Arch Clin Pediatr. 2026;3(1):8–14.

Archives of Clinical Pediatrics | Original Research

Pediatric Intensive Care Mortality Prediction Using Pediatric Risk of Mortality and Acute Physiology and Chronic Health Evaluation Scores

Nabeel Ahmad[1], Uzair Ahmed[1], Bilal Aslam[1], Muhammad Talha[1], Aamir Hussain[1], Rokish Akhtar[2], Neykzad Wadeer[3,*]

[1]Lahore General Hospital, Lahore, Pakistan

[2]Hebei North University Zhangjiakou, China

[3]CURE International Hospital, Kabul, Afghanistan

[*]Correspondence should be addressed to Neykzad Wadeer, neykzad.wadeer102@gmail.com

Received date: August 29, 2025, **Accepted date:** March 10, 2026

Citation: Ahmad N, Ahmed U, Aslam B, Talha M, Hussain A, Akhtar R, et al. Pediatric Intensive Care Mortality Prediction Using Pediatric Risk of Mortality and Acute Physiology and Chronic Health Evaluation Scores. Arch Clin Pediatr. 2026;3(1):8–14.

Abstract

Background: Accurate mortality prediction in pediatric intensive care units (PICUs) is critical for clinical decision-making, benchmarking, and guiding resource allocation. While the Pediatric Risk of Mortality (PRISM) score is widely used in children, the Acute Physiology and Chronic Health Evaluation II (APACHE II), originally developed for adults, is occasionally applied in pediatric settings despite limited validation.

Objective: To compare the predictive performance of PRISM and APACHE II scores in critically ill children admitted to a tertiary PICU in Pakistan.

Methods: We conducted a retrospective cohort study of 214 patients aged 0–18 years admitted to the PICU of Lahore General Hospital over six months. PRISM and APACHE II scores were calculated within the first 24 hours of admission. Discrimination was assessed using area under the receiver operating characteristic curve (AUC), and calibration was evaluated with the Hosmer–Lemeshow test, Brier score, and error analyses.

Results: Neither PRISM (AUC = 0.656) nor APACHE II (AUC = 0.587) achieved strong discriminatory power, and differences between the two models were not statistically significant (p = 0.388). Calibration was acceptable for both models, with marginally better performance for PRISM. However, neither score independently predicted mortality in multivariate regression analyses.

Conclusion: In this large South Asian cohort, PRISM outperformed APACHE II but both scores demonstrated limited predictive accuracy. These findings highlight the limitations of applying adult-derived indices to pediatric patients and reinforce the need for pediatric-specific, locally validated prognostic tools. Our results provide important regional evidence and underscore the urgency of multicenter efforts to develop robust, resource-appropriate models for global pediatric critical care.

Keywords: Pediatrics, Intensive care units, Pediatric, Mortality, Severity of illness index, Prognosis

Introduction

Accurate mortality risk prediction is crucial in pediatric intensive care to guide clinical decision-making, allocate resources, benchmark quality, and counsel families [1]. Traditional adult ICU scores (e.g. APACHE, SAPS) were developed in the 1980s but generally perform poorly in children due to differing physiology and disease patterns [2]. Consequently, pediatric-specific models have been developed. The Pediatric Risk of Mortality (PRISM) score is a physiology-based risk model that uses 14–17 commonly measured variables (e.g. vital signs, blood gases, laboratory values) in the first hours of PICU admission to estimate mortality risk [3]. The PRISM framework has undergone several updates (PRISM III, PRISM IV) to maintain accuracy as care improves. In practice, PRISM is widely used to calculate standardized mortality ratios and to adjust for illness severity in pediatric cohorts [4]. By contrast, the Acute Physiology and Chronic Health Evaluation (APACHE)

II score was developed for adult ICU populations (1985) and combines 12 physiologic parameters plus age and chronic health points into a logistic regression model of mortality [5]. APACHE II has been extensively validated in adult cohorts, but it has not been widely applied or recalibrated for children [6,7]. A recent commentary noted that although APACHE II could be calculated in pediatric patients, only a few small studies have examined its performance, and adult-derived thresholds may not generalize to children. For example, one single-center PICU report (n=100) found APACHE II distinguished survivors from nonsurvivors (mean scores 16.6 vs 26.1) with an area under the ROC curve (AUC) of 0.889 [8]. However, APACHE II's calibration and discrimination in broader pediatric cohorts remain uncertain. Thus, there is a clear need to directly compare PRISM and APACHE II in predicting PICU mortality. Existing pediatric studies usually focus on PRISM or the Pediatric Index of Mortality (PIM), with little direct head-to-head data on APACHE II. In PICUs, PRISM (often PRISM III or IV), PIM and the Pediatric Logistic Organ Dysfunction (PELOD) score are the standard models [4]. APACHE II's adult origins and lack of pediatric recalibration leave an evidence gap. Our retrospective cohort study aims to evaluate and compare the discriminative ability and calibration of PRISM (III or IV) versus APACHE II in a modern PICU population. By quantifying each score's predictive accuracy, we seek to inform the choice of mortality risk model for pediatric critical care.

In Pakistan's resource-constrained healthcare landscape, pediatric intensive care units (PICUs) grapple with immense clinical and logistical challenges, yet limited data exist evaluating prognostic models in this context. This retrospective cohort study represents the first institutional head-to-head comparison of PRISM and APACHE II scores at Lahore General Hospital, thereby filling a critical knowledge gap in regional pediatric critical care. While prior national studies have independently validated severity indices, none have directly contrasted them within a single cohort, underscoring the novelty and relevance of our analysis. By identifying which scoring system more accurately stratifies mortality risk, our findings have tangible implications for optimizing triage, resource allocation, and outcomes in high-acuity pediatric care. The study was conducted in accordance with STROBE guidelines and received institutional ethics committee approval, with informed consent waived due to its retrospective nature—ensuring both methodological rigor and ethical integrity.

Materials and Methods

Study design and setting

This retrospective cohort study was conducted at the Pediatric Intensive Care Unit (PICU) of Lahore General Hospital, Lahore. The study spanned a period of 6 months.

Study population

Pediatric patients aged 0 to 18 years admitted to the PICU during the study period were eligible for inclusion. Inclusion criteria required that patients had complete data necessary for calculating both the Pediatric Risk of Mortality (PRISM) score and the Acute Physiology and Chronic Health Evaluation II (APACHE II) score within the first 24 hours of PICU admission. Patients were excluded if they were transferred from another intensive care unit or if essential data required to compute either scoring system were missing or incomplete.

Data collection

Data were extracted from patient medical records using a structured data collection form. Variables collected included:

- **Demographic information**: age, sex
- **Clinical characteristics**: comorbidities, reason for PICU admission
- **Physiological parameters**: heart rate, respiratory rate, mean arterial pressure, temperature, Glasgow Coma Scale, and other relevant clinical signs
- **Laboratory findings**: arterial pH, serum bicarbonate, lactate levels, and other variables as required by the PRISM and APACHE II scoring systems
- **Outcomes and interventions**: duration of PICU stay, requirement for mechanical ventilation, and in-hospital mortality

Scoring was performed retrospectively using standard criteria available at:

PRISM: https://sfar.org/scores2/prism2.php

APACHE II: https://www.mdcalc.com/calc/1868/apache-ii-score

Statistical analysis

All statistical analyses were performed using SPSS and R software. Descriptive statistics were used to summarize baseline demographic and clinical characteristics; continuous variables were presented as means ± standard deviation or medians with interquartile ranges, while categorical variables were expressed as frequencies and percentages.

The predictive performance of the PRISM and APACHE II scores was evaluated using Receiver Operating Characteristic (ROC) curve analysis. The area under the ROC curve (AUC) was calculated for each score to assess discriminatory ability. The DeLong test was used to compare the AUCs of the two models.

Calibration was assessed using the Hosmer–Lemeshow goodness-of-fit test and Brier scores to evaluate the accuracy of predicted mortality against observed outcomes. Mean absolute error (MAE) and mean squared error (MSE) were also calculated. Statistical significance was defined as a p-value <0.05.

Results

A total of 214 patients were enrolled in the study, with a nearly equal gender distribution. There were 111 females (50.7%) and 103 males (49.3%) (**Table 1**). Pearson correlation analysis revealed a very weak positive correlation between PRISM scores and mortality (r = 0.054), and a very weak negative correlation between APACHE II scores and mortality (r = –0.055). These results suggest that neither score demonstrated a strong linear association with patient outcome in this dataset. Multivariate logistic regression showed that neither the PRISM score (β = 0.0665, p = 0.521) nor the APACHE II score (β = –0.0549, p = 0.522) was a statistically significant predictor of mortality. The model intercept was statistically significant (β = –1.8839, p = 0.048). However, the overall model fit was poor (pseudo R^2 = 0.0102), and the log-likelihood ratio test was not statistically significant (p = 0.7286). ROC curve analysis revealed that the PRISM score had an area under the curve (AUC) of 0.656, while the APACHE II score had an AUC of 0.587. Although PRISM performed better numerically, DeLong's test indicated no statistically significant difference between the two AUCs (Z = 0.864, p = 0.3877; 95% CI for AUC difference: –0.087 to 0.225). The Hosmer–Lemeshow goodness-of-fit test showed acceptable calibration for both models. For PRISM, χ^2 = 6.25 (df = 7, p = 0.5107), and for APACHE II, χ^2 = 9.47 (df = 8, p = 0.3046), indicating no significant difference between observed and predicted outcomes across risk deciles. The Brier score was marginally lower for PRISM (0.123) compared to APACHE II (0.127), suggesting slightly better predictive accuracy. However, mean absolute error (MAE) and mean squared error (MSE) favored APACHE II (MAE: 0.023; MSE: 0.00068) over PRISM (MAE: 0.040; MSE: 0.00193). The 90th percentile of absolute error was also lower for APACHE II (0.044) compared to PRISM (0.058), indicating more consistent calibration. Neither PRISM nor APACHE-II scores showed a statistically significant difference between the two groups (likely deceased vs survived), based on the p-values (p >0.05 for both). Although PRISM has a larger absolute mean difference (-2.339 vs -1.507), it is still not statistically significant, indicating insufficient evidence to conclude a real difference in severity scores between the groups using this test.

Table 1. Gender distribution in the study cohort.

Gender	Frequency	Percent	Valid Percent	Cumulative Percent
Female	111	50.7%	50.7%	50.7%
Male	103	49.3%	49.3%	100.0%
Total	214	100.0%	100.0%	—

Table 2. Group statistics.

Score	Outcome Death	N	Mean	Std. Deviation	Std. Error Mean
PRISM Score	0	186	13.66	4.753	0.604
	1	28	16.00	5.586	1.684
APACHEII Score	0	186	21.13	5.097	0.647
	1	22	22.64	4.296	1.295

This table presents the mean, standard deviation (SD), and standard error of the mean (SEM) of PRISM and APACHE II scores stratified by patient outcome (0 = survived, 1 = died).

Table 3. Independent samples test.

Score	Equal Variances	Levene's F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
PRISM Score	Assumed	0.307	0.581	-1.465	71	0.147	-2.339	1.596	-5.521	0.844
	Not Assumed			-1.307	12.700	0.214	-2.339	1.789	-6.213	1.536
APACHEII Score	Assumed	0.684	0.411	-0.923	71	0.359	-1.507	1.633	-4.764	1.749
	Not Assumed			-1.041	15.462	0.314	-1.507	1.448	-4.586	1.571

This table shows the results of the independent samples t-test for comparing PRISM and APACHE II scores between survivors and non-survivors.

Table 4. Independent samples effect sizes.

Score	Standardiser	Point Estimate	95% CI Lower	95% CI Upper
PRISM Score	Cohen's d	-0.479	-1.124	0.168
	Hedges' correction	-0.474	-1.112	0.167
	Glass's delta	-0.419	-1.075	0.257
APACHEII Score	Cohen's d	-0.302	-0.944	0.342
	Hedges' correction	-0.299	-0.934	0.339
	Glass's delta	-0.351	-1.002	0.316

This table reports the effect sizes (Cohen's d, Hedges' correction, and Glass's delta) for PRISM and APACHE II scores in relation to patient mortality.

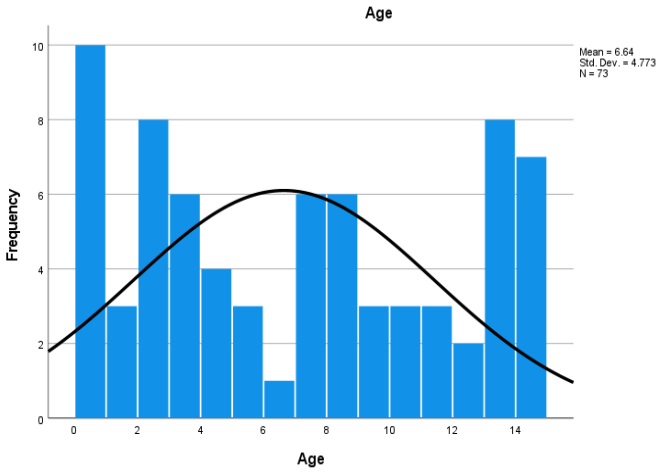

Figure 1. Age distribution histogram of PICU patients.

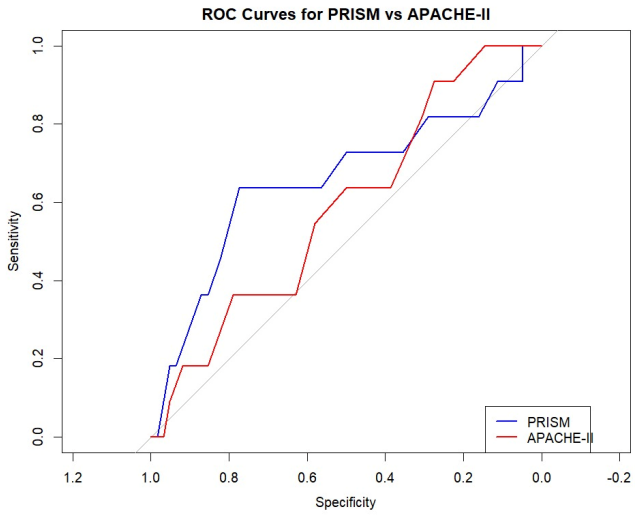

Figure 2. Receiver operating characteristic (ROC) curves comparing the predictive performance of PRISM and APACHE II scores for mortality in PICU patients.

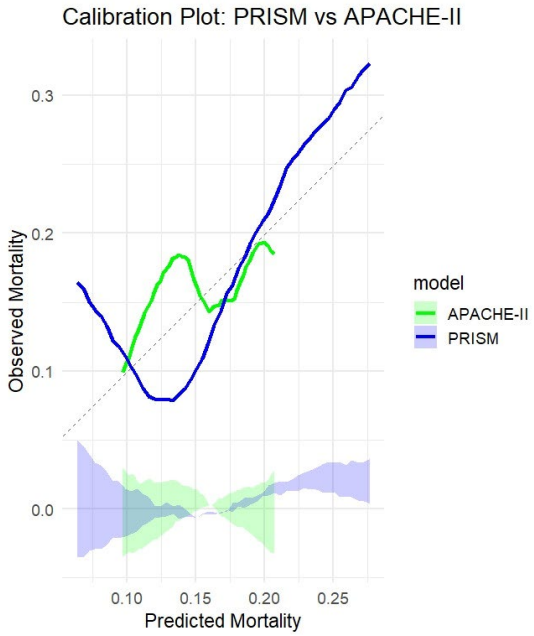

Figure 3. Calibration plot for PRISM and APACHE II scores in predicting mortality among PICU patients.

Discussion

Our study adds to the growing body of evidence that adult-derived severity indices, such as APACHE II, have limited applicability in pediatric populations. While PRISM demonstrated comparatively better calibration and discrimination, neither score achieved the level of predictive accuracy reported in larger multicenter cohorts, underscoring the importance of contextual validation. This finding is particularly relevant for low- and middle-income countries (LMICs), where resource allocation is constrained and clinicians often rely on pragmatic adaptations of available tools. By directly contrasting PRISM and APACHE II in a large South Asian cohort, our work highlights that adopting adult models without recalibration risks misclassification of mortality risk and potentially misinforms clinical decision-making in pediatric intensive care units.

Recent literature underscores that PRISM-based models generally offer good discrimination for PICU mortality, while calibration may vary across populations [9,10]. In the Brazilian multicenter cohort of Rodrigues-Santos *et al.*, PRISM exhibited an AUROC of 0.86 (95% CI 0.83–0.89), indicating strong discrimination. However, it underestimated mortality across a broad range of risk having standardized mortality ratio of 1.3:1 and demonstrated poor calibration [9]. Similarly, Alkhalifah *et al.* in a Saudi tertiary PICU reported PRISM discrimination with AUC ~0.81 [11]. A systematic review by Shen *et al.* confirmed that PRISM has good discriminatory power with pooled AUC ranging between 0.8 to 0.9 in PICU patients [10], though quality of evidence was rated low to moderate. In these studies, PRISM consistently achieved AUCs generally in the 0.80–0.90 range for mortality, comparable to or slightly higher than other pediatric scores. For example, PRISM-III showed AUC of 0.81 versus PIM-3 AUC of 0.80 in one Saudi study [11]. By contrast, literature on APACHE II in pediatric settings is sparse. As an adult metric, APACHE II is not standard in PICUs [7]. The few available pediatric data suggest APACHE II can discriminate survivors from non-survivors, but usually in small or specialized cohorts. Sankar's review notes one study reporting an AUC of 0.889 for APACHE II in 100 children [8]. However, without multicenter validation, generalization is unknown. No large PICU study has been published validating APACHE II. Adult ICU analyses have long shown APACHE II's AUC around 0.85 for general populations, but its performance on unmodified pediatric physiology is likely inferior. For example, the Pediatric Index of Mortality and PELOD-2 achieve similarly high discrimination with simpler inputs [12]. In any case, APACHE II's calibration to pediatric risk remains untested.

Clinically, PRISM scores are widely used in PICUs for research and benchmarking [4], whereas APACHE II would require adaptation. PRISM's advantage is pediatric specificity; APACHE II's potential advantage would be familiarity since adult ICUs use it but in children, chronic conditions are captured differently. Some authors argue that even newer machine-learning models or combined scores might eventually outperform static scores [13].

Our results suggest trends with modest effect sizes, highlighting the need for larger multicenter studies, they also fill an important evidence gap by providing one of the largest head-to-head comparisons of PRISM and APACHE II in pediatric critical care from an LMIC setting. The consistency of our findings with prior smaller reports strengthens the conclusion that PRISM retains greater clinical utility, while APACHE II should not be extrapolated uncritically to children. In line with previous studies, several investigations have demonstrated strong predictive performance of PRISM-based models in pediatric intensive care populations. In a multicenter prospective study conducted across four PICUs, Ekinci *et al.* evaluated multiple pediatric mortality prediction scores, including PRISM and PRISM-IV, and reported strong discriminatory performance for these models in critically ill children, supporting their utility for risk stratification in PICU settings [16]. Similarly, Kaur *et al.* assessed the performance of the PRISM III score in a tertiary PICU and found excellent discrimination for mortality prediction. In their analysis, each one-point increase in the PRISM III score was associated with a 1.25-fold increase in the odds of mortality, highlighting the strong prognostic relationship between physiologic derangement and clinical outcomes in critically ill pediatric patients [17]. Taken together, these findings from multicenter and single-center PICU cohorts reinforce the established prognostic utility of PRISM-based models, although variability in discrimination across different healthcare settings suggests that contextual validation remains essential, particularly in resource-limited environments. These findings highlight the need for pediatric-specific, locally validated risk prediction models and underscore the importance of future multicenter studies aimed at recalibrating existing scores or developing novel indices tailored to diverse healthcare settings. In doing so, our study not only cautions against the uncritical application of adult-derived scoring systems in pediatric populations but also contributes to the broader effort to improve mortality risk prediction in global pediatric critical care.

Conclusion

In this retrospective cohort of 214 critically ill children, PRISM demonstrated modestly better performance than APACHE II in predicting PICU mortality, although neither score showed strong discriminatory ability. These findings reinforce the importance of pediatric-specific severity scoring systems and highlight the limitations of applying adult-derived indices to pediatric populations without recalibration. Our study provides regional evidence from a South Asian PICU and underscores the need for locally validated prognostic tools. Future multicenter studies are warranted to refine existing models and develop context-appropriate risk prediction systems for pediatric critical care.

Conflict of Interest

None declared.

Funding

None.

Ethical Approval

This study was approved by the Institutional Review Board (IRB) of Lahore General Hospital (LGH), Lahore, Pakistan. All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

1. Niederwanger C, Varga T, Hell T, Stuerzel D, Prem J, Gassner M, et al. Comparison of pediatric scoring systems for mortality in septic patients and the impact of missing information on their predictive power: a retrospective analysis. PeerJ. 2020 Oct 5;8:e9993.
2. Pollack MM, Holubkov R, Funai T, Dean JM, Berger JT, Wessel DL, et al. The Pediatric Risk of Mortality Score: Update 2015. Pediatr Crit Care Med. 2016 Jan;17(1):2–9.
3. Sankar J. Acute physiology and chronic health evaluation II for critically ill children? Indian J Crit Care Med. 2015 Aug;19(8):446–8.
4. Rodrigues-Santos G, Prata-Barbosa A, Lima-Setta F, Silami PH, de Oliveira MB, Robaina JR, et al. Performance of Pediatric Risk of Mortality IV in Brazilian PICUs: A Multicenter Prospective Study. Critical Care Explorations. 2025 Apr 1;7(4):e1243.
5. Alkhalifah AS, AlSoqati A, Zahraa J. Performance of Pediatric Risk of Mortality III and Pediatric Index of Mortality III Scores in Tertiary Pediatric Intensive Unit in Saudi Arabia. Front Pediatr. 2022 Jul 7;10:926686.
6. Shen Y, Jiang J. Meta-Analysis for the Prediction of Mortality Rates in a Pediatric Intensive Care Unit Using Different Scores: PRISM-III/IV, PIM-3, and PELOD-2. Front Pediatr. 2021 Aug 24;9:712276.
7. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985 Oct;13(10):818–29.
8. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. Crit Care Med. 1988 Nov;16(11):1110–6.
9. 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 Feb;23(2):201–7.
10. Straney L, Clements A, Parslow RC, Pearson G, Shann F, Alexander J, et al. Paediatric index of mortality 3: an updated model for predicting mortality in pediatric intensive care*. Pediatr Crit Care Med. 2013 Sep;14(7):673–81.
11. Leteurtre S, Duhamel A, Salleron J, Grandbastien B, Lacroix J, Leclerc F; Groupe Francophone de Réanimation et d'Urgences

Pédiatriques (GFRUP). PELOD-2: an update of the PEdiatric logistic organ dysfunction score. Crit Care Med. 2013 Jul;41(7):1761–73.

12. 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 Sep;5(5):447–54.
13. Goldstein B, Giroir B, Randolph A; International Consensus Conference on Pediatric Sepsis. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. Pediatr Crit Care Med. 2005 Jan;6(1):2–8.
14. 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 Feb;23(2):201–7.
15. Teetz AK, Sarnaik AP, Taylor SL. PELOD-2: A new pediatric organ dysfunction score. Pediatr Crit Care Med. 2013;14(7):660–72.
16. Ekinci F, Yildizdas D, Horoz OO, Arslan I, Ozkale Y, Yontem A, et al. Performance and analysis of four pediatric mortality prediction scores among critically ill children: A multicenter prospective observational study in four PICUs. Arch Pediatr. 2022 Aug;29(6):407–14.
17. Kaur A, Kaur G, Dhir SK, Rai S, Sethi A, Brar A, et al. Pediatric Risk of Mortality III Score - Predictor of Mortality and Hospital Stay in Pediatric Intensive Care Unit. J Emerg Trauma Shock. 2020 Apr-Jun;13(2):146–50.