JAMA Pediatrics | Original Investigation | ARTIFICIAL INTELLIGENCE AND PEDIATRIC CARE

Derivation and Validation of Predictive Models for Early Pediatric Sepsis

Elizabeth R. Alpern, MD, MSCE; Halden F. Scott, MD, MSCS; Fran Balamuth, MD, PhD, MSCE; James M. Chamberlain, MD; Holly Depinet, MD, MPH; Lalit Bajaj, MD, MPH; Norma-Jean E. Simon, MPH, MPA; Camille P. Carter, BS; Cara Elsholz, MS; Michael Webb, MS; Diego Campos, MS; Sara J. Deakyne Davies, MPH; Lawrence J. Cook, PhD; Lyle Ungar, PhD; Robert Grundmeier, MD; for the PECARN PED Screen Study Group

IMPORTANCE Sepsis is a leading cause of death in children. Early recognition and treatment improve outcomes, but predictive models have not to date improved early diagnosis.

OBJECTIVE To develop machine learning models to estimate the probability of developing sepsis in the subsequent 48 hours.

DESIGN, SETTING, AND PARTICIPANTS This was a multisite registry for model derivation and validation using electronic health record (EHR) data from January 2016 through February 2020 and temporal validation from January 2021 through December 2022. The performance of machine learning algorithms was compared to predict development of sepsis and septic shock via logistic regression, specifically ridge regression and gradient tree boosting. Five health systems contributing to the Pediatric Emergency Care Applied Research Network were included. Emergency department (ED) visits for children aged 2 months or older to less than 18 years of age excluding patients with ED disposition of death or transfer, trauma diagnosis, or sepsis present during predictive features window. The TRIPOD-AI reporting guideline was followed, and data analysis was conducted from September 2023 to July 2025.

EXPOSURES Patient and physiologic characteristics within the first 4 hours of ED care.

MAIN OUTCOMES AND MEASURES Sepsis, defined as suspected infection with a Phoenix Sepsis Criteria (PSC) score of 2 or more or death within 48 hours of ED arrival.

RESULTS The dataset included 1 604 422 eligible visits in the training cohort and 719 298 visits in the test cohort. Performance characteristics for the PSC sepsis prediction models were AUROC of 0.92 (95% CI, 0.92-0.93) for logistic regression and 0.94 (95% CI, 0.93-0.94) for gradient tree boosting. AUROCs for PSC shock models were 0.92 or greater. The gradient tree boosting models had positive likelihood ratios ranging from 4.67 (95% CI, 4.61-4.74) to 6.18 (95% CI, 6.08-6.28) for sepsis and from 4.16 (95% CI, 4.07-4.24) to 5.83 (95% CI, 5.67-5.99) for septic shock. Predictive features included emergency severity index, age-adjusted vital signs, and medical complexity. Assessment of model performance fairness was similar for all demographic characteristics except payor; AUROC for patients with Medicaid insurance was better than for those with commercial payers.

CONCLUSIONS AND RELEVANCE Using a large multicenter population, models were developed and validated with high AUROC to predict the future development of sepsis based on EHR data collected in the ED. The models achieved positive likelihood ratios to predict sepsis and septic shock. The results highlight the opportunity for future studies that combine EHR-based models with clinical judgment to improve prediction.

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: The PECARN PED Screen Study Group members are listed at the end of the article.

Corresponding Author: Elizabeth R. Alpern, MD, MSCE, Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, 225 E Chicago Ave, Chicago, IL 60611 (ealpern@luriechildrens.org).

JAMA Pediatr. 2025;179(12):1318-1325. doi:10.1001/jamapediatrics.2025.3892
Published online October 13, 2025.

Sepsis, in which infection causes life-threatening organ dysfunction, is a leading global cause of pediatric death.[1] In the US, where more than 75 000 children are hospitalized with sepsis yearly, the in-hospital mortality is 5% to 20%.[2,3] Outcomes are worsened with delayed diagnosis and treatment, yet predictive models have not been broadly deployed successfully to improve early diagnosis of sepsis in children.[4-7]

While some models have been developed to predict mortality in pediatric sepsis, there are few that address the earlier phase of prognostic identification and treatment, before organ dysfunction is present, when a model may be most helpful for early diagnosis and treatment of high-risk patients.[8,9] Existing early, pre-intensive care unit models have suboptimal test characteristics for clinical use.[10-13] Models based on multicenter datasets have not been reported that predict organ dysfunction among children in the emergency setting, in whom organ dysfunction is not already present. New diagnostic criteria for pediatric sepsis, the Phoenix Sepsis Criteria (PSC), were recently published with the authors explicitly stating that the criteria identify infection-associated life-threating organ dysfunction but do not predict sepsis in children.[14,15] Prediction is the important next step to reduce preventable harm from sepsis.

To address the need to improve early prognostic diagnosis of sepsis in children in the ED setting, this study used multicenter electronic health record (EHR) data to derive and validate predictive models to identify patients at risk for pediatric sepsis before organ dysfunction had already occurred. Specifically, the goal of this study was to develop machine learning models using EHR data from the first 4 hours after ED arrival to estimate the probability of developing sepsis with organ dysfunction in the subsequent 48 hours.

Key Points

Question What is the predictive power of machine learning models for pediatric sepsis in the emergency department using electronic health record data to identify patients without sepsis who will develop sepsis within 48 hours?

Findings Using more than 1.6 million ED visits, models to predict Phoenix Sepsis Criteria scores were derived and validated. The gradient tree boosting models for PSC sepsis had meaningful positive likelihood ratios.

Meaning Machine learning predictive models for sepsis in the ED can identify children who have not yet developed sepsis and may be useful in future implementation work to identify children at risk.

Methods

Dataset and Population

We established a multisite registry of emergency department (ED) and inpatient EHR data (Epic and Cerner systems) from a cohort of visits at 5 health systems in the US (5 academic quaternary-care EDs and hospitals and 3 affiliated community EDs and hospitals) in the Pediatric Emergency Care Applied Research Network (PECARN). We followed the TRIPOD-AI guidance for reporting prediction models that use machine learning methods.[16] The study was approved by the institutional review boards of all study sites and the data center. We preregistered the analysis plan.[17]

The cohort included all ED visits for children age 2 months and older to less than 18 years, with the following exclusions: (1) ED disposition of death or transfer to a facility not within the database, (2) ED International Classification of Diseases 10th Revision Clinical Modification trauma diagnosis codes (S,T,V,Y prefixes), or (3) sepsis criteria met during the timeframe for monitoring predictive variables (features window). We did not include infants younger than 2 months because febrile neonate protocols largely drive care for these children with different risk levels than older children.[18-20]

The features window started at the arrival to the ED and concluded at the earlier of either 4 hours from ED arrival or ED disposition (discharge home or admission/transfer to a study hospital). Outcomes were calculated within a 48-hour window after the end of the features window (eFigure 1 in Supplement 1) including EHR data from the index and return visits within the outcome window.

The training dataset consisted of data from January 1, 2016, through February 29, 2020, and a heldout temporal validation dataset from January 1, 2021, through December 31, 2022. We excluded visits during the early months of the COVID-19 pandemic (March-December 2020) because this period was not representative of typical care.[21] Data were analyzed from September 2023 to July 2025.

Outcomes

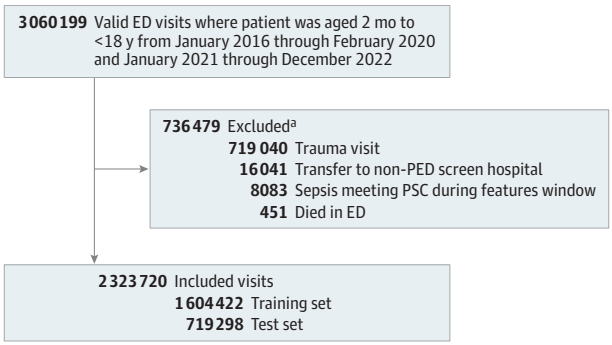
The primary outcome of the model was a composite of suspected infection and sepsis or death. Suspected infection was defined, as in past work in the ED setting, using a laboratory evaluation for infectious etiology or a chest radiograph between ED arrival and conclusion of the 48-hour outcome window, to identify patients who were evaluated for potential infection.[22] In our models, sepsis outcome was determined by suspected infection with a PSC score of 2 or greater.[14,15]

To ensure all sepsis score-qualifying organ dysfunction points occurred within the same 24-hour period, PSC scores were assessed for two 24-hour discrete time periods within the outcome window. For each visit, sepsis criteria were assessed using the worst observed vitals, laboratory values, and interventions in each period, with the higher score of the 2 periods determining the PSC score for the visit. The secondary outcome of PSC septic shock was defined as suspected infection and PSC sepsis with 1 or more PSC cardiovascular points using the same process described above.

Features

Features were categorized and models were built for each of the following: (1) patient characteristics (ie, physiologic and clinical measures excluding measures affected by clinical decisions such as laboratory test results or the number of vital signs collected) and (2) patient characteristics excluding the Emergency Severity Index (ESI), a global assessment of acuity

Figure 1. Flow Diagram



ED indicates emergency department; PED, pediatric emergency department; PSC, Phoenix Sepsis Criteria

[a]Children could be excluded for multiple reasons.

Table 1. Phoenix Sepsis Criteria (PSC) Outcomes in the Study Cohort Excluding Individuals With Sepsis in the First 4 Hours of Emergency Department Care

Dataset	Visits with outcome, No. (%) [95% CI][a]	Range across health systems, %
Training (n = 1 604 422)		
PSC sepsis	5634 (0.35) [0.34-0.36]	0.17 to 0.50
PSC shock	2446 (0.15) [0.15-0.16]	0.08 to 0.26
Test (n = 719 298)		
PSC sepsis	2639 (0.37) [0.35-0.38]	0.22 to 0.61
PSC shock	1062 (0.15) [0.14-0.16]	0.09 to 0.21

[a] The outcome definition of suspected infection and sepsis or death included 16 of 8273 patients (5634 in the training cohort and 2639 in the test cohort) with PSC sepsis (0.19%) who did not meet sepsis organ dysfunction prior to death.

and severity assigned at triage.[23] Patient features for the primary analyses included (eTable 1A and 1B in Supplement 1) (1) measures of acuity, including the Emergency Severity Index,[21] arrival mode, prior notification of arrival or referral; (2) clinical observations, including weight, age-adjusted vital signs, oxygen saturation, and pain score (all scores normalized to a 10 point scale); (3) markers for medical complexity, including ED utilization during the prior year, presence of complex chronic conditions (CCC, categorized as no CCCs vs at least 1 CCC; also categorized as an ordinal feature: no CCCs, 1 CCC, or 2 or more CCCs present in the EHR problem list),[24,25] and presence of indwelling central venous line or tracheostomy; and (4) biological variables, including age and sex assigned at birth.

Statistical Analysis

Using the training data, we tuned and compared the performance of 2 machine learning algorithms to predict sepsis and septic shock: (1) logistic regression, specifically ridge regression (L2 penalty), and (2) gradient tree boosting. Hyperparameters of our models were tuned within the training dataset by 10-fold cross-validation nested within an outer loop of 5-folds in which each hospital system was held out. The optimization target for hyperparameter tuning was maximum specificity on held out samples at a threshold of 90% sensitivity, a threshold chosen as a minimal acceptable sensitivity for screening for a potentially fatal disease. The dataset was partitioned to ensure that children with multiple ED visits were not divided across folds. For the L2 penalty (ridge) regression, we considered cost parameters of 10^{-5}, 10^{-2}, 10^{-1}, and 1. For gradient tree boosting, we used a grid search approach to select hyperparameters with a learning rate of 0.1 and the following parameters: {'n_estimators':[100, 1000], 'max_features':[1, 2, 4, 8], 'max_depth':[1, 2, 4, 8]}.

Using the hyperparameters that maximized specificity at 90% sensitivity across the majority of cross-validation folds, we refit our models using all available training data and evaluated performance in our temporal test set. We determined specificity, positive likelihood ration, and positive predictive value at both 90% sensitivity and Youden cutpoints,[26] and area under the receiver operating characteristic (AUROC). We provide plots of the receiver operating characteristic curve, the precision-recall curve, and Shapley Additive Explanations values.[27] We assessed fairness of the models to not create or exacerbate inequalities in health care provision or outcomes by assessing performance in subgroups of age, race, ethnicity, gender, language, insurance status, and site.[16] We assessed calibration by reporting Brier scores and binning predicted probabilities of sepsis into deciles to calculate the expected calibration error and plot-calibration curves. Model stability was evaluated by fitting a suite of 5 models using the training data, but withholding the data for 1 of the 5 sites in each model (ie, each model was fit using data from 4 of the 5 available sites). We then assessed the variability in performance observed in the temporal test set compared to the models fit using all available training data. All analyses were performed using Python version 3.8 with the addition of scikit-learn version 1.3.0 and gradient tree boosting (XGBoost version 1.7.6). The code for these analyses and resultant models are available online.[28]

Results

There were 3 060 199 eligible ED visits with 2 323 720 meeting inclusion criteria, with 1 604 422 visits in the training cohort and 719 298 in the test cohort (**Figure 1**). In the training cohort, 0.35% (95% CI, 0.34-0.36) of visits met PSC sepsis criteria and 0.15% (95% CI, 0.15-0.16) met PSC shock criteria in the outcome window with similar proportions in the test set (**Table 1**). PSC sepsis and shock rates varied across sites (Table 1). Demographic description of the cohort by site is in eTable 2A in Supplement 1 and by training or temporal validation datasets in eTable 2B in Supplement 1 with an overall median (IQR) age of 4.7 years (1.7-10.1) and 48.6% of visits among female individuals. The median (IQR) hospital length of stay for those visits in the overall cohort resulting in admission was 2 (1-4) days, for visits meeting PSC sepsis criteria was 6 (3-11) days, and for those visits meeting PSC shock criteria was 5 (3-10) days. The overall cohort had a mortality rate of 0.015%, with visits meeting PSC sepsis criteria having mortality of 2.164%, and visits meeting PSC shock criteria with mortality of 3.307%.

Table 2. Model Performance for Phoenix Sepsis Criteria (PSC) Outcomes in the Heldout Temporal Validation Dataset for Ridge Logistic Regression (LR) and Extreme Gradient Boosting (XGBoost) Models

	Statistic (95% CI)					
	AUROC	Sensitivity	Specificity	LR+	PPV	NNE
PSC sepsis						
90% LR	0.923 (0.918-0.928)	0.900 (0.888-0.911)	0.779 (0.778-0.780)	4.067 (4.013-4.122)	0.015 (0.014-0.015)	68 (65-71)
Youden LR		0.844 (0.830-0.858)	0.854 (0.853-0.855)	5.788 (5.688-5.889)	0.021 (0.020-0.022)	48 (46-50)
90% gradient tree boosting	0.936 (0.931-0.940)	0.900 (0.888-0.911)	0.807 (0.806-0.808)	4.674 (4.612-4.738)	0.017 (0.016-0.018)	59 (57-61)
Youden gradient tree boosting		0.858 (0.844-0.871)	0.861 (0.860-0.862)	6.179 (6.077-6.282)	0.022 (0.021-0.023)	45 (43-47)
PSC shock						
90% LR	0.923 (0.915-0.931)	0.900 (0.881-0.917)	0.778 (0.777-0.779)	4.059 (3.977-4.143)	0.006 (0.006-0.006)	168 (157-179)
Youden LR		0.848 (0.826-0.869)	0.840 (0.840-0.841)	5.314 (5.178-5.454)	0.008 (0.007-0.008)	128 (120-137)
90% gradient tree boosting	0.926 (0.918-0.934)	0.900 (0.881-0.917)	0.783 (0.782-0.784)	4.157 (4.073-4.243)	0.006 (0.006-0.007)	164 (154-174)
Youden gradient tree boosting		0.836 (0.813-0.857)	0.857 (0.856-0.857)	5.828 (5.671-5.989)	0.009 (0.008-0.009)	117 (110-125

Abbreviations: AUROC, area under the receiver operating characteristic curve; PPV, positive predictive value; LR+, positive likelihood ratio; NNE, number needed to evaluate.

Performance characteristics for each of the models predicting PSC sepsis in the temporal validation dataset were excellent, with AUROC of 0.923 (95% CI, 0.918-0.928) in the logistic regression models and 0.936 (95% CI, 0.931-0.940) in the gradient tree boosting models (**Table 2**). The specificity for the sepsis models with target sensitivity of 90% ranged from 0.779 to 0.807. These correspond to positive likelihood ratios (LR+) of 4.067 to 4.674. The Youden cutpoint models for sepsis resulted in sensitivities of at least 0.84 and specificities ranging from 0.854 to 0.861 with LR+ ranging from 5.788 to 6.179. At a sensitivity of 90%, the LR+ of the PSC sepsis gradient tree boosting model was 4.674 with the PPV of 1.7% compared to the observed overall prevalence of 0.37%. Over all models, the number needed to evaluate ranged from 45 to 68 (Table 2).

Model results for PSC shock are presented in Table 2. AUROCs of 0.923 (95% CI, 0.915-0.931) were observed for the logistic regression and 0.926 (95% CI, 0.918-0.934) for the gradient tree boosting models.

The receiver operating characteristic and precision-recall curves for the gradient tree boosting models for PSC sepsis and shock applied to the temporal validation dataset are depicted in **Figure 2**. There was minor variability in model performance by site (PSC sepsis AUROC range, 0.922 to 0.945) (Figure 2). The receiver operating characteristic and precision-recall curves for the logistic regression models applied to the temporal validation set are depicted in eFigure 2 in Supplement 1. Calibration of the gradient tree boosting and logistic regression models for the temporal validation set are shown in eFigure 3 in Supplement 1. To allow for comparison with the temporal validation model performance, model characteristics for all models in the training dataset are presented in eTable 3 in Supplement 1 with the ROC and precision-recall curves for PSC sepsis and shock in the training dataset depicted, respectively, in eFigures 4 and 5 in Supplement 1.

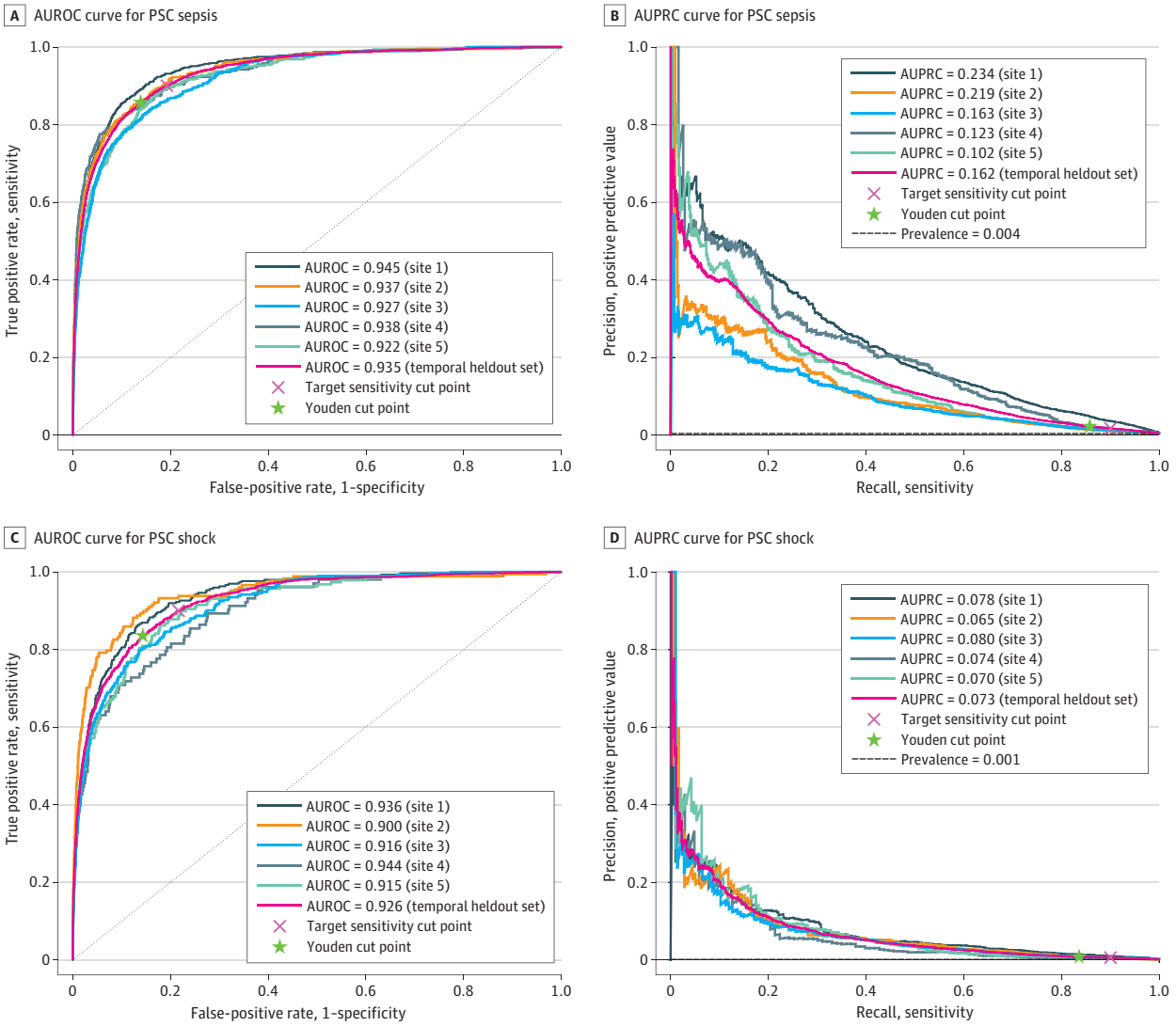
The top 20 features for the gradient tree boosting PSC sepsis models with and without Emergency Severity Index are presented in **Figure 3** with many similar predictive features, including age-adjusted vital signs, age, and markers of medical complexity. Models excluding Emergency Severity Index triage data as features had similar performance characteristics, AUROCs, and area under the precision-recall curves (eTable 4 in Supplement 1). Proportions of missingness of features in the models, training and validation sets, and by site, are presented in eTable 5A and 5B in Supplement 1, respectively.

Assessment of fairness in PSC sepsis gradient tree boosting model performance by age, gender, race, ethnicity, language, insurance, and site demonstrated AUROC for patients with Medicaid insurance of 0.940 (95% CI:,0.934-0.946) compared to commercial payer mean of 0.920 (95% CI, 0.910-0.929). The model otherwise performed similarly across patient demographic groups (eFigure 6 in Supplement 1). In our stability assessment comparing models derived on incomplete training data, performance was generally a few percentage points inferior compared to the primary analysis of models derived on the complete training set (eTable 6 in Supplement 1). The lowest performance observed in this stability analysis was for predicting PSC shock using logistic regression (AUROC of 0.906 [95% CI, 0.895 to 0.915] for the lowest performing model in our stability analysis compared to 0.923 [95% CI, 0.915-0.931] observed in the primary analysis).

Discussion

In this cohort study, using a large dataset, including a diverse population, from 5 health systems, we developed machine learning models predicting the future development of sepsis based on EHR data collected early in the ED course, prior to the development of organ system dysfunction. Our models achieved robust AUROC measurements and meaningful likelihood ratios. The strengths of these results include the successful development of models capable of early prediction of a rare outcome with excellent

Figure 2. Extreme Gradient Boosting (XGBoost) Performance for Phoenix Sepsis Criteria (PSC) Sepsis and Shock Outcomes

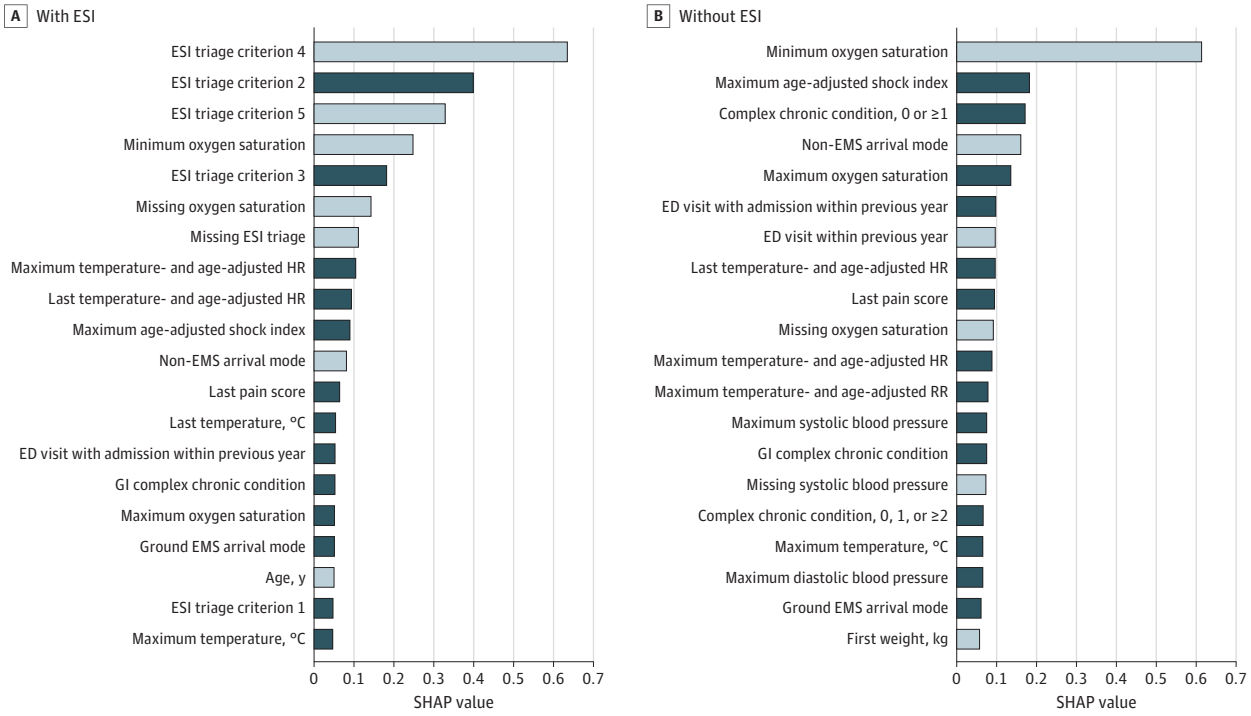


For PSC sepsis (A and B) and PSC shock (C and D), the gradient tree boosting algorithm area under the receiver operator characteristic (ROC) curves and area under the precision-recall curves (AUPRC) are shown for the temporal validation dataset.

performance characteristics. Notably, we meticulously excluded any children who already exhibited the outcome of interest at the time of presentation, and we deliberately avoided including markers of clinical judgment, which can introduce bias. To our knowledge, this study represents the first attempt to use robust machine learning models to predict the Phoenix Sepsis Criteria outcomes in a pediatric ED population that did not already have sepsis.[14,15]

Although there have been screening tools for identification of patients with sepsis or those who may have increased sepsis risk in the ED,[29-32] there are no previously published multicenter predictive models of pediatric sepsis in the ED that exclude patients who already have met sepsis criteria within their first hours of emergency care. Most work in pediatric sepsis predicts mortality in patients with identified sepsis or provides screens for risk of sepsis without excluding patients who have already developed sepsis.[25,33-35] There is a distinct and important role for both prediction of sepsis and for triggers that help identify patients who already have sepsis to allow for early treatment. Our work, however, allows for rigorous prediction, using routine EHR data, in those patients who have not yet developed sepsis. This work indicates that identification of relevant data for prediction is possible using regularly collected data from the EHR. The strengths of machine learning modeling and its implementation within the EHR is gaining acceptance in medicine, especially in instances of difficult diagnosis and prognosis.[14,36] Important predictive features in our study included Emergency Severity Index category; vital signs, including oxygen saturation; age-adjusted shock index; and markers of medical complexity. This work is a step forward to achieve the goal of up-stream identification that may allow for preemptive treatment to improve outcomes.

Figure 3. Extreme Gradient Boosting Feature Importance With and Without Emergency Severity Classification (ESI)



Top 20 features are shown (dark blue indicates a positive association with the sepsis outcome and light blue indicates an inverse association with the outcome). ED indicates emergency department; EMS, emergency medical services; GI, gastrointestinal; HR, hazard ratio; SHAP, Shapley Additive Explanations.

Prior work by Scott et al[11,12] derived and validated models in 1 health system to predict septic shock in children already identified by clinicians as having risk of sepsis, using data known at the time of ED arrival and with data at 2 hours after arrival. That work included patients identified by suspected sepsis order sets and intended to risk stratify patients who did not yet have shock but in whom clinicians were concerned for sepsis. Our work, compared to these prior studies, included an undifferentiated cohort of ED patients. With our multicenter approach, the current study had strong AUROCs for prediction of sepsis and septic shock compared to prior work.

Use of machine learning for prediction in clinical care must be balanced with understanding of fairness of the models and evaluation of biases that may be introduced in the derivation of models.[16] Fortunately, when assessing for model equity we found similar model performance, except for slightly improved performance in children with Medicaid insurance compared to those with private insurance. This may be because most of the cohort had public insurance and therefore the data used for derivation had majority representation of this community.

Limitations

Our study encountered challenges and limitations. First, despite efforts to avoid markers of clinical judgment, some outcomes related to organ dysfunction are influenced by health care processes. For example, the absence of measured oxygen saturation, an influential feature in the models, is likely a marker for lack of respiratory distress, which is not easily captured in the EHR. Our approach of treating missing data is consistent with prior work, including the Sepsis-3 and PSC derivations.[14,15,37] Second, our best-performing algorithm, gradient tree boosting, is a complex model, which may pose challenges for implementation, especially related to explainability. However, our features importance graphs help with accessibility to clinicians and recent work indicates acceptance among care providers and patients to harnessing the strength of the EHR to implement complex models.[38,39] Future work, following implementation science methodologies, will integrate the output of risk from the models into clinical workflows.[40]

The support of sepsis recognition and diagnosis has been identified as 1 of the greatest challenges to pediatric care.[41,42] Our work derives and validates models based on a multihealth system cohort with differing EHR vendors and different EHR instances within the same vendor. Our study excluded patients with identified sepsis within the feature window, to concentrate on prediction, not identification of already developed sepsis. As sepsis is a continuously evolving clinical entity, any variation of the feature window will affect the metrics. Additionally, our study included an undifferentiated cohort of patients compared to past work that used cohorts of patients with clinician concern for sepsis.[25-27] The mortality rate in our cohort was less than that in other published

work, which included global estimates or cohorts enriched by intensive care or hospitalized populations.[2,3,14] With any rare outcome, the rate of false positive predictions may pose a risk of alarm fatigue, and there will always be a need to balance sensitivity with the rate of false positive detection. The PSC sepsis gradient tree boosting model, set at 90% sensitivity, achieved a number needed to evaluate of 59, which one may argue is a reasonable starting point to identify a disease process with significant morbidity and mortality. Our models, after appropriate prospective validation, will allow for assessment and possible implementation with variable cutpoints of positive predictive value thresholds to reduce alarm fatigue. The use of the validated PSC sepsis and septic shock outcomes in our current work will allow future work to evaluate the impact of early clinical steps predicated on predictive model alerts.

Subsequent research should explore the potential integration of clinical judgment in a 2-step process, which has been previously shown to be an important component of sepsis recognition.[25] Additionally, although our work included patients treated in quaternary care pediatric EDs and affiliated community EDs from 5 distinct health systems and 2 EHR vendors, future work should broaden the scope to encompass general ED settings and prehospital care environments. Physiologic data missing not at random are concerning when using data collected in routine care to predict outcomes and could result in confounding. Reassuringly, in our analyses the missingness of data was ranked low in our measures of feature importance, but future implementation efforts should consider the potential impact of missing data if such rates are notably different than those observed in our analyses. Future evaluation of parsimonious models compared to these comprehensive models of patient and physiologic features is also worthwhile.

Conclusions

In conclusion, our study leveraged a large multicenter EHR database to identify cases of sepsis and septic shock occurring within 48 hours in children without sepsis in the initial 4 hours after ED presentation. We developed and validated models with high areas under the receiver operating characteristic curve and meaningful positive likelihood ratios. Limited positive predictive values underscore the difficulty in predicting the rare outcome of pediatric sepsis in the ED. Our research highlights the need for future studies that combine EHR-based models with clinical judgment to improve prediction.

ARTICLE INFORMATION

Accepted for Publication: August 20, 2025.

Published Online: October 13, 2025. doi:10.1001/jamapediatrics.2025.3892

Author Affiliations: Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, Illinois (Alpern); Department of Pediatrics, Children's Hospital Colorado, University of Colorado, Aurora (Scott, Bajaj); Department of Pediatrics, Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Balamuth, Grundmeier); Department of Pediatrics, Children's National Medical Center, Washington, DC (Chamberlain); Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, Ohio (Depinet); Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois (Simon); Utah Data Coordinating Center, University of Utah, Salt Lake City (Carter, Elsholz, Webb); Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Campos); Research Informatics, Children's Hospital Colorado, Aurora (Deakyne Davies); Department of Pediatrics, University of Utah, Salt Lake City (Cook); Department of Computer Science, University of Pennsylvania, Philadelphia (Ungar).

Author Contributions: Dr Grundmeier and Ms Carter had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Alpern, Scott, Balamuth, Chamberlain, Depinet, Bajaj, Elsholz, Webb, Grundmeier.
Acquisition, analysis, or interpretation of data: Alpern, Scott, Balamuth, Chamberlain, Depinet, Bajaj, Carter, Webb, Campos, Deakyne Davies, Cook, Ungar, Grundmeier.
Drafting of the manuscript: Alpern, Scott, Chamberlain, Bajaj, Cook, Grundmeier.
Critical review of the manuscript for important intellectual content: Scott, Balamuth, Chamberlain, Depinet, Bajaj, Carter, Elsholz, Webb, Campos, Deakyne Davies, Cook, Ungar, Grundmeier.
Statistical analysis: Depinet, Carter, Webb, Cook, Ungar, Grundmeier.
Obtained funding: Alpern, Scott, Bajaj.
Administrative, technical, or material support: Depinet, Bajaj, Elsholz, Campos, Deakyne Davies.
Supervision: Bajaj, Cook.

Conflict of Interest Disclosures: None reported.

Funding/Support: This work was supported by National Institute of Child Health and Human Development grant R01HD087363. PECARN is supported by the Health Resources and Services Administration of the US Department of Health and Human Services in the Maternal and Child Health Bureau under the Emergency Medical Services for Children program through the following cooperative agreements: Emergency Medical Services for Children Data Center–University of Utah, Great Lakes Emergency Medical Services for Children Research Node–Nationwide Children's Hospital, Hospitals of the Midwest Emergency Research Node–Cincinnati Children's Hospital Medical Center, Pediatric Emergency Medicine Northeast, West and South–Columbia University Medical Center, Pediatric Research in Injuries and Medical Emergencies–University of California at Davis Medical Center, Charlotte, Houston, Milwaukee Prehospital Emergency Medical Services Research Node–State University of New York at Buffalo, West/Southwest Pediatric Emergency Medicine Research–Seattle Children's Hospital, and San Francisco-Oakland, Providence, Atlanta Research Collaborative–Rhode Island Hospital/Hasbro Children's Hospital.

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.

Group Information: The PECARN PED Screen Study Group members include Elizabeth R. Alpern, MD, MSCE; Halden F. Scott, MD, MSCS; Fran Balamuth, MD, PhD, MSCE; James M. Chamberlain, MD; Holly Depinet, MD MPH; Lalit Bajaj, MD, MPH; Norma-Jean E. Simon, MS; Camille P. Carter, BS; Cara Elsholz, MS; Michael Webb, MS; Diego Campos, MS; Sara J. Deakyne Davies, MPH; Lawrence J. Cook, PhD; Lyle Ungar, PhD; Robert Grundmeier, MD; Brent Husley, BS; Cody Olsen, MS ; Aaron J. Masino, PhD; and Hari Santhanam, BS.

Disclaimer: This information or content and conclusions are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by US Health Resources and Services Administration, US Department of Health and Human Services, or the US government.

Data Sharing Statement: See Supplement 2.

REFERENCES

1. Watson RS, Carrol ED, Carter MJ, Kissoon N, Ranjit S, Schlapbach LJ. The burden and contemporary epidemiology of sepsis in children. *Lancet Child Adolesc Health*. 2024;8(9):670-681. doi:10.1016/S2352-4642(24)00140-8

2. Tan B, Wong JJ, Sultana R, et al. Global case-fatality rates in pediatric severe sepsis and septic shock: a systematic review and meta-analysis. *JAMA Pediatr*. 2019;173(4):352-362. doi:10.1001/jamapediatrics.2018.4839

3. Weiss SL, Balamuth F, Hensley J, et al. The epidemiology of hospital death following pediatric severe sepsis: when, why, and how children with

sepsis die. *Pediatr Crit Care Med*. 2017;18(9):823-830. doi:10.1097/PCC.0000000000001222

4. Weiss SL, Fitzgerald JC, Balamuth F, et al. Delayed antimicrobial therapy increases mortality and organ dysfunction duration in pediatric sepsis. *Crit Care Med*. 2014;42(11):2409-2417. doi:10.1097/CCM.0000000000000509

5. Launay E, Gras-Le Guen C, Martinot A, et al. Suboptimal care in the initial management of children who died from severe bacterial infection: a population-based confidential inquiry. *Pediatr Crit Care Med*. 2010;11(4):469-474. doi:10.1097/PCC.0b013e3181ce752e

6. Ninis N, Phillips C, Bailey L, et al. The role of healthcare delivery in the outcome of meningococcal disease in children: case-control study of fatal and non-fatal cases. *BMJ*. 2005;330(7506):1475. doi:10.1136/bmj.330.7506.1475

7. Souganidis E, Abbadessa MK, Ku B, et al. Analysis of missed sepsis patients in a pediatric emergency department with a vital sign-based electronic sepsis alert. *Pediatr Emerg Care*. 2022;38(1):e1-e4. doi:10.1097/PEC.0000000000002207

8. Schlapbach LJ, MacLaren G, Festa M, et al; Australian & New Zealand Intensive Care Society (ANZICS) Centre for Outcomes & Resource Evaluation (CORE) and Australian & New Zealand Intensive Care Society (ANZICS) Paediatric Study Group. Prediction of pediatric sepsis mortality within 1 h of intensive care admission. *Intensive Care Med*. 2017;43(8):1085-1096. doi:10.1007/s00134-017-4701-8

9. Schlapbach LJ, Straney L, Bellomo R, MacLaren G, Pilcher D. Prognostic accuracy of age-adapted SOFA, SIRS, PELOD-2, and qSOFA for in-hospital mortality among children with suspected infection admitted to the intensive care unit. *Intensive Care Med*. 2018;44(2):179-188. doi:10.1007/s00134-017-5021-8

10. Kamran F, Tjandra D, Heiler A, et al. Evaluation of Sepsis Prediction Models before Onset of Treatment. *NEJM AI*. 2024;1(3). doi:10.1056/AIoa2300032

11. Scott HF, Colborn KL, Sevick CJ, et al. Development and validation of a model to predict pediatric septic shock using data known 2 hours after hospital arrival. *Pediatr Crit Care Med*. 2021;22(1):16-26. doi:10.1097/PCC.0000000000002589

12. Scott HF, Colborn KL, Sevick CJ, et al. Development and validation of a predictive model of the risk of pediatric septic shock using data known at the time of hospital arrival. *J Pediatr*. 2020;217:145-151.e6. doi:10.1016/j.jpeds.2019.09.079

13. Le S, Hoffman J, Barton C, et al. Pediatric severe sepsis prediction using machine learning. *Front Pediatr*. 2019;7:413. doi:10.3389/fped.2019.00413

14. Schlapbach LJ, Watson RS, Sorce LR, et al; Society of Critical Care Medicine Pediatric Sepsis Definition Task Force. International consensus criteria for pediatric sepsis and septic shock. *JAMA*. 2024;331(8):665-674. doi:10.1001/jama.2024.0179

15. Sanchez-Pinto LN, Bennett TD, DeWitt PE, et al; Society of Critical Care Medicine Pediatric Sepsis Definition Task Force. Development and validation of the Phoenix criteria for pediatric sepsis and septic shock. *JAMA*. 2024;331(8):675-686. doi:10.1001/jama.2024.0196

16. 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

17. Ungar LH, Grundmeier RW. Pediatric sepsis. Published March 13, 2023. December 17, 2024. https://osf.io/qshbz/

18. Pantell RH, Roberts KB, Adams WG, et al; Subcommittee on Febrile Infants. Evaluation and management of well-appearing febrile infants 8 to 60 days old. *Pediatrics*. 2021;148(2):e2021052228. doi:10.1542/peds.2021-052228

19. Kuppermann N, Dayan PS, Levine DA, et al; Febrile Infant Working Group of the Pediatric Emergency Care Applied Research Network (PECARN). A clinical prediction rule to identify febrile infants 60 days and younger at low risk for serious bacterial infections. *JAMA Pediatr*. 2019;173(4):342-351. doi:10.1001/jamapediatrics.2018.5501

20. McDaniel CE, Kerns E, Jennings B, et al; AAP REVISE II QI Collaborative. Improving guideline-concordant care for febrile infants through a quality improvement initiative. *Pediatrics*. 2024;153(5):e2023063339. doi:10.1542/peds.2023-063339

21. Radhakrishnan L, Carey K, Hartnett KP, et al. Pediatric emergency department visits before and during the COVID-19 pandemic—United States, January 2019-January 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71(8):313-318. doi:10.15585/mmwr.mm7108e1

22. Balamuth F, Scott HF, Weiss SL, et al; Pediatric Emergency Care Applied Research Network (PECARN) PED Screen and PECARN Registry Study Groups. Validation of the Pediatric Sequential Organ Failure Assessment Score and evaluation of Third International Consensus Definitions for sepsis and septic shock definitions in the pediatric emergency department. *JAMA Pediatr*. 2022;176(7):672-678. doi:10.1001/jamapediatrics.2022.1301

23. Gilboy NTP, Travers D. *Rosenau AM Emergency Severity Index(ESI): A Triage Tool for Emergency Department Care, Version 4, Implementation Handbook 2012 Edition*. Agency for Healthcare Reseach and Quality; 2011.

24. 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(1):199. doi:10.1186/1471-2431-14-199

25. Weiss SL, Balamuth F, Chilutti M, et al. Identification of pediatric sepsis for epidemiologic surveillance using electronic clinical data. *Pediatr Crit Care Med*. 2020;21(2):113-121. doi:10.1097/PCC.0000000000002170

26. Index for rating diagnostic tests. Wiley Online Library. Accessed December 17, 2024. https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142(1950)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3

27. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011-2022). *Comput Methods Programs Biomed*. 2022;226:107161. doi:10.1016/j.cmpb.2022.107161

28. Bitbucket. Updated August 21, 2025. https://stash.utahdcc.org/stash/projects/PP/repos/predictive-model-for-early-pediatric-sepsis/browse

29. 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

30. Eisenberg M, Madden K, Christianson JR, Melendez E, Harper MB. Performance of an automated screening algorithm for early detection of pediatric severe sepsis. *Pediatr Crit Care Med*. 2019;20(12):e516-e523. doi:10.1097/PCC.0000000000002101

31. Lloyd JK, Ahrens EA, Clark D, Dachenhaus T, Nuss KE. Automating a manual sepsis screening tool in a pediatric emergency department. *Appl Clin Inform*. 2018;9(4):803-808. doi:10.1055/s-0038-1675211

32. Georgette N, Michelson K, Monuteaux M, Eisenberg M. A temperature- and age-adjusted shock index for emergency department identification of pediatric sepsis. *Ann Emerg Med*. 2023;82(4):494-502. doi:10.1016/j.annemergmed.2023.03.026

33. Zimmerman JJ, Banks R, Berg RA, et al; Life After Pediatric Sepsis Evaluation (LAPSE) Investigators. Trajectory of mortality and health-related quality of life morbidity following community-acquired pediatric septic shock. *Crit Care Med*. 2020;48(3):329-337. doi:10.1097/CCM.0000000000004123

34. Zimmerman JJ, Banks R, Berg RA, et al; Life After Pediatric Sepsis Evaluation (LAPSE) Investigators. Critical illness factors associated with long-term mortality and health-related quality of life morbidity following community-acquired pediatric septic shock. *Crit Care Med*. 2020;48(3):319-328. doi:10.1097/CCM.0000000000004122

35. Paul R, Niedner M, Riggs R, et al; IPSO Collaborative Investigators. Bundled care to reduce sepsis mortality: the Improving Pediatric Sepsis Outcomes (IPSO) Collaborative. *Pediatrics*. 2023;152(2):e2022059938. doi:10.1542/peds.2022-059938

36. Collins GS. Making the black box more transparent: improving the reporting of artificial intelligence studies in healthcare. *BMJ*. 2024;385:q832. doi:10.1136/bmj.q832

37. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801-810. doi:10.1001/jama.2016.0287

38. Stevens AF, Stetson P. Theory of trust and acceptance of artificial intelligence technology (TrAAIT): an instrument to assess clinician trust and acceptance of artificial intelligence. *J Biomed Inform*. 2023;148:104550. doi:10.1016/j.jbi.2023.104550

39. Shamszare H, Choudhury A. Clinicians' perceptions of artificial intelligence: focus on workload, risk, trust, clinical decision making, and clinical integration. *Healthcare (Basel)*. 2023;11(16):2308. doi:10.3390/healthcare11162308

40. van der Vegt AH, Scott IA, Dermawan K, Schnetler RJ, Kalke VR, Lane PJ. Implementation frameworks for end-to-end clinical AI: derivation of the SALIENT framework. *J Am Med Inform Assoc*. 2023;30(9):1503-1515. doi:10.1093/jamia/ocad088

41. Cruz AT, Lane RD, Balamuth F, et al. Updates on pediatric sepsis. *J Am Coll Emerg Physicians Open*. 2020;1(5):981-993. doi:10.1002/emp2.12173

42. Sanchez-Pinto LN, Del Pilar Arias López M, Scott H, et al. Digital solutions in paediatric sepsis: current state, challenges, and opportunities to improve care around the world. *Lancet Digit Health*. 2024;6(9):e651-e661. doi:10.1016/S2589-7500(24)00141-9