



# Machine Learning for Predicting Critical Events Among Hospitalized Children

Sierra Strutz, MS; Huan Liang, MS; Kyle Carey, MPH; Fereshteh Bashiri, PhD; Priti Jani, MD, MPH; Emily Gilbert, MD; Julie L. Fitzgerald, MD; Nicholas Kuehnel, MD; Maya Dewan, MD, MPH; L. Nelson Sanchez-Pinto, MD, MBI; Dana Edelson, MD, MS; Majid Afshar, MD, MS; Matthew Churpek, MD, MPH, PhD; Anoop Mayampurath, PhD

## Abstract

**IMPORTANCE** Unrecognized deterioration among hospitalized children is associated with a high risk of mortality and morbidity. The current approach to pediatric risk stratification is fragmented, as each hospital unit (emergency, ward, or intensive care) uses different tools for predicting specific outcomes.

**OBJECTIVE** To develop a machine learning model for the early detection of deterioration across all units, thereby enabling a unified risk assessment throughout the patient's hospital stay.

**DESIGN, SETTING, AND PARTICIPANTS** This retrospective cohort study used data from pediatric (age <18 years) admissions to inpatient and intensive care units at 3 tertiary care academic hospitals. Data were analyzed from January 2024 to March 2025.

**MAIN OUTCOMES AND MEASURES** The primary outcome was critical events, defined as invasive mechanical ventilation, administration of vasoactive medications, or death within 12 hours of an observation.

**RESULTS** The cohort included 135 621 patients (mean [SD] age, 7 [6] years; 60 376 [44.5%] female). Patient age, hospital unit, vital signs, laboratory results, and prior comorbidities were used to derive a regression-based model, an extreme gradient-boosted machine (XGB) model, and 2 deep learning models. Data from 2 hospitals were used as a derivation cohort, while patients in the third hospital constituted the hold-out external test cohort. The XGB model was the best-performing machine learning model, outperforming 2 existing ward-focused models in terms of discrimination (*C* statistic: XGB, 0.86; ward-focused models, 0.82 [*P* < .001] and 0.70 [*P* < .001]) and the number needed to alert (at an example 80% sensitivity: XGB, 6 ward-focused models: 9 and 11). The deep learning models did not exhibit improved performance. The XGB model performed better or equivalent to models trained for a specific hospital unit.

**CONCLUSIONS AND RELEVANCE** This retrospective cohort study describes the development of a novel hospitalwide model for continuously predicting the risk of critical events through the entirety of a child's stay. The model facilitated a unified framework for risk assessment in a pediatric hospital.

JAMA Network Open. 2025;8(5):e2513149. doi:10.1001/jamanetworkopen.2025.13149

## Introduction

Physiological decompensation in hospitalized children that requires the initiation of mechanical ventilation or administration of vasoactive medications increases the risk for mortality.<sup>1-6</sup> Survivors of these events remain at risk for long-term functional or neurodevelopmental impairment extending months after hospital discharge.<sup>7-12</sup> As many of these events are unrecognized,<sup>13</sup> early identification of at-risk children and timely intervention are necessary to improve outcomes.<sup>14,15</sup>

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

JAMA Network Open. 2025;8(5):e2513149. doi:10.1001/jamanetworkopen.2025.13149

## Key Points

**Question** Can a hospitalwide machine learning model accurately predict critical events among hospitalized children across emergency, ward, and intensive care units?

**Findings** In this cohort study including data from 135 621 pediatric patients, a gradient-boosted machine learning model demonstrated superior performance in predicting hospitalwide critical events, defined as mechanical ventilation, administration of vasoactive drugs, or mortality, compared with clinical standards and other machine learning models. The gradient-boosted machine learning model also showed equivalent or better performance than models trained for a specific hospital unit.

**Meaning** These findings suggest that a gradient-boosted machine learning model can continuously assess risk for children as they progress through their hospital stay, potentially improving outcomes for children.

+ [Invited Commentary](#)

+ [Supplemental content](#)

Author affiliations and article information are listed at the end of this article.

Risk prediction models have been developed to detect early signs of deterioration in hospitalized children and serve as a foundation for clinical decision support (CDS) tools. However, risk stratification in a pediatric hospital is splintered across unit-specific silos, each focused on distinct outcomes.<sup>16</sup> For example, tools in the pediatric emergency department (ED) primarily predict triage or conditions, such as sepsis.<sup>17</sup> In contrast, studies in the pediatric intensive care unit (ICU) target estimating the risk of mortality,<sup>18</sup> organ dysfunction,<sup>19,20</sup> and resuscitation events.<sup>21</sup> On the pediatric ward, tools, such as the Bedside Pediatric Early Warning System (Bedside PEWS),<sup>22,23</sup> our machine learning-based pediatric Calculated Assessment of Risk and Triage (pCART) tool,<sup>24</sup> and other scores,<sup>25,26</sup> have been developed to predict the risk of deterioration. The availability of these disparate models across hospital units causes clinicians to be presented with different risk scores targeted toward various outcomes, which shift with a patient's location within the hospital.<sup>16</sup> Moreover, implementing these specialized models outside their intended hospital unit may diminish their utility.<sup>27-29</sup> Overall, pediatric hospitals use a compartmentalized approach to risk stratification that could lead to fragmented care, hindering a cohesive assessment of deterioration through a child's hospital stay.

The objective of this study is to develop a new model for the prediction of a critical event, defined as invasive mechanical ventilation, vasoactive medication administration, or in-hospital death, within 12 hours of any vital sign or laboratory result during a child's hospitalization. We hypothesize that a machine learning model trained on multicenter electronic health record (EHR) data will accurately predict critical events in children during external validation.

## Methods

This cohort study was approved by institutional review boards at each participating site with a waiver of informed consent because it was determined to be of minimal risk. We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis plus Artificial Intelligence (TRIPOD+AI) reporting guidelines for reporting the development and evaluation of our models.<sup>30</sup>

### Study Population

We conducted a retrospective study of all pediatric (age <18 years) admissions to the ward and the ICU across 3 tertiary care centers: University of Chicago Comer Children's Hospital (hereafter, *UC*; 2009-2019), Loyola University Ronald McDonald Children's Hospital (hereafter, *Loyola*; 2006-2020), and the American Family Children's Hospital at University of Wisconsin-Madison (hereafter, *UW-Madison*; 2009-2021). Birth encounters, including neonatal ICU admissions, were excluded. Data for this study were extracted from each hospital's EHR (EPIC Systems) database.

### Outcome and Predictors

The primary outcome of interest was a critical event, defined as a binary composite outcome of invasive mechanical ventilation, vasoactive drug administration, or in-hospital mortality, within 12 hours of a vital sign or laboratory result observation in the ward, ED, or ICU.<sup>1,2,31</sup> The primary outcome was censored if the patient experienced mechanical ventilation or was administered vasoactive medications within the prior 24 hours. Outcomes in the operating room were not considered in this study. However, the censoring logic was still implemented, in that all outcomes in the 24-hour period of a mechanical ventilation or vasoactive medication event in the operating room were censored.

Only ward, ED, or ICU observations were used for model derivation and validation. Model variables included age, location of the patient (ED, ward, or ICU), vital signs, the fraction of inspired oxygen (FiO<sub>2</sub>), neurological assessment measure using the alert-verbal-pain-unresponsive scale, count of prior comorbidities (categorized as 0, 1, or >1), and results from standard laboratory tests. The complete variable list, distribution, and percentage of admissions with missing values across all sites are provided in eTable 1 in [Supplement 1](#). These measures were chosen for their ubiquity across

units, routine capture, and prior evidence.<sup>24</sup> Comorbidities were calculated using the pediatric complex chronic condition criteria.<sup>32,33</sup> The delivery method of  $\text{FiO}_2$  was not included as a variable due to inherent inconsistency in coding across sites and the added complexity of practice changes that cause the delivery method and documentation to vary over time. Descriptive statistics (*t* tests for age, Wilcoxon rank-sum tests for length of stay, and  $\chi^2$  for categorical variables) were used to compare patients who experienced the primary outcome with those who did not at each site.

### Derivation of Prediction Models

Data were split by geographic location into derivation cohort (UC and Loyola) and a hold-out external test cohort (UW-Madison). We further split the derivation data longitudinally into temporal derivation (2017 and earlier) and temporal validation (after 2017). Similar to prior work,<sup>24</sup> we used a discrete-time survival analysis framework to create a regularized logistic regression (LR) model and an extreme gradient-boosted machine (XGB) model.<sup>34-36</sup> Our approach is illustrated in eFigure 1 in Supplement 1 and explained in the eMethods and eTable 2 in Supplement 1. The window of 12 hours for predicting critical events was chosen to maintain consistency with prior work.<sup>24,37</sup> The temporal validation and the external test cohorts were not blocked when evaluating model performances. The LR model was regularized using ridge and least absolute shrinkage and selection operator regularization approaches.

We also created 2 recurrent neural network (RNN) models to explore if deep learning models perform well in predicting critical events 12 hours in advance. First, we created an RNN that used the same variables as our non-deep learning models as input, thereby testing the possibility that using longitudinal information within individual variables would improve performance. Second, we created a separate RNN that used the output of the hyperparameter-optimized XGB model and the difference from the most recent time measurement as input (referred to as the XGB-RNN tandem model). This allowed us to test whether the trend of a risk score could be used to improve the performance of predicting critical events compared with the risk score (eMethods in Supplement 1).

Details regarding hyperparameter optimization and addressing of missing values are provided in the eMethods in Supplement 1. As baseline models, we used a modified version of Bedside PEWS (ie, only using vital sign measurements) and pCART.

### Statistical Analysis

Final models were used to calculate predicted probabilities in the temporal validation and external test datasets. Model-predicted probabilities were generated for every new observation of a vital sign reading or a laboratory result. The primary metric for assessing model performance was discrimination, evaluated using the area under the receiver operating characteristic curve (AUC) on the temporal validation and external test cohorts. Model AUCs (including 95% CIs) were compared using the nonparametric DeLong method.<sup>38</sup> Additionally, we evaluated the area under the precision-recall curve of all models in the temporal validation and external test data. We then analyzed the efficiency curve of the best-performing model within the external test data by plotting the percentage of observations crossing a chosen alert threshold across different sensitivity values within the ED, ward, and ICU. We also compared the sensitivity and specificity for the best-performing model and pCART at various model thresholds on the external test cohort. Within the external test cohort, we also compared the number needed to alert (NNA) to detect a true positive at different sensitivities for the XGB model, pCART, and Bedside PEWS, calculated using maximum values per hospital admission, similar to prior work.<sup>24</sup> The NNA is calculated as the inverse of the positive predictive values. At higher sensitivities, the total number of alerts will increase. However, the rate of false positives will also increase, causing a reduction in positive predictive value and a subsequent increase in the NNA. We estimated the overall variable importance for the best-performing model using information gain. Importance was normalized to the most important variable. Variable importance was also assessed in a single-patient instance using Shapley values,

which measure the contribution of each variable for the prediction for a single observation, calculated from the DALEX package.<sup>39,40</sup>

To determine whether the best-performing hospitalwide model had better performance in predicting critical events compared with models developed using unit-specific data, we developed 3 unit-specific XGB models (eMethods in Supplement 1). Finally, we conducted a sensitivity analysis of the best-performing hospitalwide model within patient subgroups based on year of admission (before vs during the COVID-19 pandemic), patient age, and number of prior comorbidities. Analyses were conducted from January 2024 to March 2025. All analyses were performed using R software version 4.4.0 (R Project for Statistical Computing) and Python version 3.9.18 (keras 3.7.0, tensorflow 2.18.0). A 2-sided  $P < .05$  threshold was used to assess significance.

## Results

### Study Population

Our cohort included 135 621 patient admissions (mean [SD] age, 7 [6] years; 60 376 [44.5%] female) across all 3 sites. Among 55 059 patients in the UC cohort, 2541 patients (4.6%) experienced at least 1 critical event during their hospital stay. Of 38 913 patients in the Loyola cohort, 779 patients (2.0%) experienced at least 1 critical event during their hospital stay. The UW-Madison cohort, ie, our external test dataset, included 41 649 patients, of whom 2543 patients (6.1%) experienced at least 1 critical event. The derivation dataset had a total of 701 817 twelve-hour blocks, of which 3449 (0.5%) had a positive primary outcome, 615 391 (87.7%) had a negative primary outcome, and 82 977 (11.8%) were censored. **Table 1** compares the patient characteristics and hospital outcomes for all patients with and without the primary outcome.

### Model Performances

**Table 2** reports model performances on our temporal validation and external test datasets. All hospitalwide models outperformed the modified Bedside PEWS score and our previously developed pCART model in predicting the outcome. The XGB model outperformed the LR model (temporal validation: AUC, 0.89 [95% CI, 0.89-0.89] vs 0.83 [95% CI, 0.83-0.83];  $P < .001$ ; external test: AUC, 0.86 [95% CI, 0.86-0.87] vs 0.84 [95% CI, 0.84-0.84];  $P < .001$ ). The XGB model also outperformed the 2 ward-focused models in the internal (Bedside PEWS: AUC, 0.67 [95% CI, 0.67-0.67];  $P < .001$ ; pCART: AUC, 0.79 [95% CI, 0.78-0.79];  $P < .001$ ) and external (Bedside PEWS: AUC, 0.70 [95% CI, 0.70-0.70];  $P < .001$ ; pCART: AUC, 0.82 [95% CI, 0.81-0.82];  $P < .001$ ) test cohorts. Henceforth, we will refer to this model as the pediatric Critical Event Risk Evaluation and Scoring Tool (pCREST). The deep learning models (the RNN and tandem XGB-RNN) did not have improved AUC performance over pCREST in either evaluation cohort (Table 2). The pCREST model outperformed all models in terms of area under the precision-recall curve in both cohorts (eTable 3 in Supplement 1).

**Figure 1** illustrates the top 20 most important variables for pCREST ranked in descending order of importance. The most important features included the  $\text{FiO}_2$ , hospital unit location, heart rate, temperature, and respiratory rate. Platelet count, white blood cell count, and glucose measurements demonstrated high importance among laboratory values.

eFigure 2 in Supplement 1 illustrates variable importance, as indicated by Shapley values, for pCREST scores (indicated as percentile risk between 0 and 100) across several time points for a hospital stay for a patient aged 16 years as they were triaged in the ED, admitted to the ward, and subsequently moved to the ICU before experiencing a critical event. eFigure 3 in Supplement 1 indicates that efficiency curves for pCREST were similar across all units. The sensitivity and specificity measures for pCREST and pCART at various score thresholds are shown in eTable 4 in Supplement 1. The NNA at different sensitivities for pCREST, pCART, and Bedside PEWS are shown in **Figure 2**. Across most threshold-specific sensitivity measures, pCREST resulted in lower NNA compared with pCART and Bedside PEWS across the hospital. For example, at an example sensitivity of 80%,

pCREST would have resulted in 3 and 5 fewer patient-level alerts than pCART (6 vs 9) and Bedside PEWS (6 vs 11), respectively.

eTable 5 in Supplement 1 compares the performance of pCREST and 3 XGB models trained using hospital unit-specific observations within the derivation data on corresponding unit-specific observations in the external test data. As shown, pCREST performance was equivalent to the unit-specific model within the ED and ICU but performed better than the unit-specific models in the ward. The results of our subgroup analysis are provided and explained in eTables 6 to 8 in Supplement 1.

**Table 1. Site-Specific Comparisons of Characteristics and Outcomes Observed Between Patient Admissions Who Did and Did Not Experience at Least One Critical Event During Their Hospital Stay**

Characteristic or outcome	Patients, No. (%)					
	UC		Loyola		UW-Madison	
	With critical events (n = 2541)	Without critical events (n = 52 518)	With critical events (n = 779)	Without critical events (n = 38 134)	With critical events (n = 2543)	Without critical events (n = 39 106)
Age, mean (SD), y	6 (6) <sup>a</sup>	7 (6)	7 (6) <sup>a</sup>	6 (6)	5 (6) <sup>a</sup>	7 (6)
Sex						
Male	1454 (57.2)	28 863 (55.0)	463 (59.4)	21 872 (57.4)	1336 (52.5)	21 257 (54.4)
Female	1087 (42.8)	23 655 (45.0)	316 (40.6)	16 262 (42.6)	1207 (47.5)	17 849 (45.6)
Race						
Black	1565 (61.6) <sup>a</sup>	31 240 (59.5)	222 (28.5)	9245 (24.2)	289 (11.4)	3777 (9.7)
White	569 (22.4)	14 350 (27.3)	371 (47.6)	17 916 (47.0)	2057 (80.9)	32 590 (83.3)
Other <sup>b</sup>	407 (16.0)	6928 (13.2)	186 (23.9)	10 973 (28.8)	197 (7.7)	2739 (7.0)
Hispanic	326 (12.8)	6150 (11.7)	294 (37.7)	15136 (39.7)	250 (9.8)	3564 (9.1)
Prior comorbidity count						
0	1475 (58.0) <sup>a</sup>	39 693 (75.6)	466 (59.8) <sup>a</sup>	31 049 (81.4)	707 (27.8) <sup>a</sup>	20 143 (51.5)
1	230 (9.1)	5148 (9.8)	66 (8.5)	3514 (9.2)	465 (18.3)	5802 (14.8)
>1	836 (32.9)	7677 (14.6)	247 (31.7)	3571 (9.4)	1371 (53.9)	13 161 (33.7)
Length of stay, median (IQR), d	8 (3-19) <sup>a</sup>	2 (1-4)	7 (1-19) <sup>a</sup>	2 (1-3)	5 (1-14) <sup>a</sup>	2 (1-4)
Initial hospital location						
ICU	926 (36.4) <sup>a</sup>	4722 (9)	259 (33.2) <sup>a</sup>	4797 (12.6)	689 (27.1) <sup>a</sup>	2850 (7.3)
ED	1238 (48.7)	26 706 (50.9)	355 (45.6)	16 862 (44.2)	1145 (45.0)	15 503 (39.6)
Ward	223 (8.8)	13 780 (26.2)	138 (17.7)	14 544 (38.1)	436 (17.1)	13 538 (34.6)
Other <sup>c</sup>	154 (6.1)	7310 (13.9)	27 (3.5)	1931 (5.1)	273 (10.7)	7215 (18.4)
Mortality	76 (3.0)	NA	22 (2.8)	NA	41 (1.6)	NA
Mechanical ventilation	2311 (90.9)	NA	585 (75.1)	NA	1969 (77.4)	NA
Vasopressors	574 (22.6)	NA	251 (32.2)	NA	891 (35)	NA

Abbreviations: ED, emergency department; ICU, intensive care unit; Loyola, Loyola University Ronald McDonald Children's Hospital; NA, not applicable; UC, University of Chicago Comer Children's Hospital; UW-Madison, American Family Children's Hospital at University of Wisconsin-Madison.

<sup>b</sup> Includes American Indian or Alaska Native, Asian or Mideast Indian, declined/unknown, multiracial, Pacific Islander or Hawaiian Native, and other.

<sup>c</sup> Includes the operating room, interventional and diagnostic areas (eg, sedation, radiation therapy, radiology, etc.), and other indeterminate hospital locations.

<sup>a</sup> P < .001 compared with patients who did not experience critical events at that site.

**Table 2. Performance of Hospitalwide Models at Predicting Critical Events Within the Next 12 Hours Using the Temporal Validation and the External Test Datasets**

Model	AUC (95% CI)	
	Temporal validation	External test
Bedside PEWS	0.67 (0.67-0.67)	0.70 (0.70-0.70)
pCART	0.79 (0.78-0.79) <sup>a</sup>	0.82 (0.81-0.82) <sup>a</sup>
Logistic regression	0.83 (0.83-0.83) <sup>a,b</sup>	0.84 (0.84-0.84) <sup>a,b</sup>
XGB (pCREST)	0.89 (0.89-0.89) <sup>a,b</sup>	0.86 (0.86-0.87) <sup>a,b</sup>
RNN	0.87 (0.87-0.87) <sup>a,b</sup>	0.84 (0.84-0.85) <sup>a,b</sup>
XGB-RNN	0.88 (0.88-0.88) <sup>a,b</sup>	0.85 (0.85-0.86) <sup>a,b</sup>

Abbreviations: AUC, area under the receiver operating characteristic curve; Bedside PEWS, Bedside Pediatric Early Warning System; pCART, pediatric Calculated Assessment of Risk and Triage; pCREST, pediatric Critical Event Risk Evaluation and Scoring Tool; RNN, recurrent neural network; XGB, extreme gradient-boosted.

<sup>a</sup> P < .001 compared with the Bedside PEWS model.

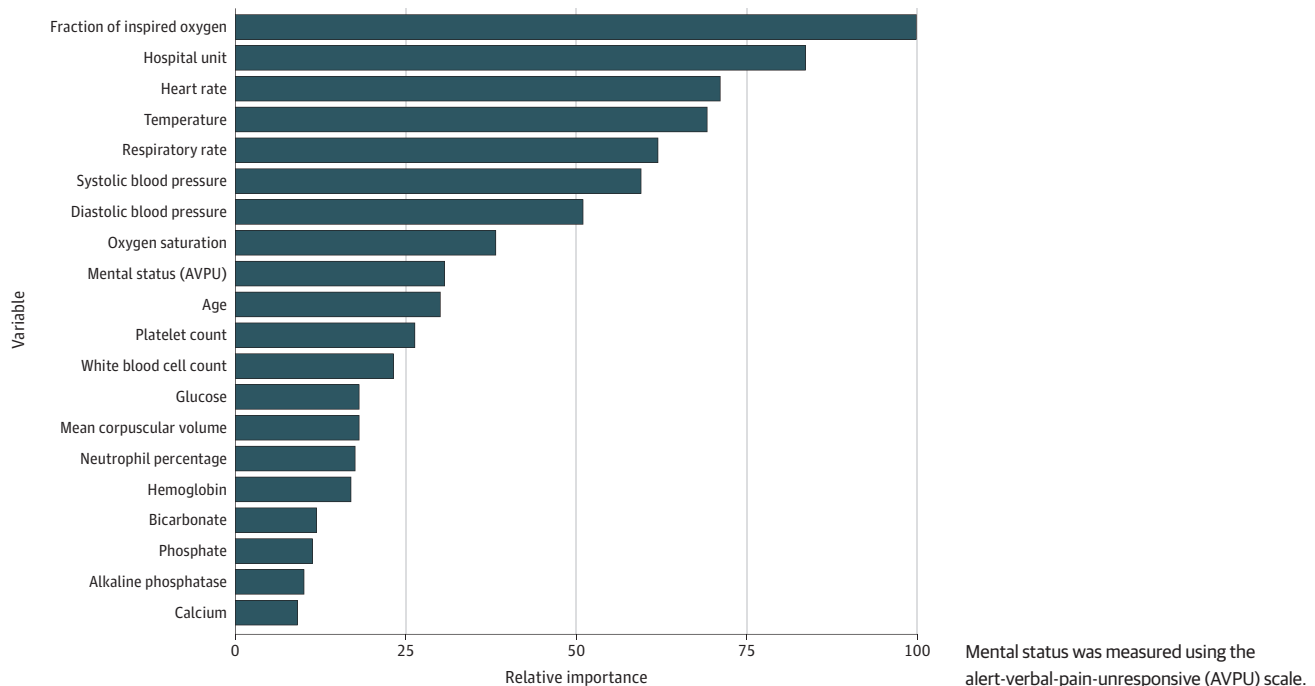
<sup>b</sup> P < .001 compared with the pCART model.

Discussion

In this cohort study, we developed a new machine learning model, pCREST, to predict the likelihood that a child will experience a critical event within 12 hours of any vital sign or laboratory result observation during hospitalization. Derived using multicenter data, pCREST outperformed other models in terms of discrimination and other clinically relevant metrics during temporal and external validation. Notably, pCREST matched or exceeded the performance of machine learning models trained to a specific unit. The availability of a single hospitalwide risk stratification model allows continuous risk monitoring throughout a patient’s stay, enabling early recognition and timely rescue of children at risk for critical events and streamlined care delivery.

Several studies have proposed risk prediction models that can be integrated into CDS tools within EHRs to identify children in the early phases of deterioration and enable timely interventions. For example, machine learning models in the ED primarily focus on aiding with decisions related to triage.<sup>41-43</sup> Risk stratification models in the pediatric ICU focus on predicting mortality or cardiac arrests,<sup>44-47</sup> and risk scores incorporated into CDS tools have been associated with decreased incidences of cardiopulmonary resuscitation events.<sup>21</sup> Similarly, ward-based models primarily predict ICU transfer by identifying indications of deterioration,<sup>24,26</sup> and subsequent implementation for real-time risk stratification in the ward was associated with positive patient outcomes.<sup>48</sup> Despite these advances, model performances have been noted to decline when validated outside their intended hospital unit.<sup>29</sup> For example, we observed in this study that pCART did not perform well in predicting critical events in patients outside the ward. Given the lack of evidence of generalizability, the current paradigm for early recognition of deterioration is structured to apply to specific outcomes and hospital units, placing patients at risk for fragmented care, which has been associated with adverse outcomes.<sup>49,50</sup> In addition, the disjointed risk assessment hampers decision-making within the health system, whose leaders do not have adequate information to guide the allocation of resources.<sup>16</sup> Transitioning to a cohesive, hospitalwide model, such as pCREST, could enhance care continuity, potentially improving outcomes for children and increasing efficiency in hospital workflows.

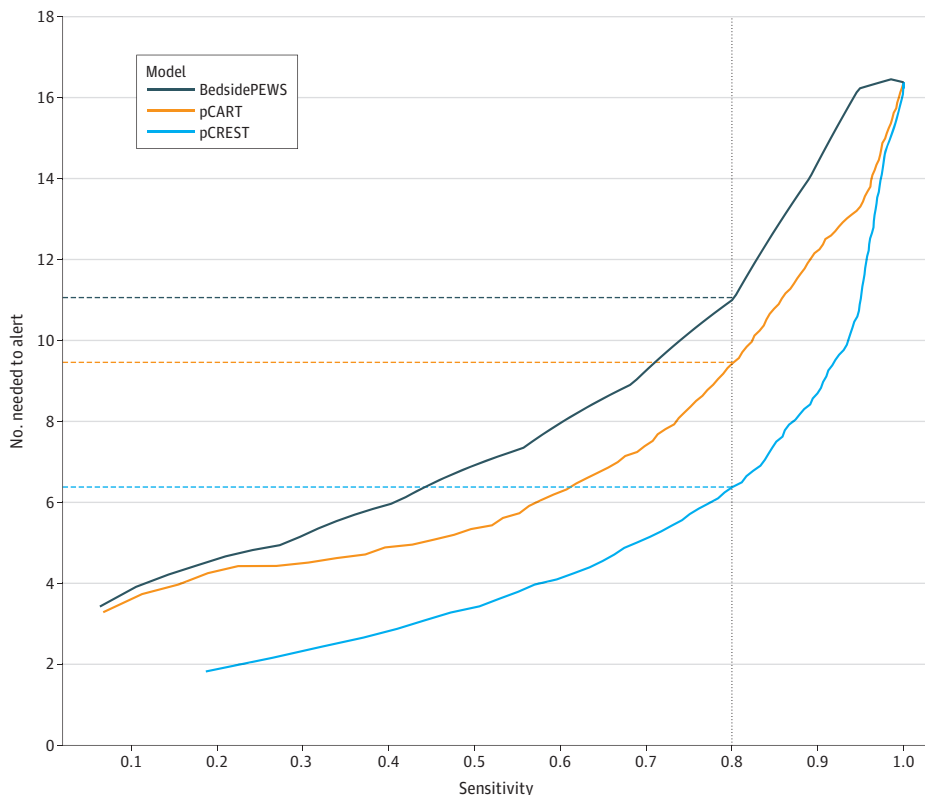
Figure 1. The Top 20 Most Important Variables, as Assessed Using Information Gain, From the pCREST Model for Predicting Critical Events



The pCREST model outperformed clinical standards and machine learning models in predicting hospitalwide critical events several hours in advance. Extension to more advanced deep learning architectures did not improve model discrimination in both the temporal validation and external test cohorts. These findings suggest that incorporating longitudinal trends in vital signs and laboratory data does not provide additional predictive value beyond using the most recent observations for identifying hospitalized children at risk for critical events. We also found that the performance of pCREST was similar to ED- and ICU-specific machine learning models and superior to ward-specific models. This result highlights the ability of pCREST to generalize to patients across units and departments that differ in event rates, clinical care, and workflows.

The numerical output of pCREST indicates the likelihood of the patient experiencing a critical event within the next 12 hours, and the output is generated with every new vital sign or laboratory result recorded in the EHR. The variables used by pCREST are age, standard physiological measurements that are routinely collected, location of the patient within the hospital, and comorbidities from prior encounters, all of which are easily extractable from the EHR and do not impede prospective implementation. The operationalization of pCREST also does not depend on imputation, as it uses the most recent observations for prediction. To enhance interpretation, pCREST outputs can be scaled to a 0 to 100 score, similar to our illustration in eFigure 2 in Supplement 1 or our implementation of pCART,<sup>48,51</sup> and incorporated into clinical practice. The scores can serve multiple purposes for the clinicians. For example, they can be used for an initial admission risk assessment or continuous real-time risk monitoring.<sup>52</sup> Each unit can establish thresholds that, when crossed, trigger alerts indicating early signs of deterioration, ensuring timely identification and rescue of patients at risk of adverse outcomes.<sup>53</sup> Additionally, scores could inform triage decisions,

Figure 2. The Number Needed to Alert for Various Sensitivity Values for the Pediatric Critical Event Risk Evaluation and Scoring Tool (pCREST), Pediatric Calculated Assessment of Risk and Triage (pCART), and Bedside Pediatric Early Warning System (Bedside PEWS) in the External Test Cohort at University of Wisconsin-Madison



such as whether to admit an ED patient to the ward or ICU or if an ICU patient can transition to a lower acuity setting. Finally, the health system can use unit-level pCREST scores to optimize the allocation of hospital resources.

The availability of a single pCREST risk score can significantly enhance communication among health care practitioners across units by establishing a shared mental model of patient risk, creating better clinician alignment regarding a patient's health.<sup>54</sup> Several studies have stressed the importance of effective communication and standardized protocols during patient handover.<sup>50,55-59</sup> pCREST scores could play a pivotal role in these programs by providing objective measures of the severity of illness that can guide care coordination among teams.

Our variable importance analysis revealed that  $\text{FiO}_2$ , heart rate, temperature, and respiratory rate were the most important in predicting pediatric critical illness.<sup>60-63</sup> The hospital unit location of the patient was considered an additional influential variable. This further supports our hypothesis that our model uses the hospital unit as a proxy for adjustment according to patient acuity or complexity, allowing us to generalize across units. However, the reasons for ICU admission are likely to be multifactorial and may vary across hospital sites and time, indicating the importance of undertaking external validation, with and without temporal splits, before pCREST can be implemented in a hospital.

### Limitations

This study has several limitations. Our study is retrospective in nature and may have limitations in the data elements collected. For example, our model uses the number of prior comorbidities as a feature, which relies on the completeness and accuracy of recorded diagnosis codes for past visits in the EHR. Before pCREST implementation, future work should address ethical concerns with machine learning models, such as ensuring fairness of operation across relevant patient groups, data security, and implementing effective explainability algorithms at the bedside. Similar to our prior work,<sup>48</sup> assessment of patient outcomes and evaluation of clinical utility after model implementation are also needed before wide adoption. Additionally, choosing critical events could fall short of a consensus outcome of deterioration throughout a pediatric hospital.<sup>16</sup> Validation for other measures of hospital deterioration remains a focus of future work.

### Conclusions

In this retrospective cohort study, We developed and externally validated a new EHR-based machine learning model called pCREST for predicting critical events among hospitalized children across hospital EDs, wards, and ICUs. Our model could be used to monitor a child's health seamlessly throughout their hospital journey, facilitating early recognition of deterioration and timely intervention.

#### ARTICLE INFORMATION

**Accepted for Publication:** March 27, 2025.

**Published:** May 30, 2025. doi:10.1001/jamanetworkopen.2025.13149

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](#). © 2025 Strutz S et al. *JAMA Network Open*.

**Corresponding Author:** Anoop Mayampurath, PhD, Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, 265H WARF Office Building, 610 Walnut St, Madison, WI 53726 ([mayampurath@wisc.edu](mailto:mayampurath@wisc.edu)).

**Author Affiliations:** Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison (Strutz, Liang, Afshar, Churpek, Mayampurath); Department of Medicine, University of Chicago, Chicago, Illinois (Carey, Edelson); Department of Medicine, University of Wisconsin-Madison, Madison (Bashiri, Afshar, Churpek, Mayampurath); Department of Pediatrics, University of Chicago, Chicago, Illinois (Jani); Department of Medicine,

Loyola University Medical Center, Chicago, Illinois (Gilbert); Department of Pediatrics, Loyola University Medical Center, Chicago, Illinois (Fitzgerald); Department of Emergency Medicine, University of Wisconsin-Madison, Madison (Kuehnel); Division of Critical Care Medicine, Department of Pediatrics, University of Cincinnati, Cincinnati, Ohio (Dewan); Department of Pediatrics, Ann & Robert H Lurie Children's Hospital of Chicago, Chicago, Illinois (Sanchez-Pinto).

**Author Contributions:** Dr Mayampurath had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Ms Strutz and Mr Liang contributed equally to the study.

**Concept and design:** Liang, Jani, Kuehnel, Churpek, Mayampurath.

**Acquisition, analysis, or interpretation of data:** Strutz, Liang, Carey, Bashiri, Gilbert, Fitzgerald, Kuehnel, Dewan, Sanchez-Pinto, Edelson, Afshar, Churpek, Mayampurath.

**Drafting of the manuscript:** Strutz, Liang, Jani, Fitzgerald, Kuehnel, Dewan, Mayampurath.

**Critical review of the manuscript for important intellectual content:** Liang, Carey, Bashiri, Jani, Gilbert, Kuehnel, Dewan, Sanchez-Pinto, Edelson, Afshar, Churpek, Mayampurath.

**Statistical analysis:** Strutz, Liang, Carey, Bashiri, Mayampurath.

**Obtained funding:** Kuehnel, Mayampurath.

**Administrative, technical, or material support:** Jani, Kuehnel, Dewan, Edelson, Mayampurath.

**Supervision:** Jani, Sanchez-Pinto, Afshar, Churpek, Mayampurath.

**Conflict of Interest Disclosures:** Dr Gilbert reported receiving grants from the National Institutes of Health (NIH) during the conduct of the study. Dr Sanchez-Pinto reported receiving grants from the NIH during the conduct of the study. Dr Edelson reported receiving personal fees from AgileMD in the form of equity and employment outside the submitted work; in addition, Dr Edelson had a patent for US-11410777-B2 issued to the University of Chicago. Dr Churpek reported receiving grants from National Institute of General Medical Sciences, National Institute of Diabetes and Digestive and Kidney Diseases, and Department of Defense outside the submitted work; in addition, Dr Churpek had a patent for a machine learning model (eCART) issued to the University of Chicago received royalties from the University of Chicago related to the eCART patent. No other disclosures were reported.

**Funding/Support:** Dr Mayampurath, Ms Strutz, Mr Carey, Dr Jani, Dr Kuehnel, and Dr Sanchez-Pinto are supported by an RO1 from the National Heart, Lung, and Blood Institute (NHLBI) (grant No. RO1HL173037). Ms Strutz is also supported by a training grant from the National Library of Medicine (grant No. 5 15LM007359). Dr Churpek is supported by a grant from the NHLBI (grant No. RO1HL157262). Dr Dewan is supported by the Agency for Healthcare Research and Quality (grant No. K08HSO26975, 1R18HSO29630, and 1R18HSO29626).

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Data Sharing Statement:** See [Supplement 2](#).

**Additional Contributions:** We thank the Center for Research Informatics (University of Chicago), the Office of Informatics and Systems Development (Loyola University), and the Clinical and Health Informatics Institute (University of Wisconsin-Madison) for assistance with data extraction and Madeline Oguss, MS (University of Wisconsin-Madison), for administrative assistance; she was not compensated outside of her usual salary.

## REFERENCES

1. Bonafide CP, Roberts KE, Priestley MA, et al. Development of a pragmatic measure for evaluating and optimizing rapid response systems. *Pediatrics*. 2012;129(4):e874-e881. doi:10.1542/peds.2011-2784
2. Hussain FS, Sosa T, Ambroggio L, Gallagher R, Brady PW. Emergency transfers: an important predictor of adverse outcomes in hospitalized children. *J Hosp Med*. 2019;14(8):482-485. doi:10.12788/jhm.3219
3. Killien EY, Keller MR, Watson RS, Hartman ME. Epidemiology of intensive care admissions for children in the US From 2001 to 2019. *JAMA Pediatr*. 2023;177(5):506-515. doi:10.1001/jamapediatrics.2023.0184
4. Berg RA, Nadkarni VM, Clark AE, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Collaborative Pediatric Critical Care Research Network. Incidence and Outcomes of Cardiopulmonary Resuscitation in PICUs. *Crit Care Med*. 2016;44(4):798-808. doi:10.1097/CCM.0000000000001484
5. Blinder J, Nadkarni V, Naim M, Rossano JW, Berg RA. Epidemiology of Pediatric Cardiac Arrest. In: da Cruz EM, Ivy D, Hraska V, Jagers J, eds. *Pediatric and Congenital Cardiology, Cardiac Surgery and Intensive Care*. Springer; 2020:1-18. doi:10.1007/978-1-4471-4999-6\_58-2

6. Magill SS, Sapiano MRP, Gokhale R, et al. Epidemiology of sepsis in US children and young adults. *Open Forum Infect Dis*. 2023;10(5):ofad218. doi:10.1093/ofid/ofad218
7. Heneghan JA, Reeder RW, Dean JM, et al. Characteristics and outcomes of critical illness in children with feeding and respiratory technology dependence. *Pediatr Crit Care Med*. 2019;20(5):417-425. doi:10.1097/PCC.0000000000001868
8. Stremler R, Haddad S, Pullenayegum E, Parshuram C. Psychological outcomes in parents of critically ill hospitalized children. *J Pediatr Nurs*. 2017;34:36-43. doi:10.1016/j.pedn.2017.01.012
9. Hopkins RO. Life after pediatric critical illness: risk factors for reduced health-related quality of life and functional decline. *Am J Respir Crit Care Med*. 2019;200(7):804-805. doi:10.1164/rccm.201905-0977ED
10. Choong K, Fraser D, Al-Harbi S, et al. Functional recovery in critically ill children, the "WeeCover" multicenter study. *Pediatr Crit Care Med*. 2018;19(2):145-154. doi:10.1097/PCC.0000000000001421
11. Pinto NP, Rhinesmith EW, Kim TY, Ladner PH, Pollack MM. Long-term function after pediatric critical illness: results from the survivor outcomes study. *Pediatr Crit Care Med*. 2017;18(3):e122-e130. doi:10.1097/PCC.0000000000001070
12. Watson RS, Beers SR, Asaro LA, et al; RESTORE-Cognition Investigators. Association of acute respiratory failure in early childhood with long-term neurocognitive outcomes. *JAMA*. 2022;327(9):836-845. doi:10.1001/jama.2022.1480
13. Sosa T, Galligan MM, Brady PW. Clinical progress note: Situation awareness for clinical deterioration in hospitalized children. *J Hosp Med*. 2022;17(3):199-202. doi:10.1002/jhm.2774
14. Lambert V, Matthews A, MacDonell R, Fitzsimons J. Paediatric early warning systems for detecting and responding to clinical deterioration in children: a systematic review. *BMJ Open*. 2017;7(3):e014497. doi:10.1136/bmjopen-2016-014497
15. Bonafide CP, Localio AR, Roberts KE, Nadkarni VM, Weirich CM, Keren R. Impact of rapid response system implementation on critical deterioration events in children. *JAMA Pediatr*. 2014;168(1):25-33. doi:10.1001/jamapediatrics.2013.3266
16. Galligan MM, Sosa T, Dewan M. The need for a standard outcome for clinical deterioration in children's hospitals. *Pediatrics*. 2023;152(4):e2023061625. doi:10.1542/peds.2023-061625
17. Balamuth F, Alpern ER, Abbadessa MK, et al. Improving recognition of pediatric severe sepsis in the emergency department: contributions of a vital sign-based electronic alert and bedside clinician identification. *Ann Emerg Med*. 2017;70(6):759-768.e2. doi:10.1016/j.annemergmed.2017.03.019
18. Straney L, Clements A, Parslow RC, et al; ANZICS Paediatric Study Group and the Paediatric Intensive Care Audit Network. Paediatric index of mortality 3: an updated model for predicting mortality in pediatric intensive care\*. *Pediatr Crit Care Med*. 2013;14(7):673-681. doi:10.1097/PCC.0b013e31829760cf
19. 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;41(7):1761-1773. doi:10.1097/CCM.0b013e31828a2bbd
20. Matics TJ, Sanchez-Pinto LN. Adaptation and validation of a pediatric sequential organ failure assessment score and evaluation of the sepsis-3 definitions in critically ill children. *JAMA Pediatr*. 2017;171(10):e172352. doi:10.1001/jamapediatrics.2017.2352
21. Dewan M, Soberano B, Sosa T, et al. Assessment of a situation awareness quality improvement intervention to reduce cardiac arrests in the PICU. *Pediatr Crit Care Med*. 2022;23(1):4-12. doi:10.1097/PCC.0000000000002816
22. Parshuram CS, Duncan HP, Joffe AR, et al. Multicentre validation of the bedside paediatric early warning system score: a severity of illness score to detect evolving critical illness in hospitalised children. *Crit Care*. 2011;15(4):R184. doi:10.1186/cc10337
23. Parshuram CS, Dryden-Palmer K, Farrell C, et al; Canadian Critical Care Trials Group and the EPOCH Investigators. Effect of a pediatric early warning system on all-cause mortality in hospitalized pediatric patients: the EPOCH randomized clinical trial. *JAMA*. 2018;319(10):1002-1012. doi:10.1001/jama.2018.0948
24. Mayampurath A, Sanchez-Pinto LN, Hegermiller E, et al. Development and external validation of a machine learning model for prediction of potential transfer to the PICU. *Pediatr Crit Care Med*. 2022;23(7):514-523. doi:10.1097/PCC.0000000000002965
25. Romaine ST, Potter J, Khanijau A, et al. Accuracy of a modified qSOFA score for predicting critical care admission in febrile children. *Pediatrics*. 2020;146(4):e20200782. doi:10.1542/peds.2020-0782

26. Rust LOH, Gorham TJ, Bambach S, et al. The Deterioration Risk Index: developing and piloting a machine learning algorithm to reduce pediatric inpatient deterioration. *Pediatr Crit Care Med*. 2023;24(4):322-333. doi:10.1097/PCC.0000000000003186
27. Kowalski RL, Lee L, Spaeder MC, Moorman JR, Keim-Malpass J. Accuracy and monitoring of Pediatric Early Warning Score (PEWS) scores prior to emergent pediatric intensive care unit (ICU) transfer: retrospective analysis. *JMIR Pediatr Parent*. 2021;4(1):e25991. doi:10.2196/25991
28. Nielsen KR, Migita R, Batra M, Gennaro JLD, Roberts JS, Weiss NS. Identifying high-risk children in the emergency department. *J Intensive Care Med*. 2016;31(10):660-666. doi:10.1177/0885066615571893
29. Seiger N, Maconochie I, Oostenbrink R, Moll HA. Validity of different pediatric early warning scores in the emergency department. *Pediatrics*. 2013;132(4):e841-e850. doi:10.1542/peds.2012-3594
30. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378
31. Liang H, Carey KA, Jani P, et al. Association between mortality and critical events within 48 hours of transfer to the pediatric intensive care unit. *Front Pediatr*. 2023;11:1284672. doi:10.3389/fped.2023.1284672
32. Feudtner C, Feinstein JA, Zhong W, Hall M, Dai D. Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatr*. 2014;14:199. doi:10.1186/1471-2431-14-199
33. Feinstein JA, Russell S, DeWitt PE, Feudtner C, Dai D, Bennett TD. R package for pediatric complex chronic condition classification. *JAMA Pediatr*. 2018;172(6):596-598. doi:10.1001/jamapediatrics.2018.0256
34. Singer JD, Willett JB. It's about time: using discrete-time survival analysis to study duration and the timing of events. *J Educ Stat*. 1993;18(2):155-195.
35. Suresh K, Severn C, Ghosh D. Survival prediction models: an introduction to discrete-time modeling. *BMC Med Res Methodol*. 2022;22(1):207. doi:10.1186/s12874-022-01679-6
36. Sloma M, Syed FJ, Nemati M, Xu KS. Empirical comparison of continuous and discrete-time representations for survival prediction. *Proc Mach Learn Res*. 2021;146:118-131.
37. Mayampurath A, Jani P, Dai Y, Gibbons R, Edelson D, Churpek MM. A vital sign-based model to predict clinical deterioration in hospitalized children. *Pediatr Crit Care Med*. 2020;21(9):820-826. doi:10.1097/PCC.0000000000002414
38. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845. doi:10.2307/2531595
39. Sundararajan M, Najmi A. The many Shapley values for model explanation. Paper presented at: 37th International Conference on Machine Learning. July 12-18, 2020; Virtual. Accessed April 22, 2025. <https://proceedings.mlr.press/v119/sundararajan20b.html>
40. Biecek P. DALEX: explainers for complex predictive models in R. *J Mach Learn Res*. 2018;19(84):1-5.
41. Zachariasse JM, Nieboer D, Maconochie IK, et al. Development and validation of a paediatric early warning score for use in the emergency department: a multicentre study. *Lancet Child Adolesc Health*. 2020;4(8):583-591. doi:10.1016/S2352-4642(20)30139-5
42. Goto T, Camargo CA Jr, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw Open*. 2019;2(1):e186937. doi:10.1001/jamanetworkopen.2018.6937
43. Hwang S, Lee B. Machine learning-based prediction of critical illness in children visiting the emergency department. *PLoS One*. 2022;17(2):e0264184. doi:10.1371/journal.pone.0264184
44. Trujillo Rivera EA, 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(5):344-352. doi:10.1097/PCC.0000000000002910
45. Lee B, Kim K, Hwang H, et al. Development of a machine learning model for predicting pediatric mortality in the early stages of intensive care unit admission. *Sci Rep*. 2021;11(1):1263. doi:10.1038/s41598-020-80474-z
46. Kim SY, Kim S, Cho J, et al. A deep learning model for real-time mortality prediction in critically ill children. *Crit Care*. 2019;23(1):279. doi:10.1186/s13054-019-2561-z
47. Dewan M, Muthu N, Shelov E, et al. Performance of a clinical decision support tool to identify PICU patients at high risk for clinical deterioration. *Pediatr Crit Care Med*. 2020;21(2):129-135. doi:10.1097/PCC.0000000000002106

48. Mayampurath A, Carey K, Palama B, et al. Machine learning-based pediatric early warning score: patient outcomes in a pre- versus post-implementation study, 2019-2023. *Pediatr Crit Care Med*. 2025;26(2):e146-e154. doi:10.1097/PCC.0000000000003656
49. Vincent JL. The continuum of critical care. *Crit Care*. 2019;23(suppl 1):122. doi:10.1186/s13054-019-2393-x
50. Patton LJ, Tidwell JD, Falder-Saeed KL, Young VB, Lewis BD, Binder JF. Ensuring safe transfer of pediatric patients: a quality improvement project to standardize handoff communication. *J Pediatr Nurs*. 2017;34:44-52. doi:10.1016/j.pedn.2017.01.004
51. McCaffery K, Carey KA, Campbell V, et al. Predicting transfers to intensive care in children using CEWT and other early warning systems. *Resusc Plus*. 2023;17:100540. doi:10.1016/j.resplu.2023.100540
52. Zackoff MW, Iyer S, Dewan M. An overarching approach for acute care delivery: extension of the acute care model to the entire inpatient admission. *Transl Pediatr*. 2018;7(4):246-252. doi:10.21037/tp.2018.09.14
53. Dewan M, Sanchez-Pinto LN. Crystal balls and magic eight balls: the art of developing and implementing automated algorithms in acute care pediatrics. *Pediatr Crit Care Med*. 2019;20(12):1197-1199. doi:10.1097/PCC.0000000000002147
54. Brady PW, Wheeler DS, Muething SE, Kotagal UR. Situation awareness: a new model for predicting and preventing patient deterioration. *Hosp Pediatr*. 2014;4(3):143-146. doi:10.1542/hpeds.2013-0119
55. Ong MS, Coiera E. A systematic review of failures in handoff communication during intrahospital transfers. *Jt Comm J Qual Patient Saf*. 2011;37(6):274-284. doi:10.1016/S1553-7250(11)37035-3
56. Starmer AJ, Spector ND, Srivastava R, et al; I-PASS Study Group. Changes in medical errors after implementation of a handoff program. *N Engl J Med*. 2014;371(19):1803-1812. doi:10.1056/NEJMsa1405556
57. Nasarwanji MF, Badir A, Gurses AP. Standardizing handoff communication: content analysis of 27 handoff mnemonics. *J Nurs Care Qual*. 2016;31(3):238-244. doi:10.1097/NCQ.0000000000000174
58. Cornell P, Gervis MT, Yates L, Vardaman JM. Impact of SBAR on nurse shift reports and staff rounding. *Medsurg Nurs*. 2014;23(5):334-342.
59. Shahid S, Thomas S. Situation, Background, Assessment, Recommendation (SBAR) communication tool for handoff in health care—a narrative review. *States Health*. 2018;4(1):7. doi:10.1186/s40886-018-0073-1
60. Kim GE, Choi SH, Park M, et al. Spo<sub>2</sub>/FiO<sub>2</sub> as a predictor of high flow nasal cannula outcomes in children with acute hypoxemic respiratory failure. *Sci Rep*. 2021;11(1):13439. doi:10.1038/s41598-021-92893-7
61. Darnell R, Brown A, Laing E, et al; Protocolised Evaluation of Permissive Blood Pressure Targets Versus Usual Care (PRESSURE) Trial Investigators on behalf of the U.K. Paediatric Critical Care Society Study Group (PCCS-SG). Protocol for a randomized controlled trial to evaluate a Permissive Blood Pressure Target Versus Usual Care in critically ill children with hypotension (PRESSURE). *Pediatr Crit Care Med*. 2024;25(7):629-637. doi:10.1097/PCC.0000000000003516
62. Lockwood J, Reese J, Wathen B, et al. The Association between fever and subsequent deterioration among hospitalized children with elevated PEWS. *Hosp Pediatr*. 2019;9(3):170-178. doi:10.1542/hpeds.2018-0187
63. Bonafide CP, Brady PW, Keren R, Conway PH, Marsolo K, Daymont C. Development of heart and respiratory rate percentile curves for hospitalized children. *Pediatrics*. 2013;131(4):e1150-e1157. doi:10.1542/peds.2012-2443

## SUPPLEMENT 1.

### eMethods.

**eTable 1.** Distribution and Admission-Level Missingness Rates of All Evaluated Features for Each Patient Cohort

**eTable 2.** A List of All Evaluated Hyperparameters, Search Values Considered, and the Optimal Hyperparameters Selected to Train Each Respective Model

**eTable 3.** Internal and External Area Under the Precision-Recall Curve Metrics of All Models Evaluated in this Study

**eTable 4.** Sensitivity and Specificity of Different Cutoffs for pCREST and pCART for Patients With a Critical Event Compared With Those Not Experiencing Any Event in the External Validation Data

**eTable 5.** Comparison of the Performance of Gradient-Boosted Machine Models Trained Using Unit-Specific Data and pCREST for Predicting Pediatric Critical Events Within the Next 12 Hours in the External Test Data

**eTable 6.** Performance of pCREST Within Individual Years From 2018-2021 Within the External Test Cohort

**eTable 7.** Evaluation of pCREST Performance in Patient Subgroups Stratified by Patient Age Within the External Test Cohort

**eTable 8.** Evaluation of pCREST Performance in Patient Subgroups Stratified by the Number of Prior Comorbidities Within the External Test Cohort

**eFigure 1.** Illustration of the Discrete-Time Survival Analysis Framework for Blocking Derivation Data into 12-Hour Intervals for Deriving the Logistic Regression and the Gradient-Boosted Machine Learning Models

**eFigure 2.** Explanation of pCREST Predictions Using Shapley Values for a Test Patient (Age 16 Years) From the UW-Madison Cohort

**eFigure 3.** Efficiency Curves for Predicting Critical Events in the Next 12 hours Using pCREST in the External Validation Cohort for Each Unit

**SUPPLEMENT 2.**

**Data Sharing Statement**