OPEN

Prediction of Impending Septic Shock in Children With Sepsis

Ran Liu, BS[1,2]
Joseph L. Greenstein, PhD[1,2]
James C. Fackler, MD[3]
Jules Bergmann, MD[3]
Melania M. Bembea, MD, MPH, PhD[3,4]
Raimond L. Winslow, PhD[1,2]

OBJECTIVES: Sepsis and septic shock are leading causes of in-hospital mortality. Timely treatment is crucial in improving patient outcome, yet treatment delays remain common. Early prediction of those patients with sepsis who will progress to its most severe form, septic shock, can increase the actionable window for interventions. We aim to extend a time-evolving risk score, previously developed in adult patients, to predict pediatric sepsis patients who are likely to develop septic shock before its onset, and to determine whether or not these risk scores stratify into groups with distinct temporal evolution once this prediction is made.

DESIGN: Retrospective cohort study.

SETTING: Academic medical center from July 1, 2016, to December 11, 2020.

PATIENTS: Six-thousand one-hundred sixty-one patients under 18 admitted to the Johns Hopkins Hospital PICU.

INTERVENTIONS: None.

MEASUREMENTS AND MAIN RESULTS: We trained risk models to predict impending transition into septic shock and compute time-evolving risk scores representative of a patient's probability of developing septic shock. We obtain early prediction performance of 0.90 area under the receiver operating curve, 43% overall positive predictive value, patient-specific positive predictive value as high as 62%, and an 8.9-hour median early warning time using Sepsis-3 labels based on age-adjusted Sequential Organ Failure Assessment score. Using spectral clustering, we stratified pediatric sepsis patients into two clusters differing in septic shock prevalence, mortality, and proportion of patients adequately fluid resuscitated.

CONCLUSIONS: We demonstrate the applicability of our methodology for early prediction and stratification for risk of septic shock in pediatric sepsis patients. Through analyses of risk score evolution over time, we corroborate our past finding of an abrupt transition preceding onset of septic shock in children and are able to stratify pediatric sepsis patients using their risk score trajectories into low and high-risk categories.

KEY WORDS: cluster analysis; electronic health records; intensive care units, pediatric; machine learning; sepsis; shock, septic

Sepsis and septic shock are leading causes of in-hospital mortality for both adults and children worldwide (1, 2). Kumar et al (3) showed in adult septic shock patients that every hour of delayed treatment is associated with an 8% increase in mortality. Weiss et al (4) found this same association in pediatric patients, and that treatment delays remain common.

The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) (5) reflect the most recent understanding of adult sepsis as organ dysfunction caused by dysregulated immune response to infection. However, consensus definitions for pediatric sepsis were last updated in 2005 (6) and closely modeled the adult Sepsis-2 definitions (7). Recently, individual groups have proposed age-adjusted Sepsis-3 criteria for children (8–10), and an international expert group

DOI: 10.1097/CCE.0000000000000442

published new guidelines for the treatment of pediatric septic shock and its associated organ dysfunction (11).

Several computational approaches for early prediction of sepsis and septic shock using electronic health record (EHR) data have been developed with the aim of reducing treatment delays in adults (12–15). We predicted impending transition from sepsis to septic shock based on the hypothesis that there exists a physiologically distinct state of sepsis, which we termed "preshock," and that entry into this state presages the onset of septic shock (13, 15). Through analyses of the temporal evolution of patient state, we stratified adult sepsis patients by outcomes and interventions received and discovered that entry into preshock was marked by a rapid shift in both patient physiology and risk occurring within a 30–60 minutes timeframe (16).

In this study, we evaluated age-adjusted Sepsis-3 criteria and applied our previously published method for early prediction of septic shock to patients admitted to an academic, quaternary center PICU. We corroborated our past finding of an abrupt transition preceding septic shock onset in children and stratified sepsis patients using their risk score trajectories into low- and high-risk categories.

METHODS

Study Population

We conducted a retrospective observational cohort study of all patients admitted to the Johns Hopkins PICU beginning July 1, 2016, discharged before December 11, 2020. Patients 18 years old and older were excluded. The Johns Hopkins Medicine Institutional Review Board approved the study (Protocol IRB00258534) with a waiver of consent. We also present results on the publicly available pediatric intensive care (PIC) dataset (17), containing EHR data from patients admitted to ICUs at the Children's Hospital, Zhejiang University School of Medicine, between 2010 and 2018 (**Results in an Independent Cohort**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647).

Data Extraction and Processing

Raw data were sourced from an EHR data report and included patient demographics, encounter diagnosis codes, admit-discharge-transfer codes (with patient room/bed assignments), provider-entered flowsheets (which included nurse-validated vitals and respiratory therapist-validated ventilator settings and measurements), a subset of laboratory results, medication orders, and medication administrations. Missing values were imputed using a Bayesian structural time series model (18, 19) trained for each feature.

Labeling Clinical States

Suspected infection was determined using the presence of concomitant orders for antibiotics and blood cultures, as specified by Seymour et al (20). Comorbidities (**Table S1**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647) were computed according to the Pediatric Complex Chronic Condition Classification (21, 22). Labels were reevaluated at each new observation of clinical data, and if no prior observations of a feature were available, the patient was assumed to be within normal ranges.

According to the Goldstein consensus criteria (6), sepsis is defined as suspected infection and two or more age-adjusted systemic inflammatory response syndrome (SIRS) criteria (**Labeling Clinical States**, **Table S2**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Septic shock is defined as sepsis with cardiovascular dysfunction.

According to the Sepsis-3 criteria, sepsis is defined as organ dysfunction consequent to suspected infection. We determine organ dysfunction as a 2-point rise in age-adjusted Sequential Organ Failure Assessment (SOFA) score, as defined by Matics et al (9) by using Pediatric Logistic Organ Dysfunction-2 (PELOD-2) (23) cut offs for mean arterial pressure (MAP) and creatinine, respectively (**Table S3**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647), and by a 2-point rise or a 6-point rise in PELOD-2 (23). Septic shock patients are sepsis patients adequately fluid resuscitated, administered vasopressors, and exhibiting serum lactate greater than 2 mmol/L. Fluid resuscitation was determined using the 2020 Surviving Sepsis Campaign (SSC) pediatric guidelines (11), defined as 40 mL/kg of fluids in a 3-hour window, or having attained the resuscitation target of MAP at the fifth percentile or higher for age, estimated as 1.5 × age in years + 40 mm Hg (24).

Risk Modeling and Prediction

Risk models were built as previously described, using 26 features extracted from EHR data (13). MAP and heart rate were normalized to percentile values by age (24, 25). In order to characterize the preshock state, XGBoost and

generalized linear models (GLMs) were trained using data from sepsis in patients who do not develop septic shock and from a time window spanning 100 minutes prior to septic shock onset to 1 minute prior in septic shock patients. Lasso regularization was used for feature selection in GLM. We also compared the use of Cox proportional hazards modeling (26) to compute the risk score, as well as the use of age-adjusted SOFA score alone.

All four scores (XGBoost, GLM, Cox, SOFA) were calculated at each time where there are EHR data. Prediction of impending septic shock occurs when a patient's risk score first exceeds a threshold value, determined from training data as the threshold (for a given model) corresponding to the point on the receiver operating characteristic (ROC) curve closest to the top left. One result is generated per hospital admission: a true positive is a patient who develops septic shock and whose risk score exceeds the threshold prior to septic shock onset; a true negative is a patient who never develops septic shock and whose risk score always remains below the threshold. Early warning time (EWT) is defined as the difference between the time when the risk score crosses threshold and time of septic shock onset. CIs for performance criteria were estimated using bootstrap, for 100 iterations: bootstrapped datasets were generated by sampling hospital admissions with replacement.

Stratification of Sepsis Patients

Stratification of sepsis patients by risk score trajectories was performed as previously described and is a separate analysis from early prediction (16). Risk score trajectories for each patient were computed by applying the XGBoost model at each of the 12 hours following time of threshold crossing. Spectral clustering was applied to stratify patients into clusters with similar risk trajectories following early prediction (27, 28) (**Spectral Clustering**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647).

RESULTS

The dataset contains EHR data from 6,560 distinct patients and 9,330 hospital admissions. Excluding patients 18 years old and older yields our analysis set of 6,161 patients and 8,630 hospital admissions, with overall sepsis prevalence of 17.58%, as determined by age-adjusted Sepsis-3 criteria (**Table 1**) (8, 9). Availability of data and central tendency measures are given in **Table S4** (Supplemental Digital Content 1, http://links.lww.com/CCX/A647) and **Table S5** (Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Data entries are comprised of timestamp-value pairs, with identifier (ID) numbers for patients, hospital admissions, and PICU stays, as well as an ID indicating the feature associated with each value (**Table S6**, **S12**, and **S13**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Seventy percent of patients were uniformly sampled into the training set, and the remaining 30% reserved for testing.

Baseline Statistics

We applied four sets of diagnostic criteria for distinguishing sepsis and septic shock: the 2005 Goldstein

TABLE 1.
Baseline Statistics of the Dataset

Most Severe Clinical State Reached	No Sepsis	Sepsis Without Septic Shock	Sepsis Leading to Septic Shock	Overall
Admissions, *n* (%)	7,113 (82.42)	1,203 (13.94)	314 (3.64)	8,630 (100)
Patients, *n* (%)	4,921 (79.87)	938 (15.22)	302 (4.90)	6,161 (100)
PICU stays, *n* (%)	7,487 (78.74)	1,522 (16.01)	500 (5.26)	9,509 (100)
In-hospital mortality, *n* (%)	61 (0.86)	51 (4.24)	74 (23.57)	186 (2.16)
Gender, %	55.76 male, 44.24 female	55.78 male, 44.22 female	54.14 male, 45.86 female	55.70 male, 44.30 female
Median ICU stay length, d (interquartile range)	1.55 (0.90–2.91)	5.90 (2.47–15.45)	12.09 (3.59–33.46)	1.80 (0.95–3.91)
Mean age, yr (SD)	6.44 (5.73)	4.12 (5.39)	4.27 (5.80)	6.04 (5.75)

Sepsis cohorts are determined using age-adjusted Sequential Organ Failure Assessment score.

consensus criteria (6) and three others evaluated by Schlapbach et al (8). Using these labels, we computed the prevalence and mortality of each cohort (**Table 2**). Depending on the criteria, sepsis prevalence varies between 7.65% and 26.02%, and septic shock prevalence ranges from 2.32% to 17.80%. We find that sepsis patients labeled using age-adjusted Sepsis-3 criteria have greater mortality than those labeled using the Goldstein criteria. When using age-adjusted SOFA scores to determine clinical state labels, mortality in both sepsis cohorts (sepsis without shock and septic shock patients) is higher than in the corresponding cohort determined using the SIRS-based Goldstein criteria. Defining organ dysfunction by an increase of 6 points or greater in PELOD-2 score results in the lowest prevalence of sepsis and the highest mortality in both sepsis cohorts. The Goldstein criteria result in the highest prevalence of sepsis but the lowest mortality in both sepsis cohorts.

Early Prediction of Septic Shock

Figure 1 shows risk score trajectories using XGBoost (29) from patients with sepsis who do (Fig. 1*A*) and do not (Fig. 1*B*) develop septic shock. The risk score of the patient who does not develop shock remains below the threshold, whereas the patient who ultimately develops shock has a risk score which rapidly increases above the threshold in advance of septic shock onset.

Early prediction of septic shock was evaluated for two machine learning methods, GLM (30) and XGBoost (29). **Figure 2** shows the ROC curves, precision-recall curves, and performance metrics using age-adjusted SOFA scores for labels. **Figure S6** (Supplemental Digital Content 1, http://links.lww.com/CCX/A647) shows the calibration curves for these two models. We evaluated early prediction using all clinical criteria and found that best performance was obtained with labels using age-adjusted SOFA scores (**Fig. S1**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). In the held-out test set, XGBoost yields greatest performance of 0.90 area under the ROC curve (AUC), compared with 0.87 with GLM, 0.82 with the Cox proportional hazards model, and 0.72 with SOFA score (**Table S7**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). With XGBoost, the threshold chosen yields a median EWT of 8.9 hours and 43% average positive predictive value (PPV) (77 true positives, 98 false positives, 481 true negatives, 13 false negatives). Feature importance is given in **Table S8** (Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Early prediction using both XGBoost and GLM risk scores yields greater performance than a Cox proportional hazards model, and all three models yield greater performance than SOFA score alone. Applying fitted models to the external dataset obtained from PIC, we obtain moderate performance in early prediction of septic shock in an independent cohort (**Fig. 3**). In this dataset, XGBoost also yields the greatest performance (**Table S14**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647), with 0.82 AUC, 48 hours median EWT, and 22% overall PPV (280 true positives, 989 false positives, 2438 true negatives, 72 false negatives).

TABLE 2.
Comparison of Diagnostic Criteria

Criteria	Nonsepsis, *n* (%)	Sepsis Without Shock, *n* (%)	Septic Shock, *n* (%)
Goldstein	6,385 (73.99)	709 (8.22)	1,536 (17.80)
	50 (0.78)	**8 (1.13)**	**128 (8.33)**
Age-adjusted Sequential Organ Failure Assessment	7,113 (82.42)	1,203 (13.94)	314 (3.64)
	61 (0.86)	**51 (4.24)**	**74 (23.57)**
PELOD-2 (2 points)	6,480 (75.09)	1,777 (20.59)	373 (4.32)
	50 (0.77)	**58 (3.26)**	**78 (20.91)**
PELOD-2 (6 points)	7,970 (92.35)	460 (5.33)	200 (2.32)
	77 (0.97)	**43 (9.35)**	**66 (33.00)**

PELOD, Pediatric Logistic Organ Dysfunction.

Hospital admission counts and proportion of all admissions represented by each cohort in parentheses. Counts for age-adjusted Sequential Organ Failure Assessment correspond to the admission counts given in Table 1. Hospital admission counts ending in mortality are given in bold, with the proportion of all admissions within each cohort ending in mortality in parentheses.

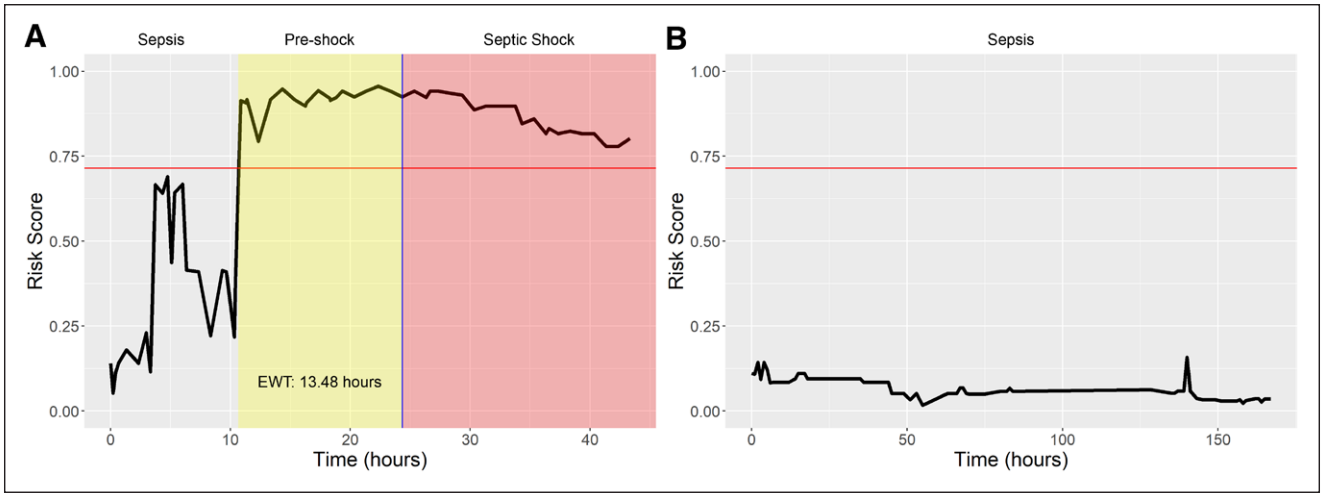

Figure 1. Example risk trajectories for (**A**) a patient who developed septic shock and (**B**) a nonshock sepsis patient. Threshold for early prediction is indicated by the *red horizontal line* (with a value of 0.715), and time of septic shock onset is indicated by the *blue vertical line*.

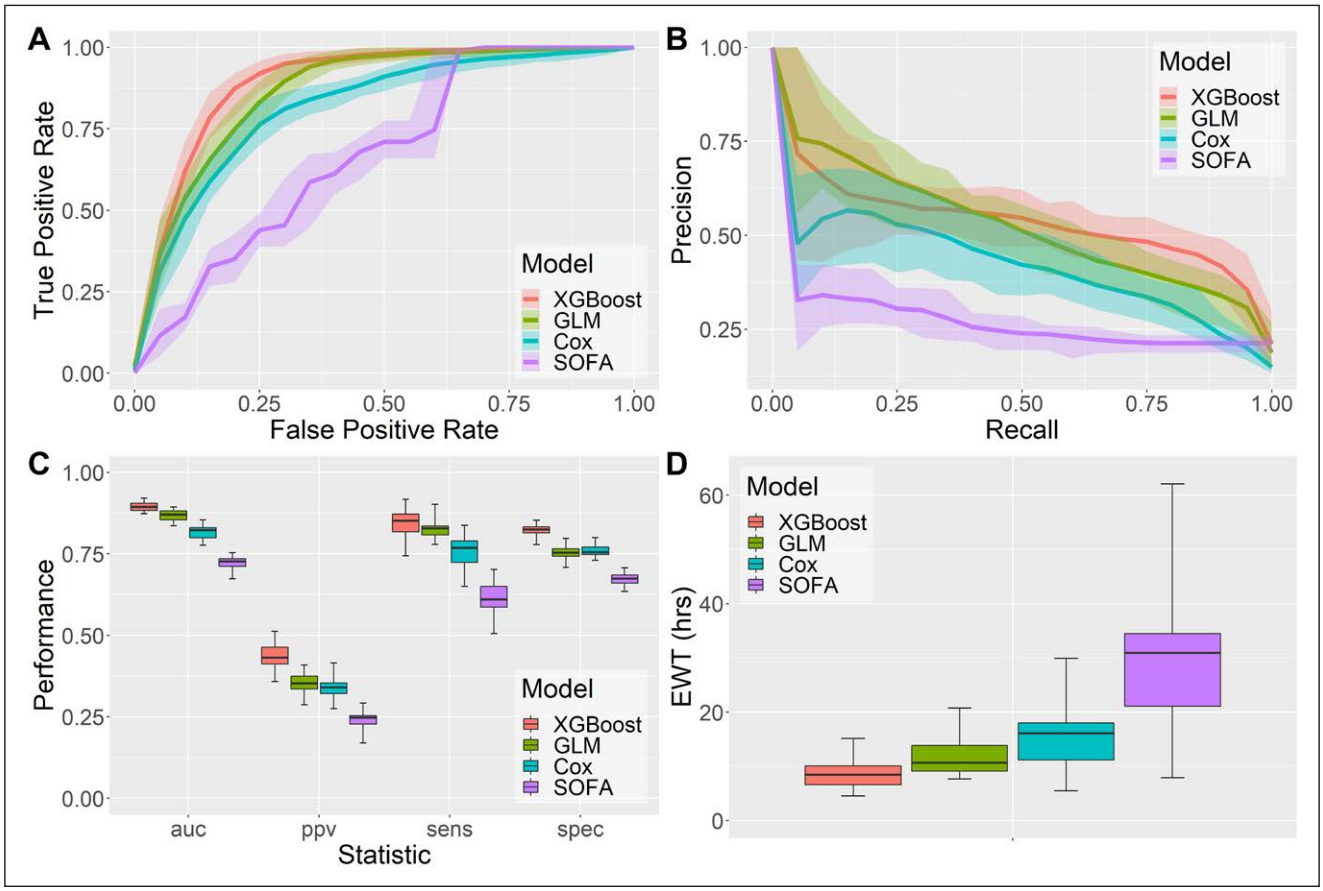

Figure 2. Performance of early prediction of septic shock in pediatric patients: (**A**) receiver operating characteristic (ROC) curves and 90% CIs, (**B**) precision-recall curves and 90% CIs, (**C**) area under the ROC curve (AUC), positive predictive value (PPV), sensitivity (sens), and specificity (spec), (**D**) early warning time (EWT) for generalized linear model (GLM), XGBoost, Cox, and Sequential Organ Failure Assessment (SOFA) score, with box plots indicating minimum, maximum, median, and first and third quartiles.

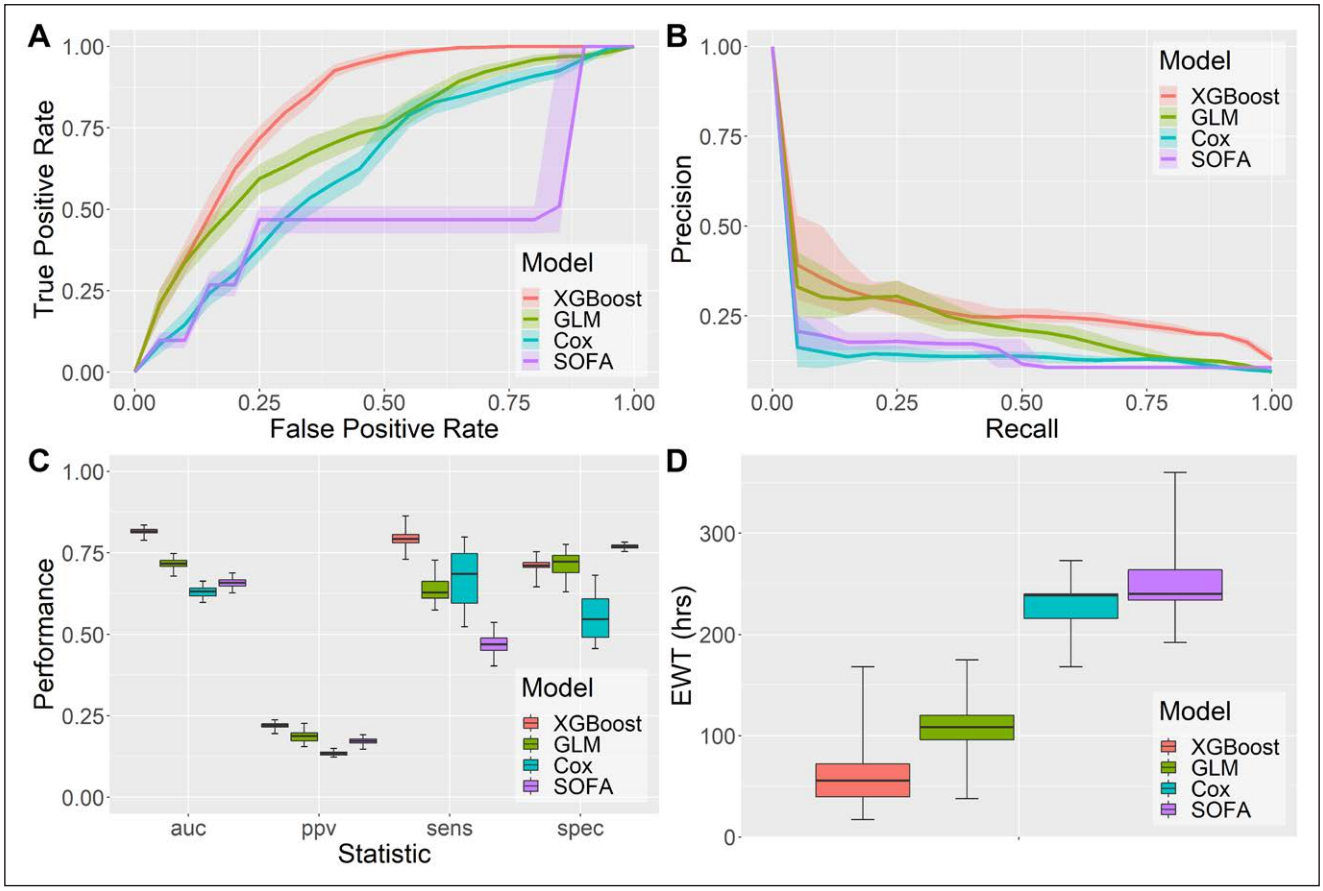

Figure 3. Performance of early prediction of septic shock in the pediatric intensive care dataset: (**A**) receiver operating characteristic (ROC) curves and 90% CIs, (**B**) precision-recall curves and 90% CIs, (**C**) area under the ROC curve (AUC), positive predictive value (PPV), sensitivity (sens), and specificity (spec), (**D**) early warning time (EWT) for generalized linear model (GLM), XGBoost, Cox, and Sequential Organ Failure Assessment (SOFA) score with box plots indicating minimum, maximum, median, and first and third quartiles.

Stratification of Sepsis Patients

As previously described (13), our approach to early prediction allows calculation of a patient-specific PPV based on the first value of risk score that exceeds the threshold. These values were binned into quintiles, and PPV was computed for patients in each bin, estimating the probability that a prediction of impending septic shock onset for a patient whose risk score falls into the specified range is a true positive (**Table S9**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). In higher quintiles, the likelihood that predictions are true positives is greater than in lower quintiles, with PPV as high as 62%.

We repeated our analyses of risk score trajectories for stratification of sepsis patients (16). Spectral clustering (27) of risk score trajectories in the window surrounding early prediction yielded two clusters (**Fig. 4**) (**Fig. S2**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Patient risk trajectories are indistinguishable before time of early prediction. Risk scores increase abruptly at the time of threshold crossing for all patients, and clusters diverge subsequently. Separation between the two clusters is quantified by Kullback-Leibler (KL) divergence (31) in each time window (**Fig. S3**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). KL divergence quantifies the separation between two probability distributions: two identical (and thus indistinguishable) distributions will have a KL divergence of 0, and a pair of distributions which have a greater degree of overlap will have a lower KL divergence than a pair of distributions with less overlap. The evolution of lactate and Glasgow Coma Scale (GCS) (**Fig. S4**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647), two important physiologic features in the risk models (Table S8, Supplemental Digital Content 1, http://links.lww.com/CCX/A647), are similar to that of risk

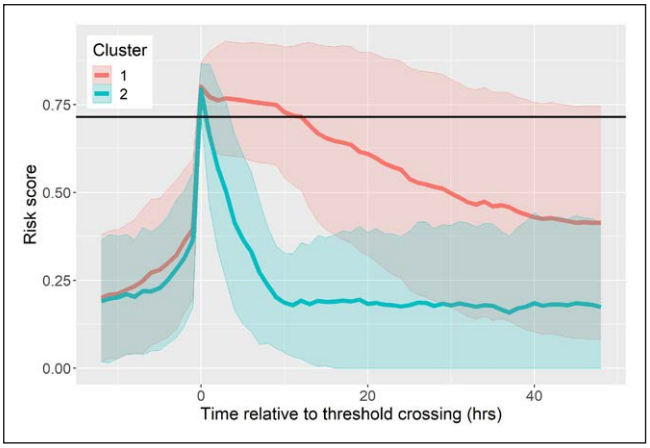

Figure 4. Results of spectral clustering applied to risk score trajectories that cross the threshold for early prediction (*black horizontal line*). *Solid lines* indicate mean risk scores within each cluster, whereas the *shaded areas* indicate an interval of 1 sd from the mean.

score, with increasing divergence between risk clusters following threshold crossing. Clusters stratify by prevalence of septic shock, mortality, and the proportion of patients who are adequately fluid resuscitated prior to time of early warning (**Table S10**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647).

DISCUSSION

Pediatric Sepsis Criteria

The shortcomings of SIRS-based criteria for sepsis are well-known. SIRS (Table S2, Supplemental Digital Content 1, http://links.lww.com/CCX/A647) is not specific for infection, and over 90% of adult intensive care patients meet the criteria for SIRS (32–34). The Sepsis-3 criteria for adults redefined sepsis as infection resulting in organ dysfunction, determined by an increase of at least 2 points in SOFA score (35, 36). Consequently, there has been strong interest in redefining pediatric sepsis on the basis of organ dysfunction. Leclerc et al (10) suggested the use of PELOD-2 scores in children with suspected infection, and Matics et al (9) suggested an age-adjusted SOFA score (Table S3, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Pediatric Sepsis-3 criteria based on age-adjusted SOFA score or PELOD-2 produce labels which result in greater performance in early prediction than the SIRS-based Goldstein criteria (Fig. S1, Supplemental Digital Content 1, http://links.lww.com/CCX/A647), indicating that those criteria yield clinical states which are more physiologically distinct. Furthermore, we corroborate the findings of other groups that age-adjusted SOFA and PELOD-2 have greater validity in stratifying patients by metrics of disease severity than the SIRS-based Goldstein criteria (**Fig. S5**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647) (8).

Prediction Method

As in adults, the goal of early prediction in children is to provide clinicians with a time window of intervention, enabling more timely treatment and possibly preventing development of shock. Model performance using XGBoost (29) is higher than in our prior study of adult patients and is higher than that using GLM (30). XGBoost uses gradient boosting of decision trees and can learn nonlinear associations between features and risk. This likely yields its improved performance compared with GLM, a finding also consistent with our previous results.

In our previous work, we introduced the notion of patient-specific PPV, where patient risk scores are stratified based on the first postthreshold crossing value of risk (13). Because this study contains fewer patients than MIMIC-III and eICU (6,161 vs 38,418 vs 139,367), and the prevalence of sepsis and septic shock is lower in children, we stratified positive predictions into quintiles, rather than deciles. However, we demonstrate that our method for estimating the reliability of a positive prediction remains applicable in pediatric sepsis patients. Ultimately, we envision the application of this methodology in a prospective, real-time setting, where clinicians are privy to the predictions generated by the model, as well as patient-specific PPVs. This enables clinicians to act only on highly reliable predictions or to take different courses of action informed with the likelihood that a patient will develop septic shock. The goal is to give clinicians information that is novel, rather than merely to confirm what may already be apparent (37). Inquiry into how clinicians incorporate information from risk scores into clinical decision-making is needed. We suspect that if a clinician with a low suspicion of septic shock is presented with a low risk score, they may be more confident in not administering additional fluids or vasopressors. Conversely, if the risk score is high, and the clinician concurs that the patient is at high risk, then earlier administration of fluids and vasopressors may potentially mitigate sepsis-related morbidity.

Stratification of Patients

Previously, we found that entry into preshock was marked by a rapid transition from low to high risk. Prior to entry, patient physiology was indistinguishable between the low- and high-risk clusters (16). However, after entry into preshock, risk score trajectories diverged and stratified patients by risk of septic shock, mortality, time to septic shock onset, and treatments received. We confirm our past finding that very rapid transitions in risk score occur in pediatric as well as adult sepsis patients, with changes in physiology reflecting changes in risk score (Fig. S4, Supplemental Digital Content 1, http://links.lww.com/CCX/A647), and that risk score trajectories can stratify pediatric sepsis patients by risk of septic shock, mortality, and the proportion of patients adequately fluid resuscitated at time of threshold crossing.

We found no statistically significant difference in EWT or the proportion of patients treated with vasopressors prior to entry into preshock (Table S10, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Pediatric patients in our single-center study may be more homogenous and receive more uniform care than adult patients in the eICU dataset, who may be admitted with more confounding conditions and may receive different treatments across hundreds of different hospitals. Nonetheless, these findings support our postulation of the preshock state and that entry into septic shock is extremely rapid in both adult and pediatric sepsis patients. The rapid nature of the transition further indicates the necessity of automated methods for the detection and prediction of septic shock.

Risk score is computed using physiologic variables, and thus, the evolution of these variables leads to dynamic variations in risk score. However, the discernable magnitude of changes in physiologic variables upon entry into the preshock state is smaller than in risk score itself, as quantified by KL divergence (Fig. S3, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Entry into the preshock state is often reflected by a change in many physiologic variables rather than a single variable, and thus, this shift in patient state is best captured by risk score. We choose spectral clustering (27, 28) in order to cluster the time series data. Spectral clustering identifies clusters such that distance between members of the same cluster is minimized. This methodology can produce clusters with nonlinear decision boundaries and has good empirical performance on a variety of data (38).

Model Interpretation

It is possible to determine the importance of features in both XGBoost and GLM models (Table S8, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). For XGBoost, gain is the normalized average increase in performance resulting from the addition of a feature. Coverage is the normalized frequency at which decision tree nodes which split on a feature are reached. Frequency is the normalized proportion of decision trees where features appear. Exponentiated GLM coefficients can be interpreted as odds ratios. For example, the exponentiated coefficient of lactate is 5.80. Therefore, a patient with serum lactate 1 SD above the population mean is over five times as likely to develop septic shock as a patient with average serum lactate.

Both models share their top three features: lactate, respiratory SOFA, and GCS. Lactate is the most important feature in both models, as was true in our previous study. These findings align with literature on sepsis pathophysiology. Elevated serum lactate indicates reduced tissue perfusion and predicts mortality in patients with infections (39, 40). Increased respiratory SOFA is associated with respiratory dysfunction (35). GCS is associated with neurologic function, known to be affected in pediatric sepsis (41, 42).

Limitations

We algorithmically determine clinical labels according to proposed pediatric Sepsis-3 definitions. Therefore, limitations of these labels are also limitations of the study. For example, the Infectious Disease Society of America notes that determining septic shock using adequate fluid resuscitation as part of the criteria results in ambiguity and disagreement on time of onset (43). Deutschman (44) remarks that the Sepsis-3 criteria may not encompass the entire pathophysiology of sepsis, which may become life-threatening through mechanisms other than organ dysfunction. These limitations would persist in our analysis of age-adjusted Sepsis-3 in pediatric sepsis patients.

Availability and frequency of EHR data influence the accuracy of our labels and predictions. This is particularly important in the PIC dataset, where median

time between observations for most features is 24 hours (Tables S10 and **S11**, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Furthermore, GCS, a component of both SOFA and PELOD-2, is unavailable in PIC, potentially resulting in cases of neurologic organ dysfunction uncaptured by the labeling criteria. Central venous pressure is also unavailable in PIC. Sparsity of data is a major cause of degraded performance in the PIC dataset although it is not possible to determine whether this is because observations were infrequently made or simply because entry of data into the EHR occurred infrequently. Practices regarding the frequency of tests and data entry vary between different centers of care. However, more frequent measurements, particularly of features with high predictive value, may yield not only improved model performance but also be of general clinical value: Vincent et al (45) suggest that lactate should be measured once every 1–2 hours, and the 2018 SSC update for adults (46) adopts the recommendation of a lactate measurement in the first hour for sepsis patients.

Some features are also not measured in many patients in both the Johns Hopkins and PIC datasets. Lactate, which is the most important feature in both the GLM and XGBoost risk models, is only measured in 37% of patients (Table S4, Supplemental Digital Content 1, http://links.lww.com/CCX/A647). Agniel et al (47) showed that the presence and timing of orders for laboratory tests, independent of test results, was associated with mortality. This would be a potential source of bias in measured laboratory values and affect our computed relative importance of features.

Last, our dataset is limited to patients from a single center of care. Due to manifold deficiencies in the PIC dataset, we do not consider that results obtained therein meet the standard of external validation. However, at present, there are no other publicly available PIC datasets that can be used for this purpose. Although we corroborate the findings of other research groups in other cohorts of patients and achieve moderate performance in an independent cohort, greater validity could be achieved via a multicenter analysis encompassing a greater number of patients treated within diverse settings.

ACKNOWLEDGMENT

We would like to thank Drs. Nauder Faraday and Adam Sapirstein for valuable discussion.

1 Institute for Computational Medicine, The Johns Hopkins University, Baltimore, MD.

2 Department of Biomedical Engineering, The Johns Hopkins University School of Medicine & Whiting School of Engineering, Baltimore, MD.

3 Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

4 Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore, MD.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccejournal).

Supported, in part, by National Science Foundation EECS 1609038 and the National Institutes of Health UL1 TR001079.

The authors have disclosed that they do not have any potential conflicts of interest.

This work was performed at The Johns Hopkins University.

For information regarding this article, E-mail: rwinslow@jhu.edu

REFERENCES

1. Watson RS, Carcillo JA, Linde-Zwirble WT, et al: The epidemiology of severe sepsis in children in the United States. *Am J Respir Crit Care Med* 2003; 167:695–701
2. Angus DC, Linde-Zwirble WT, Lidicker J, et al: Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001; 29:1303–1310
3. Kumar A, Roberts D, Wood KE, et al: Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006; 34:1589–1596
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; 42:2409–2417
5. Singer M, Deutschman CS, Seymour CW, et al: The third international consensus definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; 315:801–810
6. Goldstein B, Giroir B, Randolph A: International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics*. *Pediatric Critical Care Medicine* 2005; 6:2–8
7. Levy MM, Fink MP, Marshall JC, et al; SCCM/ESICM/ACCP/ATS/SIS: 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. *Crit Care Med* 2003; 31:1250–1256
8. Schlapbach LJ, Straney L, Bellomo R, et al: 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; 44:179–188
9. 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:e172352
10. Leclerc F, Duhamel A, Deken V, et al; Groupe Francophone de Réanimation et Urgences Pédiatriques (GFRUP): Can the pediatric logistic organ dysfunction-2 score on day 1 be used

in clinical criteria for sepsis in children? *Pediatr Crit Care Med* 2017; 18:758–763

11. Weiss SL, Peters MJ, Alhazzani W, et al: Surviving sepsis campaign international guidelines for the management of septic shock and sepsis-associated organ dysfunction in children. *Pediatr Crit Care Med* 2020; 21:e52–e106
12. Henry KE, Hager DN, Pronovost PJ, et al: A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015; 7:299ra122
13. Liu R, Greenstein JL, Granite SJ, et al: Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. *Sci Rep* 2019; 9:6145
14. Calvert JS, Price DA, Chettipally UK, et al: A computational approach to early sepsis detection. *Comput Biol Med* 2016; 74:69–73
15. Liu R, Greenstein JL, Sarma SV, et al: Natural language processing of clinical notes for improved early prediction of septic shock in the ICU. *Annu Int Conf IEEE Eng Med Biol Soc* 2019; 2019:6103–6108
16. Liu R, Greenstein JL, Fackler JC, et al: Spectral clustering of risk score trajectories stratifies sepsis patients by clinical outcome and interventions received. *Elife* 2020; 9:e58142
17. Zeng X, Yu G, Lu Y, et al: PIC, a paediatric-specific intensive care database. *Sci Data* 2020; 7:14
18. Brockwell PJ, Brockwell PJ, Davis RA, et al: Introduction to Time Series and Forecasting. Springer, 2016
19. Harvey AC: Forecasting, structural time series models and the Kalman filter. Cambridge, UK, Cambridge University Press, 1990
20. Seymour CW, Liu VX, Iwashyna TJ, et al: Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:762–774
21. Feinstein JA, Russell S, DeWitt PE, et al: R package for pediatric complex chronic condition classification. *JAMA Pediatr* 2018; 172:596–598
22. Feudtner C, Feinstein JA, Zhong W, et al: Pediatric complex chronic conditions classification system version 2: Updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatr* 2014; 14:199
23. Leteurtre S, Duhamel A, Salleron J, et al; Groupe Francophone de Réanimation et d'Urgences Pédiatriques (GFRUP): PELOD-2: An update of the PEdiatric logistic organ dysfunction score. *Crit Care Med* 2013; 41:1761–1773
24. Haque IU, Zaritsky AL: Analysis of the evidence for the lower limit of systolic and mean arterial pressure in children. *Pediatr Crit Care Med* 2007; 8:138–144
25. Bryan S, Saint-Pierre Larose M, Campbell N, et al: Resting blood pressure and heart rate measurement in the Canadian health measures survey, cycle 1. *Health Rep* 2010; 21:71–78
26. Tibshirani R: The lasso method for variable selection in the Cox model. *Stat Med* 1997; 16:385–395
27. Ng AY, Jordan MI, Weiss Y: On spectral clustering: Analysis and an algorithm. *In:* Advances in Neural Information Processing Systems, NeurIPS 2001, Vancouver, BC, Canada, December 3–8, 2001, pp 849–856
28. De Bie T, Cristianini N, Rosipal R: Eigenproblems in pattern recognition. In: Handbook of geometric computing
29. Chen T, Guestrin C: Xgboost: A scalable tree boosting system. *In:* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM, San Francisco, CA, August 13–17, 2016, pp 785–794
30. Tibshirani R: Regression shrinkage and selection via the lasso. *J R Stat Soc Series B (Methodological)* 1996;58:267–288
31. Kullback S, Leibler RA: On Information and Sufficiency. *The Annals of Mathematical Statistics* 1951; 22:79–86
32. Vincent JL, Opal SM, Marshall JC, et al: Sepsis definitions: Time for change. *Lancet* 2013; 381:774–775
33. Sprung CL, Sakr Y, Vincent JL, et al: An evaluation of systemic inflammatory response syndrome signs in the sepsis occurrence in acutely ill patients (SOAP) study. *Intensive Care Med* 2006; 32:421–427
34. Kawasaki T: Update on pediatric sepsis: A review. *J Intensive Care* 2017; 5:47
35. Vincent JL, Moreno R, Takala J, et al: The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the european society of intensive care medicine. *Intensive Care Med* 1996; 22:707–710
36. Donnelly JP, Safford MM, Shapiro NI, et al: Application of the third international consensus definitions for sepsis (sepsis-3) classification: A retrospective population-based cohort study. *Lancet Infect Dis* 2017; 17:661–670
37. Beaulieu-Jones BK, Yuan W, Brat GA, et al: Machine learning for patient risk stratification: Standing on, or looking over, the shoulders of clinicians? *NPJ Digit Med* 2021; 4:62
38. Yan D, Huang L, Jordan MI: Fast approximate spectral clustering. *In:* Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28–July 1, 2009, pp 907–916
39. Mikkelsen ME, Miltiades AN, Gaieski DF, et al: Serum lactate is associated with mortality in severe sepsis independent of organ failure and shock. *Crit Care Med* 2009; 37:1670–1677
40. Trzeciak S, Dellinger RP, Chansky ME, et al: Serum lactate as a predictor of mortality in patients with infection. *Intensive Care Med* 2007; 33:970–977
41. Kaur J, Singhi P, Singhi S, et al: Neurodevelopmental and behavioral outcomes in children with sepsis-associated encephalopathy admitted to pediatric intensive care unit: A prospective case control study. *J Child Neurol* 2016; 31:683–690
42. Miao H, Shi J, Wang C, et al: Continuous renal replacement therapy in pediatric severe sepsis: A propensity score-matched prospective multicenter cohort study in the PICU. *Crit Care Med* 2019; 47:e806–e813
43. Klompas M, Winslow DL, Strich JR, Sweeney DA, et al: Infectious Diseases Society of America Position Paper: Recommended revisions to the national severe sepsis and septic shock early management bundle (SEP-1) sepsis quality measure. *Clinical Infectious Diseases* 2020; 72:541–552
44. Deutschman CS: Imprecise medicine: The limitations of sepsis-3. *Crit Care Med* 2016; 44:857–858
45. Vincent JL, Quintairos E Silva A, Couto L Jr, et al: The value of blood lactate kinetics in critically ill patients: A systematic review. *Crit Care* 2016; 20:257
46. Levy MM, Evans LE, Rhodes A: The surviving sepsis campaign bundle: 2018 uupdate. *Crit Care Med* 2018; 46:997–1000
47. Agniel D, Kohane IS, Weber GM: Biases in electronic health record data due to processes within the healthcare system: Retrospective observational study. *BMJ* 2018; 361:k1479