COMPARISON OF THREE PROGNOSTIC SCORES (PRISM, PELOD AND PIM 2) AT PEDIATRIC INTENSIVE CARE UNIT UNDER PAKISTANI CIRCUMSTANCES

Ahmad Usaid Qureshi, Agha Shabbir Ali, Tahir Masood Ahmad
Children's Hospital and Institute of Child Health, Lahore, Pakistan

Background: To compare the performance of the Pediatric Risk of Mortality (PRISM), the Pediatric Index of Mortality 2 (PIM 2) and Pediatric Logistic Organ Dysfunction (PELOD) scores at general pediatric intensive care unit in a developing country setting, investigating the relation between observed and predicted mortality. **Method:** A contemporary cohort study was undertaken at Pediatric Intensive Care Unit (PICU), Children's Hospital, Institute of Child Health, Lahore, Pakistan. 131 consecutive admissions fulfilling the inclusion criteria were enrolled in the study. PRISM, PIM 2 and PELOD calculations were performed as set out by original articles, using the published formulae. Statistical analysis included Standardized Mortality Rate (SMR), Hosmer Lemeshow goodness of fit test, receiver operating curve (ROC) characteristics and Spearman's correlation test. **Results:** 139 patients were admitted to PICU. 38 presented exclusion criteria. 29 (28.7%) patients died. Estimated mortality was; PRISM: 19.7(19.5%), PIM: 21.01(20.5%) and PELOD:18.4(18.3%). SMR was 1.47 (SD ± 0.19), 1.4 (SD ± 0.19) and 1.57 (SD ± 0.19), respectively. PRISM had better calibration (x^2 = 7.49, p = 0.49) followed by PIM 2 (x^2 = 9.65, p = 0.29). PIM 2 showed best discrimination with area under ROC = 0.88 (0.81-0.94) followed by PRISM 0.78 (0.67-0.89) and PELOD 0.77 (0.68-0.87). Spearman's correlation r between PRISM and PIM 2 returned 0.74 (p < 0.001). **Conclusion:** PRISM as well as PIM 2 is validated for PICU setting in Pakistani circumstances. PELOD performed poorly. PIM 2 has advantages over PRISM for stratification of patients in clinical trials.

Key words: Prognostic score; PRISM; PIM 2; PELOD; Mortality

INTRODUCTION

Pediatric intensive care unit (PICU) is an important component of tertiary pediatric care services. PICU aim at promoting qualified care for critically ill children. These units are points of major technology transfer and constitute one of the main consumers of hospital budgets [1]. There are relatively few efficiently equipped PICUs in Pakistan. Most of the units lack in technical equipments. The staff available in PICU settings is usually not adequately trained and clinical experience regarding pediatric intensive care is limited. Therefore, when patients with varying prognoses and degrees of clinical severity are being treated, the final result of employing the resources available at such units is often uncertain [2].

Intensive care scoring systems are devised to determine probable outcome of the patients being admitted into the ICU [3]. ICU scoring system provides health care administrator an outlook regarding patient prognosis. Thereby, decisions regarding cost effectiveness of financial and equipment assistance to the patient become more purposeful and focused.

ICU scoring systems are also important while conducting clinical trials to remove the bias by selecting patients with similar severity of illness. PICU scoring systems are mostly studied for developed nations settings. Data from developing nations has conflicting results [4,5,6]. PICU scoring system has to be validated for ICU setting in Pakistan, as various factors affect the general outcome of patients in respect to survival or mortality.

In PICUs, worldwide, pediatric risk of mortality (PRISM) and Pediatric index of mortality (PIM) are being used. PIM 2 and PRISM III are the newer versions of these scoring systems. Both have shown better performance according to their authors but high cost of attaining the software has limited the use of PRISM III even in developed nations. Pediatric Logistic Organ Dysfunction (PELOD) has recently been validated with good discrimination.

In this study, we compared the performance of the PRISM, PELOD and the PIM-2 at a general PICU, investigating the relationship between observed outcome (death/survival) and the mortality and survival rates estimated by the three scores. Thus we aimed to validate the best scoring system for PICU setting in Pakistan.

MATERIAL AND METHOD

This study was conducted at PICU, Children's Hospital and Institute of child health, Lahore.

This unit comprises of a 9-bedded medical ICU, 4 bedded cardiology ICU and 4 bedded surgical ICU within a 340-bedded tertiary care centre. It admits pediatric patients <18 years of age, from both medical and surgical subspecialties. There are at least

4 doctors on duty each day who answer to calls at the PICU as well as the Accident and Emergency department, thus minimizing the little delay that exists before ICU admissions. The volume of patients admitted to PICU averages about 29 patients per month.

This was a cohort study from April to June' 2006, including cases between 1 month and 18 years of age. 101 consecutive cases admitted to the PICU were studied. Cases expiring within first 8 hours or discharged before 24 hours of ICU admission were excluded from the study.

Data for calculating scores and predictive outcome were recorded prospectively and with the techniques set out for each score (PRISM, first 24 hours after admission, PIM 2, one hour after admission and PELOD at admission to the ICU). The PRISM, PELOD and PIM 2 scores were calculated using the formulae available in their original articles [7,8]. Demographic data was collected in order to characterize the sample, including age at admission, sex, and nutritional status. The outcome for all cases was documented as survival or death. Length of hospital stay at the unit was also recorded.

Simple descriptive analysis was utilized for the groups and subgroups under study (mean, median, standard deviation). Comparison of the general similarity between observed mortality and that estimated by the standardized mortality rate (SMR) was calculated [9]. For aptness of the three models, the Hosmer-Lemeshow goodness-of-fit test was employed to test the agreement between observed and expected mortality (calibration). The capacity for discrimination between survivors and moribund patients was made using the typical area under a receiver operating characteristic curve (ROC curve) and quantitative correlation between the results of the scores were analyzed using the Spearman test. Data was analyzed using Statistical Program for Social Science (SPSS Inc., Chicago, IL, USA).

RESULTS

139 patients were admitted to PICU during the duration of study. 38 patients were excluded due to discharge or expiry before 24 hours of admission. No patient was excluded due to lack of data. These included CNS diseases (n=29, 28.7%), respiratory diseases (n=19, 18.8%), hepatobiliary and gastrointestinal problems (n=12, 11.9%) and cardiovascular (n=11, 10.9%). The rest of the patients had renal, metabolic problems or septicemia. Non-surgical patients comprised 86% of all cases. Majority of patients, 56 (55.4%) had malnutrition, below 5th centile for weight for age and 35(34.7%) were below 5th centile for weight for height parameter. 29(28.7 %) patients expired out of total 101 individuals. Estimated mortality using PIM2 was 21.01 (20.5%), PRISM, 19.7 (19.5%) and PELOD, 18.4 (18.3%) respectively. This corresponds with SMR (CI. 95%) of 1.4 (SD ± 0.19) for PIM2, 1.47 (SD ± 0.19) for PRISM and 1.57 (SD ± 0.19) for PELOD.

The similarities between observed and estimated risk of mortality for the three models using x^2 goodness of fit test at five mortality risk intervals was determined for calibration (Table 2).

Table 3 shows performance of the prognostic scoring systems. PRISM had best calibration(x^2 = 7.49, p = 0.49) followed by PIM 2 (x^2 = 9.65, p = 0.29). PIM 2 showed best discrimination using area under ROC=0.88 (0.81-0.94) followed by PRISM 0.78 (0.67-0.89) and PELOD 0.77 (0.68-0.87).

PIM2 and PRISM revealed positive and significant correlation, with Spearman's correlation r=0.74 (p<0.001). PELOD also showed significant but lower correlation with r=0.69 (p<0.001) and r=0.64 (p<0.001) with PIM 2 and PRISM respectively.

Table 1. Characteristic of patients admitted to PICU

Characteristics	
Male gender	60.4% (n=61/101)
Mean age in months; mean (median)	45.66 (18.0)
Weight in kg; mean(median)	12.6 (9.4)
Duration of stay (days); mean (median)	12.16 (10.0)

DISCUSSION

In this study, the observed mortality was 28.7%. This mortality was much higher than the documented rates at other PICUs where validation of prognostic scores has been undertaken. Higher number of non surgical patients (86%) admitted to PICU was one of the main factors. Also, the level of clinical instability in admitted patients was much higher. This was evident by up to three times higher mean expected death rate as compared to previous studies [5].

Table 2. Calibration of models in five intervals of mortality risk

	RISK %	Total number	Observed survival	Expected survival	Observed mortality	Expected mortality
PIM2	0-1	16	16	15.93	0	0.07
	>1-5	23	22	22.28	1	0.72
	>5-15	23	19	20.90	4	2.10
	>15-30	11	6	8.53	5	2.47
	>30	28	9	12.67	19	15.33
	Total	101	72	80.31	29	20.69
PRISM	0-1	9	8	7.95	1	0.05
	>1-5	30	25	24.22	5	0.88
	>5-15	32	27	29.32	5	2.78
	>15-30	8	5	6.22	3	1.88
	>30	22	7	7.92	15	14.08
	Total	101	72	75.63	29	19.67
PELOD	0-1	36	33	35.91	3	0.09
	>1-5	25	17	24.63	8	0.37
	>5-15	0	0	0	0	0
	>15-30	23	17	18	6	4.59
	>30	17	5	3.61	12	13.39
	Total	101	72	82.15	29	18.44

Table 3. Performance of three PICU prognostic scoring systems

Performance of the models			
	PIM 2	**PRISM**	**PELOD**
Mean of mortality risk; % (SD)	20.49±24.72	19.49±26.21	18.26±29.99
Median of mortality; %	8.5	7.4	1.3
Estimated mortality; n	20.69	19.67	18.44
Standardized mortality rate (SMR) (CI 95%)	1.4 (0.77-2.0)	1.47(0.9-2.0)	1.57 (1.0-2.1)
Hosmer Lemeshow goodness-of-fit test; x^2 (p)	9.65 (p=0.29)	7.49 (p= 0.49)	20.03 (p=0.006)
Area under ROC (CI 95%)	0.88 (0.81-0.95)	0.78 (0.67-0.89)	0.77 (0.68-0.87)
Standard error AUC	0.035	0.056	0.05

Malnutrition was a major issue in our data with 55.4% patients below 5th centile on weight for age plot and 34.7% below 5th centile on weight for height plot. However, malnutrition could not be established as an independent prognostic factor. Similar results have been documented in a study from India [10]. Mean duration of stay for patients in PICU (12.6 days) as well as hospital stay was similar to data from most PICUs in developing countries, though higher than developed countries. The difference was mainly associated with surgical post recovery patients who generally require a short stay in intensive care setting.

In this single unit study, the individual performance of the three scoring systems underestimated general mortality using SMR (PIM 2 predicted 71.4% and PRISM 67.9%). The difference was not significant in both these scores. PELOD showed poor performance predicting only 63.8% with significant difference between observed and expected mortality (p = 0.001). SMR was not significantly higher for patients with malnutrition. It was in contrast with findings of Thukral et al.[10].

The power of calibration was tested in all the three scoring systems using Hosmer Lemeshow goodness of fit test. The value of p was required to be greater than 0.05 for good calibration of the model. Both PRISM and PIM 2 showed good calibration, the predicted results were similar to those observed. PRISM had slightly better power of calibration ($x^2 = 7.49$, $p = 0.49$, $d = 8$) than PIM 2 ($x^2 = 9.65$, $p = 0.29$, $d = 8$). PELOD showed poor calibration ($x^2 = 20.03$, $p < 0.01$, $d = 7$).

A discriminatory power of 0.90 or more is considered excellent, 0.80-0.89 as good and 0.70-0.79 as fair discriminatory performance by scoring model. The discriminatory power was evaluated using ROC curve, showing good discrimination for PIM 2 (AUC, 0.88) while fair performance by PRISM and PELOD (AUC, 0.78 and 0.77 respectively) (Figure 1). When patients with cardiac problem were excluded from study data, the discrimination was improved for all models (Figure 2). The discrimination

was found to be excellent for PIM2 (0.92), good for PRISM (0.81) but still fair for PELOD (0.79). This finding suggested that none of the models could cater for cardiac patients. Discrimination was better for PIM 2 in malnourished patients (AUC, 0.95).

In this study, we tried to validate the three scoring systems; PIM2, PRISM and PELOD. We compared the three systems for both discrimination and calibration. Although still debatable, both functions are important in validation of any generic scoring system [11]. Both functions gain importance in respective objective for which the scoring system is used. Discrimination is important while distinguishing the outcome either survival or moribund among the admitted patients. Calibration is more important while comparing expected and observed outcome at various intervals of severity. Thus discrimination and calibration are both important while validating prognostic scoring systems. Our study demonstrated PRISM showing better calibration though having only slight edge over PIM2. PIM 2, however, had a much better discriminatory power compared to PRISM. The discrimination was better in all disease spectrums. Cardiac patients showed poor discrimination for both models. Excluding cardiac patients, PIM 2 had excellent discrimination (>90% AUC). PELOD showed both lower discrimination and poor calibration.

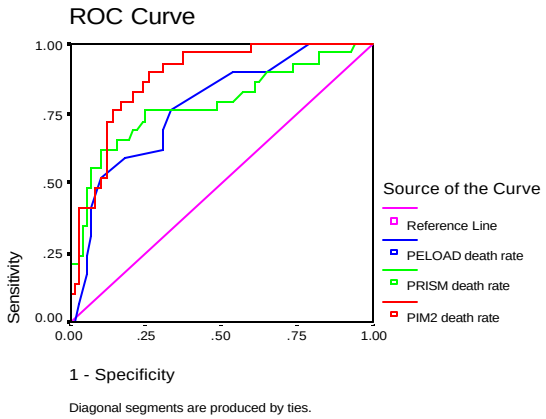

Figure 1 - Superposition of three receiver operating curves (ROC)

Area Under the Curve					
	Area	Std. Error[a]	Asymptotic Sig[b].	Asymptotic 95% Confidence Interval	
Test Result Variable(s)				Lower Bound	Upper Bound
PIM2 death rate	0.88	0.035	0.000	0.81	0.95
PRISM death rate	0.78	0.056	0.000	0.67	0.89
PELOD death rate	0.77	0.050	0.000	0.68	0.87

a Under the nonparametric assumption
b Null hypothesis: true area = 0.5

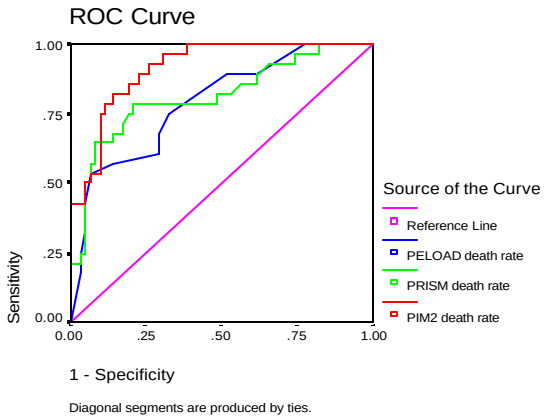

Figure 2. Superposition of three receiver operating curves (ROC) excluding cardiac patients

Area Under the Curve					
	Area	Std. Error[a]	Asymptotic Sig[b]	Asymptotic 95% Confidence Interval	
Test Result Variable(s)				Lower Bound	Upper Bound
PIM2 Death rate	0.915	0.028	0.000	0.860	0.971
PRISM Death rate	0.813	0.053	0.000	0.710	0.916
PELOD Death rate	0.792	0.051	0.000	0.692	0.892

a Under the nonparametric assumption
b Null hypothesis: true area = 0.5

PRISM and PIM 2 both showed good over all predictive performance. The results were in agreement with Martha VF[5] and Slater A[8], who also showed good performance by PRISM and PIM 2 in PICU in developing country. The correlation between PRISM and PIM 2 was better (r = 0.74) than documented between PRISM and PIM by Martha VF[5]. Thukral A[10] found very similar results while evaluating PRISM and PIM 2 in a developing country setting. The scores underestimated the mortality; SMRs(CI 95%) using PRISM and PIM 2 models were 1.20 (0.94-1.50) and 1.57 (1.24-1.59), respectively. They also documented AUC 0.80 (0.74-0.86) and 0.81 (0.75-0.87) for PRISM and PIM 2, respectively[10]. Various other studies have shown similar good discrimination but poor calibration for PRISM and PIM 2 [12]. Poor calibration has been attributed to various factors like poor performance of medical system, particularly if observed mortality is higher than expected values. This is more important in developing countries where resources are more limited. Other factors include different case mix [13], disease pattern [14] and failure of the scoring system equation to model the actual situation accurately [15].

PIM 2 had an edge over PRISM having fewer variables making assessment more convenient. As the resources are limited in developing countries like Pakistan, this could mean economically more acceptable. Moreover, assessors act as confounding factor due to improper training of physicians in PICU setting. Fewer variables would also make the uniform training of PICU staff more convenient.

Co-morbid conditions as well as diagnosis at admission to PICU invariably affect the outcome of patients. Wells et al[6] demonstrated much poor performance by PRISM score, both in terms of calibration and discrimination. Poor performance could have been due to different demographic profile, disease distribution or availability of infrastructure including trained personnel as well as equipment. He also attributes difficulty in achieving same outcome for similar level of instability but having different pathological processes. PIM 2 includes low and high risk categories which help in better discriminatory power of the model. Newer version of PRISM, PRISM III has also included the risk groups but as mentioned elsewhere, the use of this prognostic system is likely to remain limited in developing countries as long as monitory concerns persist.

Stratification for inclusion of children in clinical trials was an important concern to be evaluated in this study. Although PIM 2 and PRISM both had good performance, PIM 2 appears superior because the evaluation is done at 1hour of admission against 24hours for PRISM. If the evaluation is to be delayed for 24 hours to stratify sample in clinical trial, important time is lost and the situation becomes impracticable. Early evaluation in PIM 2 allows for commencing the intervention required in clinical trial early and effectively.

CONCLUSION

Both PRISM and PIM2 had good discrimination and calibration in PICU in Pakistani circumstances. Both predictive models are validated for given circumstances. Due to fewer variables and early calculation (at 1 hour), PIM 2 is more helpful in stratification of patients for clinical trials. PELOD has poor performance and is not validated in PICU setting in Pakistan. Malnutrition, being an important factor for consideration in developing countries, is also best addressed by PIM2.

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. 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.
3. 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.
4. Singhal D, Kumar N, Puliyel J.M, Singh S.K, Srinivas V: Prediction of mortality by application of prism score in intensive care unit. Indian Pediatrics 2001; 38: 714-719
5. Martha VF, Garcia PCR, 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.
6. Wells M, Riera-Fanego JF, Luyt DK, Dance M, Lipman: 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.
7. Pollack MM, Ruttimann UE, Getson PR: The Pediatric Risk of Mortality (PRISM) score. Crit Care Med. 1988; 16:1110-6.
8. Slater A, Shann F, Pearson G: PIM2: a revised version of the Paediatric Index of Mortality. Intensive Care Med. 2003; 29: 278-85.
9. Flora, JD: A method for comparing survival of burn patients to a standard survival curve. J Trauma. 1978;18:701-8
10. Thukral A, Lodha R, Irshad M, Arora NK: Performance of Pediatric Risk of Mortality (PRISM), Pediatric Index of Mortality (PIM), and PIM2 in a pediatric intensive care unit in a developing country. Pediatr Crit Care Med. 2006 ;(4):356-61.
11. Randolph AG, Guyatt GH, Calvin JE, Doig G, Richardson WS. Understanding articles describing clinical prediction tools. Evidence Based Medicine in Critical Care Group. Crit Care Med 1998, 26: 1603-12
12. Bertolini G, Ripamonti D, Cattaneo A, Apolone G: Pediatric risk of mortality: an assessment of its performance in a sample of 26 Italian intensive care units. Crit Care Med. 1998;(8):1427-32.
13. Pappachan JV, Millar B, Bennett ED, Smith GB: Comparison of outcome from intensive care admission after adjustment for case mix by the APACHE III prognostic system. Chest 1999; 115:802-810.
14. Markgraf R, Deutschinoff G, Pientka L, Scholten T: Comparison of acute physiology and chronic health evaluations II and III and simplified acute physiology score II: a prospective cohort study evaluating these methods to predict outcome in a German interdisciplinary intensive care unit. Crit Care Med 2000, 28:26-33.
15. Livingston BM, MacKirdy FN, Howie JC, Jones R, Norrie JD: Assessment of the performance of five intensive care scoring models within a large Scottish database. Crit Care Med 2000, 28:1820-1827.

Address for Correspondence: Dr. Ahmad Usaid Qureshi, Medical Officer(Postgraduate Trainee), Department of Paediatric Medicine, Children's Hospital and Institute of Child Health, Ferozepur Road, Lahore.
Phone: 92-42-5164842, 92-333-4485306
E-mail: qureshiahmad@yahoo.com