data: PIM vs PRISM in infants and children pos cardiac surgery in a UK PICU. *Intensive Care Med*. 2000;26:145.
- 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.
- 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.
- Slater A, Shann F, Pearson G. Paediatric Index of Mortality (PIM) Study Group. PIM2: a revised version of the Paediatric Index of Mortality. *Intensive Care Med*. 2003;29:278-85.
- Shann F. Are we doing a good job: PRISM, PIM and all that. *Intensive Care Med*. 2002;28:105-7.
- Marcin JP, Pollack MM. Review of the methodologies and applications of scoring systems in neonatal and pediatric intensive care. *Pediatr Crit Care Med*. 2000;1:20-7.
- Slater A, Shann F, ANZICS Paediatric Study Group. The suitability of the Pediatric Index of Mortality (PIM), PIM2, the Pediatric Risk of Mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand. *Pediatr Crit Care Med*. 2004;5:447-54.
- Castellanos-Ortega A, Delgado-Rodriguez M, Llorca J, Sanchez Buron P, Mencia Bartolome S, Soult Rubio A, et al. A new prognostic scoring system for meningococcal septic shock in children. Comparison with three other scoring systems. *Intensive Care Med*. 2002;28:341-51.
- Ozer EA, Kizilgunesler A, Sarioglu B, Halicioglu O, Sutcuoglu S, Yaprak I. The Comparison of PRISM and PIM Scoring Systems for Mortality Risk in Infantile Intensive Care. *J Trop Pediatr*. 2004;50:334-8.
- Flora, JD. A method for comparing survival of burn patients to a standard survival curve. *J Trauma*. 1978;18:701-8.
- Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: John Wiley; 1989.

25. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148:839-43.
26. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993;39:561-77.
27. Moreno R, Matos R. The new scores: what problems have been fixed, and what remain? *Curr Opin Crit Care*. 2000;6:158-65.
28. Mourouga P, Goldfrad C, Rowan KM. Does it fit? Assessment of scoring systems. *Curr Opin Crit Care*. 2000;6:176-80.
29. Wells M, Riera-Fanego JF, Luyt DK, Dance M, Lipman J. Poor discriminatory performance of the Pediatric Risk of Mortality (PRISM) score in a South African intensive care unit. *Crit Care Med*. 1996;24:1507-13.
30. Rowan KM, Angus DC. Don't let perfection be the enemy of the good: it's for optimism over the role of severity scoring systems in intensive care unit performance measurement. *Curr Opin Crit Care*. 2000;6:153-4.

Correspondence:

Pedro Celiny Ramos Garcia
Rua Curupaiti, 62
CEP 90820-090 – Porto Alegre, RS, Brazil
Tel.: +55 (51) 3266.5121
E-mail: celiny@terra.com.br