



## OPEN ACCESS

## EDITED BY

Tyler John Loftus,  
University of Florida, United States

## REVIEWED BY

Melissa Aczon,  
Children's Hospital of Los Angeles,  
United States  
Daniel Ehrmann,  
Hospital for Sick Children, University of  
Toronto, Canada

## \*CORRESPONDENCE

Anita K Patel  
apatel4@childrensnational.org

<sup>†</sup>These authors have contributed equally to this work

<sup>‡</sup>These authors have contributed equally to this work and share first authorship

## SPECIALTY SECTION

This article was submitted to Pediatric Critical Care, a section of the journal Frontiers in Pediatrics

RECEIVED 23 August 2022

ACCEPTED 26 October 2022

PUBLISHED 01 December 2022

## CITATION

Patel AK, Trujillo-Rivera E, Morizono H and Pollack MM (2022) The criticality Index-mortality: A dynamic machine learning prediction algorithm for mortality prediction in children cared for in an ICU. *Front. Pediatr.* 10:1023539. doi: 10.3389/fped.2022.1023539

## COPYRIGHT

© 2022 Patel, Trujillo-Rivera, Morizono and Pollack. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The criticality Index-mortality: A dynamic machine learning prediction algorithm for mortality prediction in children cared for in an ICU

Anita K Patel<sup>1\*††</sup>, Eduardo Trujillo-Rivera<sup>2†</sup>, Hiroki Morizono<sup>3††</sup> and Murray M. Pollack<sup>1††</sup>

<sup>1</sup>Department of Pediatrics, Division of Critical Care Medicine, Children's National Health System, George Washington University School of Medicine and Health Sciences, Washington, DC, United States, <sup>2</sup>Department of Bio-Informatics, Children's National Health System, George Washington University School of Medicine and Health Sciences, Washington, DC, United States, <sup>3</sup>Department of Pediatrics, Children's National Research Institute, George Washington University School of Medicine and Health Sciences, Washington, DC, United States

**Background:** The Criticality Index-Mortality uses physiology, therapy, and intensity of care to compute mortality risk for pediatric ICU patients. If the frequency of mortality risk computations were increased to every 3 h with model performance that could improve the assessment of severity of illness, it could be utilized to monitor patients for significant mortality risk change.

**Objectives:** To assess the performance of a dynamic method of updating mortality risk every 3 h using the Criticality Index-Mortality methodology and identify variables that are significant contributors to mortality risk predictions.

**Population:** There were 8,399 pediatric ICU admissions with 312 (3.7%) deaths from January 1, 2018 to February 29, 2020. We randomly selected 75% of patients for training, 13% for validation, and 12% for testing.

**Model:** A neural network was trained to predict hospital survival or death during or following an ICU admission. Variables included age, gender, laboratory tests, vital signs, medications categories, and mechanical ventilation variables. The neural network was calibrated to mortality risk using nonparametric logistic regression.

**Results:** Discrimination assessed across all time periods found an AUROC of 0.851 (0.841–0.862) and an AUPRC was 0.443 (0.417–0.467). When assessed for performance every 3 h, the AUROCs had a minimum value of 0.778 (0.689–0.867) and a maximum value of 0.885 (0.841,0.862); the AUPRCs had a minimum value 0.148 (0.058–0.328) and a maximum value of 0.499 (0.229–0.769). The calibration plot had an intercept of 0.011, a slope of 0.956, and the  $R^2$  was 0.814. Comparison of observed vs. expected proportion of deaths revealed that 95.8% of the 543 risk intervals were not statistically significantly different. Construct validity assessed by death and survivor risk trajectories analyzed by mortality risk quartiles and 7 high and low risk diseases confirmed *a priori* clinical expectations about the trajectories of death and survivors.

**Conclusions:** The Criticality Index-Mortality computing mortality risk every 3 h for pediatric ICU patients has model performance that could enhance the clinical assessment of severity of illness. The overall Criticality Index-Mortality

framework was effectively applied to develop an institutionally specific, and clinically relevant model for dynamic risk assessment of pediatric ICU patients.

#### KEYWORDS

prediction algorithm, dynamic outcome prediction, mortality prediction, criticality index, machine learning

## Introduction

Models computing mortality risk have not been traditionally used for individual patient risk assessment. Static scores such as PRISM and PIM that utilize fixed time periods for data collection at the beginning on the intensive care unit stay (ICU) stay were developed and calibrated on populations of interest primarily for hospital quality assessment and benchmarking; they were not designed to aid in the assessment of individual patients or to enhance clinical decision making (1, 2). Since they are calibrated to produce a risk score at a single time point, they fail to provide updated assessments reflecting the changing clinical status of individual patients; some have warned against their use for individual patients (3–5). When the data are collected or risk is computed at the bedside, there are data reliability issues as well as an additional risk of human error. However, there is a long-standing desire to have a frequently updated patient assessment methodology as evidenced by investigations of intermittently updated static scoring systems (6–10).

Physiologic patterns often precede acute deterioration of patients and may go unrecognized (11–15). This is the foundation of the recent emphasis on early warning systems (15). An automated, frequently updated assessment of a patient's severity of illness could alert health care providers to a patient's changing status prior to clinical recognition (16). For ICU patients, serious deteriorations often follow a sequence of both deteriorations and improvements until there is overt decompensation, rather than a linear progression to decompensation (14). Alerts that identify an unrecognized problem or reinforce a provider's suspicions could result in monitoring or therapy changes that mitigate mortality and morbidity risks.

Despite the increasing sophistication of monitoring and therapeutic technologies, and the data-rich environment of the ICU, current monitoring strategies in clinical use are still challenged to accurately assess changing patient status. The aim of this study was to assess the performance of a dynamic method of updating mortality risk every 3 h using the Criticality Index-Mortality (CI-M) neural network methodology.

## Methods

The data came from Children's National Hospital, a 323-bed academic hospital with 48 bed Pediatric ICU beds and 24

Cardiac ICU beds. The patient sample included all admissions to the Pediatric and Cardiac ICUs from January 1, 2018 to February 29, 2020. The end date was chosen to reflect care prior to the novel coronavirus pandemic since ICU admission and discharge guidelines and isolation practices were in flux for much of this time. Exclusions included patients over 21 years of age and those admitted to the neonatal ICU. This study was approved by the Institutional Review Board (protocol Pro00015931). Modeling was performed under the direction of Eduardo Trujillo-Rivera PhD.

## Dataset

The dataset was extracted from the electronic health record (EHR). Descriptive data included age, gender, dates of hospital and ICU admission, admission type (with elective, emergent, and urgent), diagnoses, race, and hospital outcome (survival, death) during or following an ICU admission. Only data from the first ICU admission was included for patients with multiple admissions. Modeling included all independent variables from the original multi-institutional CI-M model representing physiology, treatment, and care intensity. (Trujillo Rivera et al., 2022)<sup>17</sup> This included age, gender, 30 laboratory tests, 6 vital signs, 1113 and medications classified into 143 categories. Positive pressure ventilation (invasive and non-invasive) was identified by procedure codes, and specific times with and without ventilation were identified. Variables reflecting positive pressure ventilation included the number of hours of mechanical ventilation at the time of risk assessment, the number of hours since discontinuation at the time of risk assessment, and the proportion of time that positive pressure ventilation used. Medication data were extracted from the medication administration record using start and end times. Medication classes were categorized using the Multum™ system (18). A comprehensive list of independent variables and details of medication classification have been published and included in **Supplementary Appendices S1, S2** (17). Other data included diagnostic information categorized by the International Classification of Diseases 10th Edition (ICD-10) (19). Diagnoses were used for descriptive purposes, but not for modelling, because they were determined at discharge. All discharge diagnoses for each patient were categorized by system of dysfunction (**Table 1**).

**TABLE 1 Population characteristics** the mann whitney U test was used to compare the distribution of numerical quantities between survivors and deaths. The Santner and Snell exact test was used to compare proportions (percentages) between survivors and deaths. A single encounter could have diagnoses in multiple groups.

Characteristic	All	Survivors	ICU Deaths	P-values <sup>d</sup>
N	8,399	8,087	312	
Age (Months) <sup>a</sup>	38 (12, 130)	37 (11,130)	50 (1,132)	0.034
Female ( <i>n</i> (%))	3,789 (45.1)	3,650 (45.13)	139 (44.55)	0.841
Hospital LOS (hours) <sup>a</sup>	83.10 (52.33,151.79)	82.50 (51.88,147.06)	143.97 (65.78,323.60)	<0.001
ICU LOS (hours) <sup>a</sup>	35 (20, 69)	34 (20, 67)	78 (43, 178)	<0.001
Admission Type <sup>b</sup>				
Elective	2,450 (29.2)	2,399 (29.66%)	51 (16.35%)	<0.001
Emergency	5,666 (67.5)	5,427 (67.11%)	239 (76.60%)	<0.001
Urgent	282 (3.4)	261 (3.23%)	21 (6.73%)	0.225
Hospital Mortality	312 (3.7)			
Positive Pressure Ventilation, <i>n</i> (%)	3,118 (37.1)	2,882 (35.64)	236 (75.64)	<0.001
Congenital Disease <i>n</i> (%)	1,975 (23.5%)	1,891 (23.4%)	84 (26.8%)	0.243
Systems of Dysfunction <sup>c</sup> <i>n</i> (%)				
Respiratory	4,781 (56.9%)	4,563 (56.4%)	218 (69.4%)	<0.001
Endocrine, Nutritional, Metabolic, and Immune	1,343 (16.0%)	1,249 (15.4%)	94 (29.9%)	<0.001
Gastrointestinal	843 (10.0%)	770 (9.5%)	73 (23.2%)	<0.001
Infectious	1,127 (13.4%)	1,052 (13.0%)	75 (23.9%)	<0.001
Injury and Poisoning	971 (11.6%)	898 (11.1%)	73 (23.2%)	<0.001
Neurological	2,177 (25.9%)	2,034 (25.2%)	143 (45.5%)	<0.001
Neoplasms	434 (5.2%)	377 (4.7%)	57 (18.1%)	<0.001
Hematologic	609 (7.3%)	541 (6.7%)	68 (21.7%)	<0.001
Cardiovascular	1,235 (14.7%)	1,113 (13.8%)	122 (38.9%)	<0.001
Musculoskeletal	663 (7.9%)	636 (7.9%)	27 (8.6%)	0.799
Dermatologic	359 (4.3%)	330 (4.1%)	29 (9.2%)	0.073
Genitourinary	516 (6.1%)	455 (5.6%)	61 (19.4%)	<0.001
Ophthalmologic	238 (2.8%)	210 (2.6%)	28 (8.9%)	0.0286
ENT	182 (2.2%)	181 (2.2%)	1 (0.3%)	0.504
Psychiatric	568 (6.8%)	540 (6.7%)	28 (8.9%)	0.437

LOS, length of stay.

<sup>a</sup>Median (25<sup>th</sup> percentile, 75<sup>th</sup> percentile).

<sup>b</sup>Emergency refers to all admissions/transfers that are classified as needing the ICU. Urgent refers to all patients who were intubated or actively decompensating that required immediate transfer/admission to the ICU. These metrics were supplied for descriptive purposes, and not used for modeling.

<sup>c</sup>A single encounter could have diagnoses in multiple groups.

<sup>d</sup>Survivors and deaths were compared using univariate analysis.

The ICU course was discretized into consecutive 3-hour time periods truncated at 11.4 days (273 h) when the sample size was reduced to 97 survivors and 10 deaths to ensure an appropriate sample for calibration testing. Deaths could have occurred during or following the ICU admission; however, they had to occur during the specific hospital admission to be included as a death in the study population presented in **Table 1**. We randomly selected 75% of patients for training, 13% for validation, and 12% for testing. The training set was used for model development, the validation set was used to fine-tune parameters and prevent over-fitting, and the test set was used to evaluate model performance. Counts of survivors and deaths per 3-hour time period in the training/validation

and test set are provided in **Supplementary Appendix S3** for reference.

Vital signs, laboratory data, and medication data were standardized to values from 0 to 1 using the maximum and minimum values of the training set. Consistent with other machine learning models, the data for each time period were forward imputed using the last available data if new data were not obtained (16, 17, 20–23). For the first time period, if vital signs or laboratory data were not obtained, we used the medians of the first time periods across all training patients adjusted by 9 age groups. The imputed values by age groups are reported in **Supplementary Appendix S4**. For modeling, imputed values were identified by setting the count to zero.

Vital signs and laboratory values for each time period were summarized using the count, averages, standard deviations, maximums, and minimums. Medication administration data included the number of medications by class but not dose. Medication history was summarized in each time period by the number of medications in a class, the proportion of time periods the medication class was previously administered, and the number of administered medications within the class.

## Machine learning methodology

A single neural network for classification of hospital outcome as survival or death was developed. The model used variables up to 6 h in the past for all patients if these data were available. The objective of each training epoch was to maximize the Mathew's correlation coefficient at a cut point of 0.5 while minimizing the binary cross entropy between the predicted score and the patient's outcome. The initial neural network had a single hidden layer, and a logit output. We consecutively increased the number of nodes while monitoring the Mathew's correlation coefficient, sensitivity, specificity, precision, and negative predictive value at cut points of 0.15, 0.5, and 0.9 in the training and validation sets. If overfitting was detected, regularization included L2 regularization and layer node dropouts with parameters tuned to maintain similar metrics on the training and validation sets. We increased the number of nodes while monitoring for overfitting. When there were no additional gains on the performance metrics, we added another hidden layer and repeated the process, increasing the number of hidden layers until our regularization attempts were unsuccessful in avoiding overfitting. The best model as measured by the Mathew's correlation coefficient without overfitting was kept. We assessed models using up to four previous time periods and found that performance was maximized with the current time period and the change between the present and previous time period.

The final model had three hidden layers with 32 nodes each, bias and ReLU activation functions, L2 regularization with a parameter 0.02, and drop out nodes of 0.05. The neural network was calibrated to risk of mortality using a logistic regression with covariates including the logit transformed score of the neural network.

## Statistics

The performance was assessed in the test sample and only test sample data are reported. Discrimination was assessed by the area under the receiver operating characteristic curve (AUROC) and the area under the precision recall curve (AUPRC) for all time periods combined and as function

of time (24). AUPRC was computed with integral approximation. The model was calibrated using the training and validation datasets. Calibration was assessed on the test dataset with calibration plots of the differences between the observed and expected proportions of deaths in the risk intervals. The composite calibration plot had 543 risk intervals with a minimum of 49 patients in each risk interval. We computed the regression line for observed vs. expected proportions of deaths and tested the difference between proportions in each interval with the Santner and Snell exact test and report the percentage of intervals with no statistical difference (25, 26). Optimal calibration plot performance includes intercept = 0, slope = 1,  $R^2 = 1$ , and  $\leq 5.0\%$  of risk intervals with a statistically difference ( $p < 0.05$ ) between the observed and expected proportions. In addition, we computed the additional performance metrics of precision, number needed to evaluate, accuracy, and negative predictive value at a cut point of 0.5 and performed a net benefit analysis (27) for the proposed model of both treated and untreated patients.

Construct validity was assessed by plotting population trajectories and mortality risk changes in consecutive time periods for selected groups. First, we plotted the mortality risk trajectories for survivors and deaths in the total sample for specific mortality risk groups. Mortality risk groups were determined from the first 3-hour time period using mortality risk quartiles. A priori based on analyses of clinical trajectories from the national Criticality Index-Mortality model (28), we expected that the mortality risk of deaths in the highest risk quartile would remain high and other death groups would increase or remain relatively constant over time and the mortality risk of survivors in the highest survivor risk quartile group would decrease and the risk of other survivor risk groups would decrease or remain relatively constant over time. Second, we plotted the mortality risk trajectories using diagnostic groups with traditionally low (diabetic ketoacidosis, bronchiolitis) and high risk (congenital cardiovascular conditions, traumatic brain injury, bone marrow transplant, pediatric acute respiratory distress syndrome). We hypothesized that deaths in the low-risk diagnoses would follow a similar clinical trajectory to deaths in the low-risk quartile – they would increase overtime, while survivors would maintain relatively low and constant mortality risks. Whereas, deaths in the high-risk diagnoses would rapidly increase their mortality risks or begin at a high mortality risk and remain high, while survivors' mortality risk trajectories would remain relatively constant or improve over time (28).

## Variable importance

We explored the variable importance using a Local Interpretable Model-Agnostic Explanation (LIME) approach (LIME R package) (29). Briefly, the LIME approach assumes

that every model performs like a linear prediction model for each prediction and the hierarchy of covariate importance is preserved. The collection of individual linear models provides an interpretation of the covariate importance in the final model. We ranked the most important variables across all predictions by computing the percentage of times each variable was among the 30 most important covariates.

## Results

There were 8,399 pediatric ICU admissions with 312 (3.7%) deaths (Table 1). Admissions to the ICU were 29.2% elective, 67.5% emergent, and 3.4% urgent. The most common systems of dysfunction were respiratory, cardiovascular, neurologic, infectious, and endocrine/nutrition/metabolic/immune. Median age was 38.1 months (IQR 11.6–129.9), 45.1% were female; median hospital length of stay (LOS) was 83.1 (IQR 52.3–151.8), ICU LOS was 35.3 (IQR 20.01–69.1), and 37.1% of patients received positive pressure ventilation. Deaths were more likely to be older, emergent admissions with longer LOS's, and receive positive pressure ventilation.

The composite assessment of discrimination across all time periods found an AUROC of 0.851 (0.841–0.862) and a AUPRC was 0.443 (0.417–0.467) (Figure 1A,B respectively). The AUROCs and AUPRC (Figure 1C) were also calculated for every 3-hour time period. The AUROCs had a minimum value of 0.778 (0.689–0.867) at hour 3 and a maximum value of 0.885 (0.841,0.862) at hour 81 and the individual AUPRCs had a minimum value 0.148 (0.058–0.328) at hour 3 and a maximum value of 0.499 (0.229–0.769) at hour 195. Additional performance metrics are reported in Supplementary Appendix S5, Table J. Most notable is the precision of the model at 0.594, with a number needed to evaluate of 1.7, accuracy of 94.7% and negative predictive value of 0.958. The net benefit analysis also demonstrates the potential positive impact (Supplementary Appendix S6: Net Benefit Analysis of the Criticality Index-Mortality Models for both Treated and Untreated Patients). The calibration plot had an intercept of 0.011, a slope of 0.956, and the  $R^2$  was 0.814 (Figure 2). Comparison of the observed vs. expected proportion of deaths in the risk intervals revealed that 95.8% of the 543 risk intervals were not statistically significant (Figure 2).

Construct validity assessed by death and survivor risk trajectories (Figure 3A) revealed the *a priori* expectations of risk trajectories were correct (Methods). Risks for deaths in the highest risk quartile were high and remained high and risks for deaths in the other 3 risk quartiles increased over time. All survivor risk quartiles had mortality risks that remained relatively constant over time. Construct validity assessed by the mortality risk trajectories of different disease pathologies (Figure 3B) followed the expected pattern of deaths having higher mortality risks than survivors. Mortality

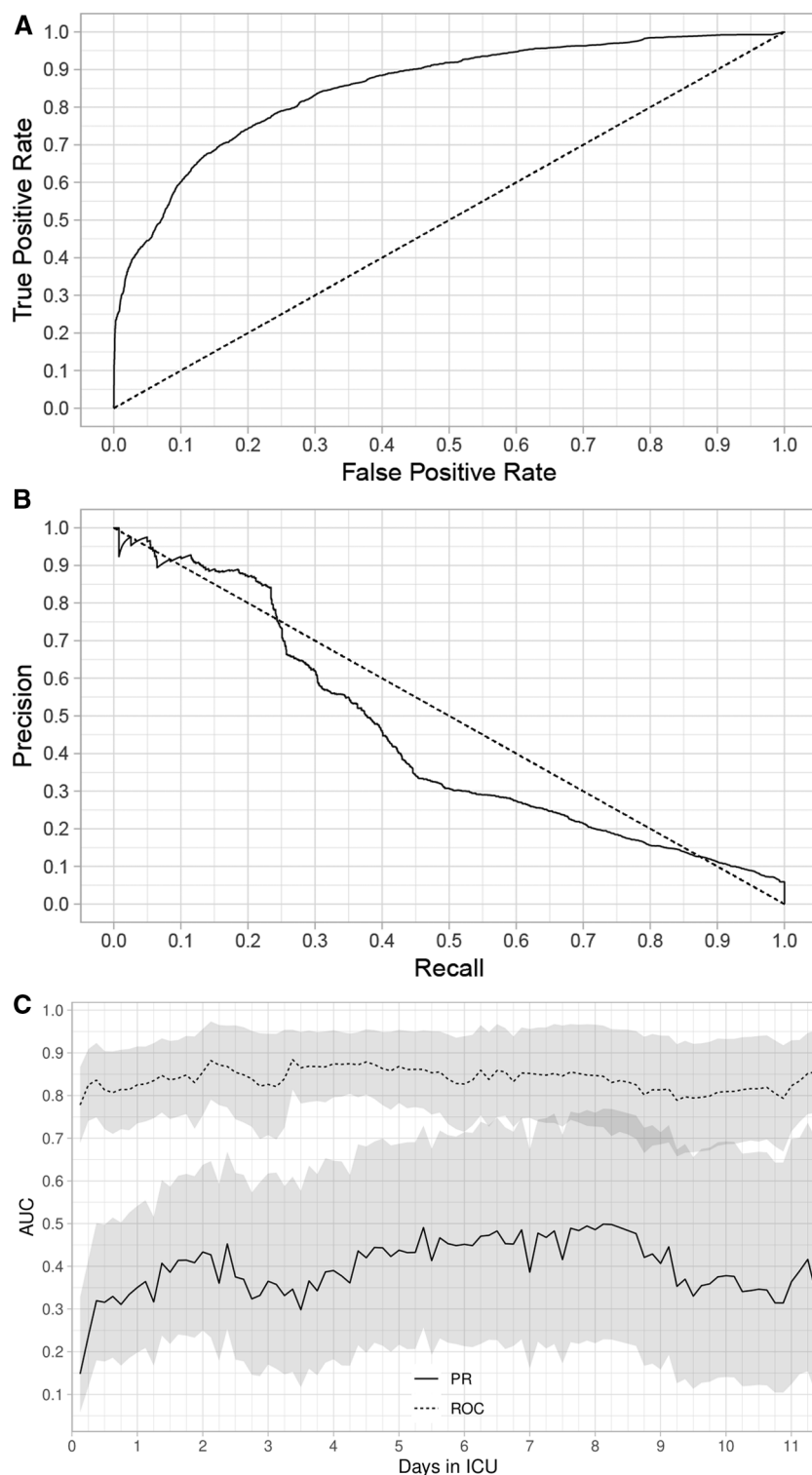
risk trajectories exhibited different patterns across differing pathologies that were consistent with clinical experience.

The 20 most frequently important covariates for risk prediction using the LIME approach are shown in Figure 4. The 10 most frequently important variables were the length of time in the ICU prior to the risk prediction, duration since mechanical ventilation was discontinued, hours on mechanical ventilation, the maximum, minimum and average coma scores, age, and the minimum, average, and maximum neutrophil counts. Other variables of frequent importance included medication classes (anti-infectives, anti-neoplastics, general anesthetics, respiratory inhalants and nasal lubricants). Overall, these variables reflect disease pathologies such as neurologic conditions (coma scores), respiratory conditions (mechanical ventilation, respiratory inhalants, and lubricants).

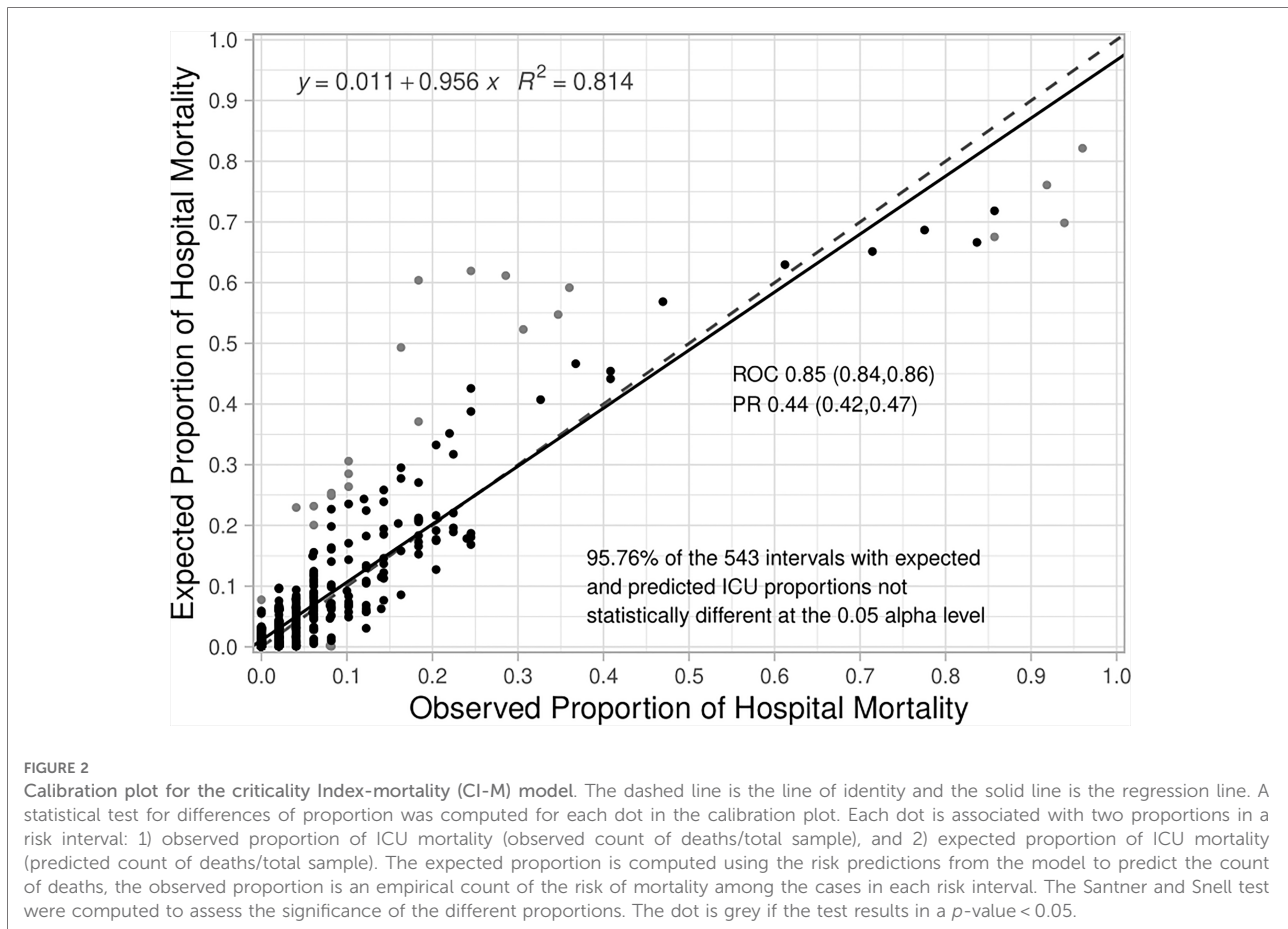
## Discussion

The CI-M, a neural network which uses physiology, therapy, and intensity of care to compute mortality risk for pediatric ICU patients, was effectively used to develop an institution-specific and clinically relevant model for mortality risk assessment every 3 h. Model performance, while it varied over time, indicated potential to enhance clinical assessment of mortality risk in the PICU. Discrimination assessed across all time periods found an AUROC of 0.851 (0.841–0.862) and a AUPRC was 0.443 (0.417–0.467). The calibration plot had an intercept of 0.011, a slope of 0.956, and the  $R^2$  was 0.814. Comparison of the observed vs. expected proportion of deaths in 543 the risk intervals revealed that 95.8% were not statistically significant. Construct validity assessed by death and survivor risk trajectories analyzed by mortality risk quartiles and by 7 high and low risk diseases confirmed risk patterns consistent with clinical expectations. Importantly, the worst performance was during the initial time intervals when historical data was less available, and during the later times when less data was available for modeling. These issues would be amenable to improvement as more institutional data, including patients and patient variables, are acquired. Additionally, since the evaluation cohort is constantly changing as patients leave the ICU, a larger data set would also enable a more complete assessment of performance, including the characteristics of patients remaining in the ICU at different time points coupled with focused performance metrics for those time points.

The approach taken to develop an institutional-specific model was straightforward and applicable to institutions with a range of data sciences expertise, including those without research expertise. We used the core variable list and modeling methodology used to develop the CI-M. Since the original CI-M was developed from a national database with a limited number of data elements, we were able to make minor



**FIGURE 1**  
 Area under the receiver operating characteristic (AUROC) curves (A), area under the precision recall (AUPRC) curves (B), and AUROC and AUPRC computed every 3 h. Outcomes were survival and death. (A) The AUROC for all time periods through 11.4 days. (B) The AUPRC for all time periods through 11.4 days. (C) AUROC and AUPRC calculated every 3 h through 11.4 days. The shaded regions are pointwise 95% confidence intervals. We applied boost strap techniques to the test set with 5,000 stratified bootstrap replicates to compute the confidence intervals.

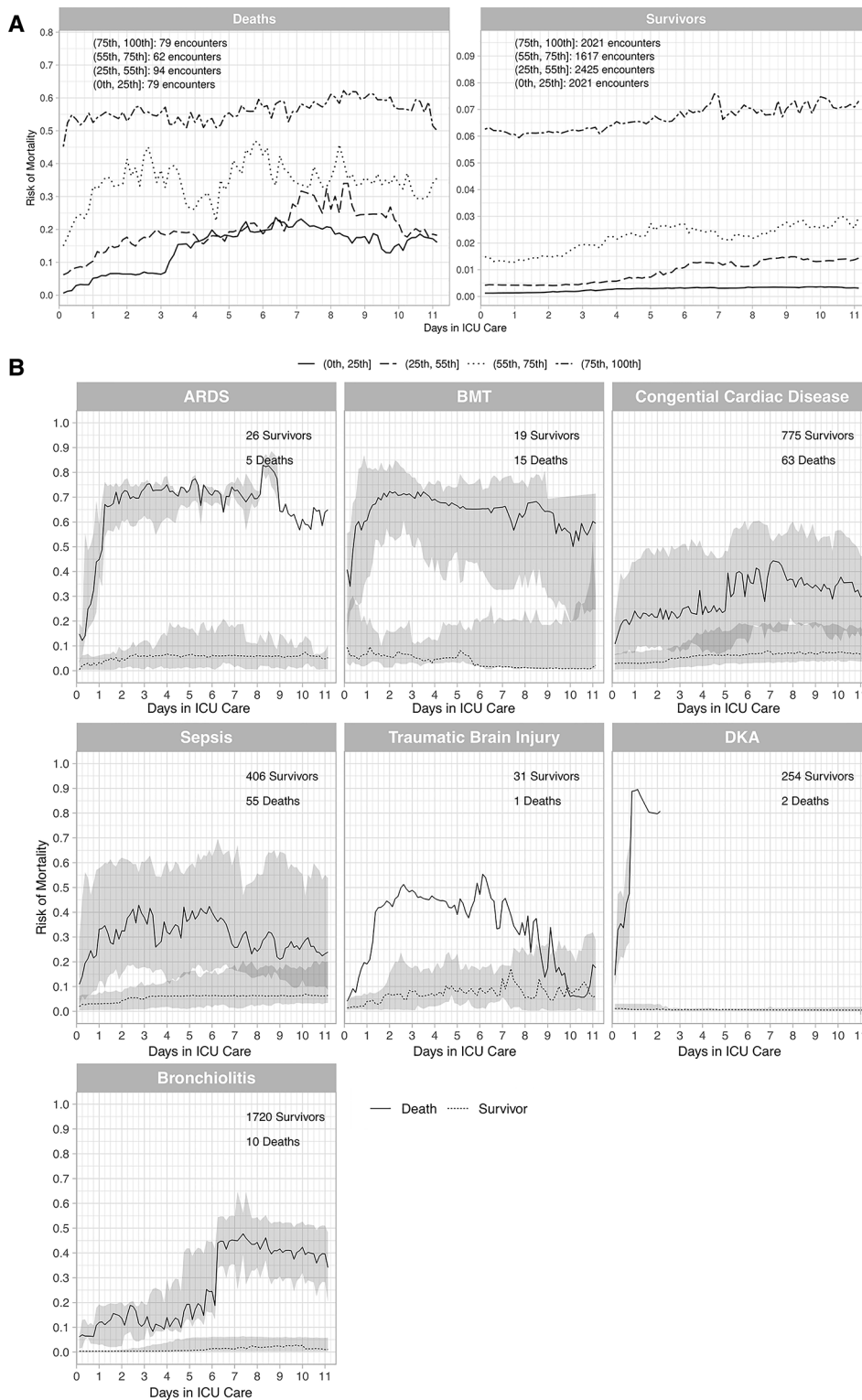


enhancements to the dataset relevant to our institution. Developing an institution specific model also enabled us to fine-tune risk computations to outcomes that might have been influenced by institution-specific clinical practice (30–32). Developing institution-specific models has the further potential advantage for “continuous improvement” by adjusting variables and outcomes as institution-specific practices change and adding variables to improve performance.

While neither current risk assessment nor prediction methods have significantly enhanced the ability of bedside caregivers to recognize early patterns of deterioration (33), the methodology described in this article has the potential to identify patients by their changing risk status who are deteriorating, improving, or remaining stable. If successful, it could improve clinical decision-making by supplementing the limitations of cognitive processing and by reducing medical errors (34). Medical errors are often based in heuristics and are more likely to occur in high-pressure, high-stakes decisions, particularly when dealing with incomplete information, such as assessing a deteriorating patient (35–37). The response to risk change will need to be individualized based on the patient’s condition, clinical expectations, and the magnitude of the change. For example, some patients will

need immediate evaluation with a change in therapeutic plans while others may be considered for ICU discharge because of the documented improvement.

Health care personnel benefiting most from measures of risk change are expected to be those with less experience or training. Experienced intensivists are excellent at assessing patients using clinical snapshots and their risk assessments for discharge outcomes are generally accurate and reliable (38–40). The wide-spread use of the multiple static models used to estimate objective mortality risks (2, 41, 42) might have helped to calibrate clinical assessments to objective risk predictions for discharge outcomes. Unfortunately, even experienced physicians’ clinical judgements have not been calibrated to assess dynamic risk changes that occur over short time intervals. The ability and reliability of health care professions, no matter what their training or experience, to assess these small changes is unknown. Contiguous changes in risk for ICU patients are generally small (17) and may be dependent on a small changes in a large number of variables. The ability to successfully integrate this large amount of changing information on a frequent basis lies beyond the capabilities of most care givers, especially those who are less experienced (38). Furthermore, the model’s high precision of 0.594 resulting in a number needed to



**FIGURE 3**

Trajectories for deaths and survivors by risk quartiles (A), and trajectories for deaths and survivors by diagnostic groups (B). (A) Death and survivor risk trajectories. Risk was computed with the Criticality Index-Mortality (CI-M). Median risk for each quartile for survivors and deaths used CI-M predictions. Patients were stratified using the first risk prediction at 3 h. The shaded areas are 95% confidence intervals (CIs). The trajectories were constructed from the total sample with 314 deaths and 8,087 survivors. (B) Death and survivor risk trajectories for different conditions. Risk was computed with the CI-M predictions and displayed for deaths and survivors. The numbers of survivors and deaths are shown in each panel. ARDS = Acute Respiratory Distress Syndrome (ARDS), BMT = Bone Marrow Transplants, DKA = Diabetic Ketoacidosis (DKA).

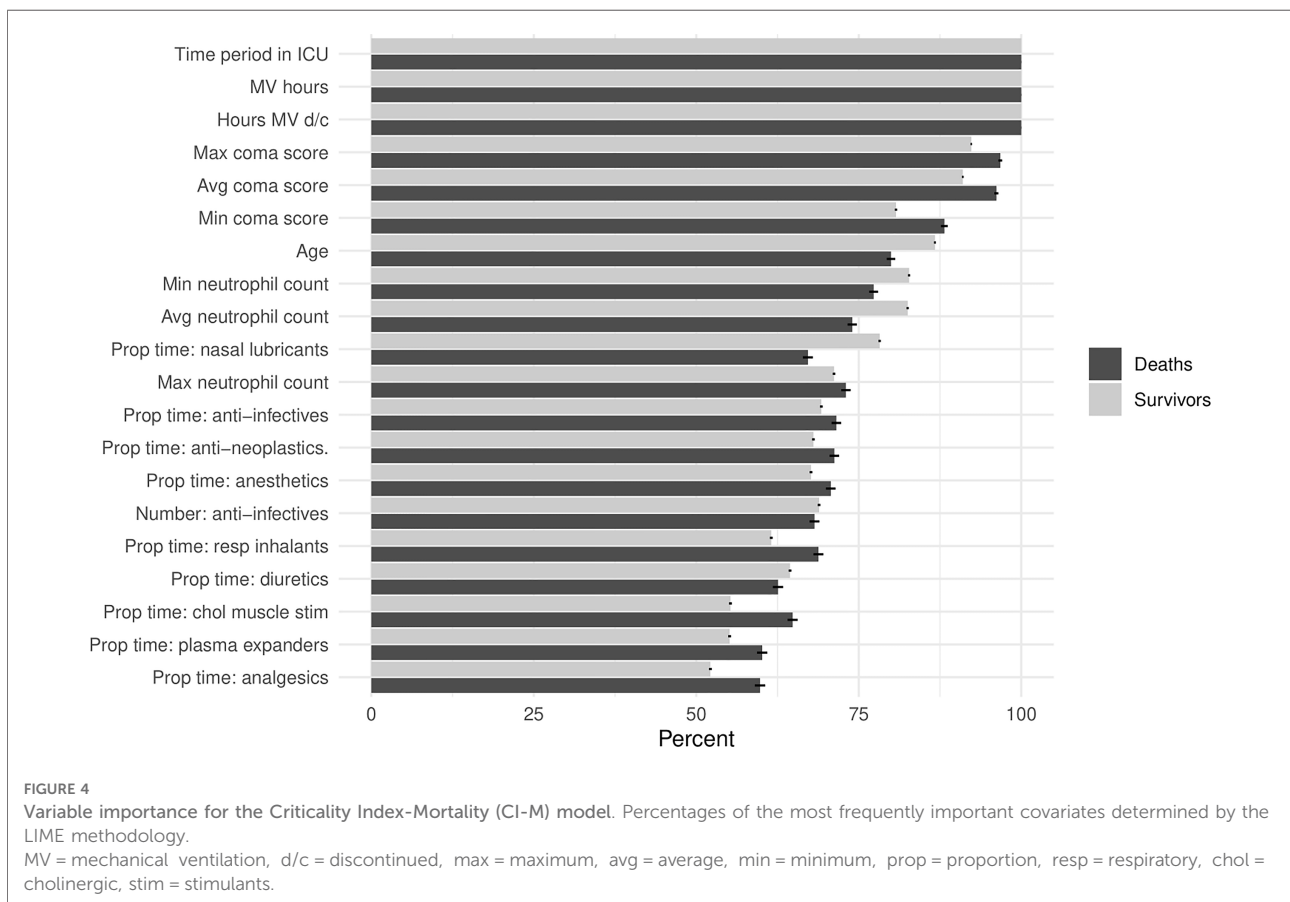


evaluate of 1.7 patients coupled with an accuracy of 94.7% suggests that any potential burden of increased patient assessments to an ICU care team is negligible with a high benefit if a model-detected clinical deterioration can lead to earlier clinical intervention. The potential benefit is confirmed by the net benefit analysis (Supplementary Appendix S6: Net Benefit Analysis of the Criticality Index-Mortality Models for both Treated and Untreated Patients). Conversely, with a high negative predictive value of 0.958, the model can potentially be utilized to direct clinical reassessments away from low risk patients, and towards those that need them most; this potential model utilization is particularly important during periods of high census. Therefore, the addition of frequently updated risk assessments for children in ICUs could result in the detection of clinical deterioration or improvement that might have been unappreciated, providing an opportunity for earlier interventions and the potential for improved outcomes.

Construct validity in this study and others was assessed using clinical trajectories, overall and within clinical entities, which were consistent with clinical expectations (17, 28). The validity for the Criticality Index, however, has been further explored. First, the variables based on therapy, physiology and intensity used in the Criticality Index are fundamental to severity of illness. In the 1960's and 1970's, The Clinical Classification

System was proposed as a measure of severity based on intensity of care and the Therapeutic Intervention Scoring System, a measure of severity based on therapies, was an early quantitative assessment of severity of illness (43). In the 1980's and 1990's, physiology-based systems were developed and are the early iterations of the APACHE and PRISM systems (44, 45). Second, the Criticality Index is highly correlated with patients' care areas that are representative of severity of illness (22). Third, clinical instability measured with the Criticality Index demonstrates increased volatility in deaths compared to survivors, consistent with the observations of Yoon et al. (14). And fourth, changes in the Criticality Index over time periods as short as 6 h strongly correlate with clinical deterioration and clinical improvement (17).

There are limitations to this study. First, while there is value in an institution specific algorithm, there is a limitation because single site datasets will usually have a relatively small sample size with the potential for overfitting the model. A larger dataset could have improved our model performance, especially during the time periods of prolonged ICU stays. Second, we did not exhaust all issues and possible modeling approaches available for machine learning. For example, data missingness is a routine issue for many machines learning approaches and, although we did not find any systematic bias,



it has the potential to introduce bias and could have influenced the LIME analysis. Third, we were unable to explore the details of the risk trajectories for the specific disease processes because most high-risk conditions are infrequently represented in multi-disciplinary pediatric ICUs. Fourth, we did not use diagnoses for our risk assessments because these were not easily available in the EHR for the individual time intervals. However, the LIME assessment of important variables demonstrated how inclusion of medications and therapies provided proxies for this information. Fifth, this study, as well as other pediatric ICU studies with mortality as the outcome, will usually have relatively few deaths. The AUROC has limited relevance with very unbalanced datasets and some recommend using the AUPRC under these circumstances (24). Sixth, the data set pre-dated the novel coronavirus pandemic and, if used clinically, the algorithm should be re-evaluated and potentially be retrained. This emphasizes the need to frequently evaluate and re-calibrate predictive models, especially if they used for clinical care. Finally, performance metrics for models such as the CI-M that frequently update risk should be assessed with the understanding that the evaluation cohort is changing in time. For example, patients dying early in their ICU stay during the resuscitation phase of illness will generally have substantial physiologic instability while those dying after long ICU stays may have an illness profile of chronic dysfunction.

In conclusion, the Criticality Index-Mortality computing mortality risk every 3 h for pediatric intensive care unit patients has potential to enhance the clinical assessment of severity of illness and clinical care. The methods are applicable to other institutions.

## Data availability statement

The datasets presented in this article are not readily available because; The data can be identified to patient cases, and is therefore not appropriate for sharing. Requests to access the datasets should be directed to; [apatel4@childrensnational.org](mailto:apatel4@childrensnational.org).

## Ethics statement

The studies involving human participants were reviewed and approved by The Children's National Institutional Review

## References

1. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. *Crit Care Med.* (1996) 24:743–52. doi: 10.1097/00003246-199605000-00004
2. 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

Board. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

## Author contributions

All authors contributed equally to this work. AKP and ET-R share first authorship for this manuscript. All authors contributed to the article and approved the submitted version.

## Funding

Funding was provided by the national Institute for Child Health and Human Development K23 Award number HD105978-01. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute for Child Health and Human Development or the National Institutes of Health.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fped.2022.1023539/full#supplementary-material>.

pediatric intensive care\*. *Pediatr Crit Care Med.* (2013) 14:673–81. doi: 10.1097/PCC.0B013E31829760CF

3. Sauthier M, Landry-Hould F, Leteurtre S, Kawaguchi A, Emeriaud G, Jouvet P. Comparison of the automated pediatric logistic organ dysfunction-2 versus manual pediatric logistic organ dysfunction-2 score for critically ill

- children. *Pediatr Crit Care Med.* (2020) 21:E160–9. doi: 10.1097/PCC.0000000000002235
4. Vincent JL, Opal SM, Marshall JC. Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions. *Crit Care Med.* (2010) 38:283–7. doi: 10.1097/CCM.0B013E3181B785A2
  5. Booth FV, Short M, Shorr AF, Arkins N, Bates B, Qualy RL, et al. Application of a population-based severity scoring system to individual patients results in frequent misclassification. *Crit Care.* (2005) 9:R522. doi: 10.1186/CC3790
  6. Badawi O, Liu X, Hassan E, Amelung PJ, Swami S. Evaluation of ICU risk models adapted for use as continuous markers of severity of illness throughout the ICU stay. *Crit Care Med.* (2018) 46:361–7. doi: 10.1097/CCM.0000000000002904
  7. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the electronic medical record. *J Biomed Inform.* (2013) 46:837–48. doi: 10.1016/j.jbi.2013.06.011
  8. Rothman MJ, Tepas 3rd JJ, Nowalk AJ, Levin JE, Rimar JM, Marchetti A, et al. Development and validation of a continuously age-adjusted measure of patient condition for hospitalized children using the electronic medical record. *J Biomed Inform.* (2017) 66:180–93. doi: 10.1016/j.jbi.2016.12.013
  9. Leteurtre S, Duhamel A, Deken V, Lacroix J, Leclerc F. Daily estimation of the severity of organ dysfunctions in critically ill children by using the PELOD-2 score. *Crit Care.* (2015) 19:1–6. doi: 10.1186/S13054-015-1054-Y
  10. Leteurtre S, Duhamel A, Grandbastien B, Proulx F, Cotting J, Gottesman R, et al. Daily estimation of the severity of multiple organ dysfunction syndrome in critically ill children. *CMAJ.* (2010) 182:1181–7. doi: 10.1503/CMAJ.081715
  11. Berlot G, Pangher A, Petrucci L, Bussani R, Lucangelo U. Anticipating events of in-hospital cardiac arrest. *Eur J Emerg Med.* (2004) 11:24–8. doi: 10.1097/00063110-200402000-00005
  12. Moss TJ, Lake DE, Calland JF, Enfield KB, Delos JB, Fairchild KD, et al. Signatures of subacute potentially catastrophic illness in the ICU: model development and validation. *Crit Care Med.* (2016) 44:1639–48. doi: 10.1097/CCM.0000000000001738
  13. Hillman KM, Bristow PJ, Chey T, Daffurn K, Jacques T, Norman SL, et al. Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med.* (2002) 28:1629–34. doi: 10.1007/S00134-002-1496-Y
  14. Yoon JH, Pinsky MR. Predicting adverse hemodynamic events in critically ill patients. *Curr Opin Crit Care.* (2018) 24:196–203. doi: 10.1097/MCC.0000000000000496
  15. Romero-Brufau S, Whitford D, Johnson MG, Hickman J, Morlan BW, Therneau T, et al. Using machine learning to improve the accuracy of patient deterioration predictions: mayo clinic early warning score (MC-EWS). *J Am Med Inform Assoc.* (2021) 28:1207–15. doi: 10.1093/JAMIA/OCAA347
  16. Aczon MD, Ledbetter DR, Laksana E, Ho LV, Wetzel RC. Continuous prediction of mortality in the PICU: a recurrent neural network model in a single-center dataset. *Pediatr Crit Care Med.* (2021) 22:519–29. doi: 10.1097/PCC.0000000000002682
  17. Rivera EA T, Chamberlain JM, Patel AK, Morizono H, Heneghan JA, Pollack MM. Dynamic mortality risk predictions for children in ICUs: development and validation of machine learning models. *Pediatr Crit Care Med.* (2022) 23:344–52. doi: 10.1097/PCC.0000000000002910
  18. UMLS Metathesaurus - MMSL (Multum) - Synopsis. n.d. <https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MMSL/> (accessed July 24, 2018).
  19. Center for Health Statistics N. ICD-10-CM Official Guidelines for Coding and Reporting 2018.
  20. Trujillo Rivera EA, Chamberlain JM, Patel AK, Zeng-Treitler Q, Bost JE, Heneghan JA, et al. Predicting future care requirements using machine learning for pediatric intensive and routine care inpatients. *Crit Care Explor.* (2021) 3:e0505. doi: 10.1097/CCE.0000000000000505
  21. Ma J, Lee DKK, Perkins ME, Pisani MA, Pinker E. Using the shapes of clinical data trajectories to predict mortality in ICUs. *Crit Care Explor.* (2019) 1:e0010. doi: 10.1097/CCE.0000000000000010
  22. Rivera EAT, Patel AK, Chamberlain JM, Workman TE, Heneghan JA, Redd D, et al. Criticality: a new concept of severity of illness for hospitalized children. *Pediatr Crit Care Med.* (2021) 22:e33–43. doi: 10.1097/PCC.0000000000002560
  23. Mohamadlou H, Panchavati S, Calvert J, Lynn-Palevsky A, Le S, Allen A, et al. Multicenter validation of a machine-learning algorithm for 48-h all-cause mortality prediction. *Health Informatics J.* (2020) 26:1912–25. doi: 10.1177/1460458219894494
  24. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE.* (2015) 10:e0118432. doi: 10.1371/journal.pone.0118432
  25. Mehrotra DV, Chan ISF, Berger RL. A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. *Biomet.* (2003) 59:441–50. doi: 10.1111/1541-0420.00051
  26. Silva Mato A, Martín Andrés A. Simplifying the calculation of the P-value for Barnard's Test and its derivatives. *Stat Comput.* (1997) 7:137–43. doi: 10.1023/A:1018573716156
  27. Rousson V, Zumbo T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. *BMC Med Inform Decis Mak.* (2011) 11:1–9. doi: 10.1186/1472-6947-11-45
  28. Rivera EAT, Patel AK, Zeng-Treitler Q, Chamberlain JM, Bost JE, Heneghan JA, et al. Severity trajectories of pediatric inpatients using the criticality Index. *Pediatr Crit Care Med.* (2021) 22:e19–32. doi: 10.1097/PCC.0000000000002561
  29. Pedersen TL, Benesty M. Local Interpretable Model-Agnostic Explanations 2021.
  30. Meert KL, Keele L, Morrison W, Berg RA, Dalton H, Newth CJL, et al. End-of-life practices among tertiary care PICUs in the United States: a multicenter study. *Pediatr Crit Care Med.* (2015) 16:e231–8. doi: 10.1097/PCC.0000000000000520
  31. Lago PM, Piva J, Garcia PC, Troster E, Bouso A, Sarno MO, et al. End-of-life practices in seven Brazilian pediatric intensive care units. *Pediatr Crit Care Med.* (2008) 9:26–31. doi: 10.1097/01.PCC.00000298654.92048.BD
  32. Brajer N, Cozzi B, Gao M, Nichols M, Revoir M, Balu S, et al. Prospective and external evaluation of a machine learning model to predict in-hospital mortality of adults at time of admission. *JAMA Netw Open.* (2020) 3:e1920733. doi: 10.1001/jamanetworkopen.2019.20733
  33. Hayes MM, Chatterjee S, Schwartzstein RM. Critical thinking in critical care: five strategies to improve teaching and learning in the intensive care unit. *Ann Am Thorac Soc.* (2017) 14:569–75. doi: 10.1513/AnnalsATS.201612-1009AS
  34. Kahneman D. *Thinking, fast and slow.* New York: Farrar, Straus and Giroux (2001).
  35. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* (1979). (1974) 185:1124–31. doi: 10.1126/science.185.4157.1124
  36. Croskerry P. From mindless to mindful practice—cognitive bias and clinical decision making. *N Engl J Med.* (2013) 368:2445–8. doi: <https://doi.org/10.1056/NEJMp1303712>
  37. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Qual Saf.* (2013) 22(Suppl 2):ii58–64. doi: 10.1136/bmjqs-2012-001712
  38. Gutierrez G. Artificial intelligence in the intensive care unit. *Crit Care.* (2020) 24:1–9. doi: 10.1186/S13054-020-2785-Y
  39. Marcin JP, Pollack MM, Patel KM, Sprague BM, Ruttimann UE. Prognostication and certainty in the pediatric intensive care unit. *Pediatrics.* (1999) 104:868–73. doi: 10.1542/peds.104.4.868
  40. 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. doi: 10.1097/00003246-200008000-00050
  41. Karam O, Demaret P, Duhamel A, Shefler A, Spinella PC, Stanworth SJ, et al. Performance of the Pediatric logistic organ dysfunction-2 score in critically ill children requiring plasma transfusions. *Ann Intensive Care.* (2016) 6:1–7. doi: 10.1186/S13613-016-0197-6
  42. 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) 17:2–9. doi: 10.1097/pcc.0000000000000558
  43. Cullen DJ, Civetta JM, Briggs BA, Ferrara LC. Therapeutic intervention scoring system: a method for quantitative comparison of patient care. *Crit Care Med.* (1974) 2:57–60. doi: 10.1097/00003246-197403000-00001
  44. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med.* (1981) 9:591–7. doi: 10.1097/00003246-198108000-00008
  45. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. *Crit Care Med.* (1988) 16:1110–6. doi: 10.1097/00003246-198811000-00006