

## **Development and Validation of the Phoenix Criteria for Pediatric Sepsis and Septic Shock**

L. Nelson Sanchez-Pinto, MD, MBI\*

Department of Pediatrics (Critical Care) and Preventive Medicine (Health & Biomedical Informatics), Northwestern University Feinberg School of Medicine; Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA

Tellen D. Bennett, MD, MS\*

Departments of Biomedical Informatics and Pediatrics (Critical Care Medicine), University of Colorado School of Medicine; Children's Hospital Colorado, Aurora, CO, USA

Peter E. DeWitt, PhD\*\*

Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, CO, USA

Seth Russell, MS\*\*

Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, CO, USA

Margaret N. Rebull, MA

Department of Biomedical Informatics, University of Colorado School of Medicine; Aurora, CO, USA

Blake Martin, MD

Departments of Pediatrics (Critical Care Medicine) and Biomedical Informatics, University of Colorado School of Medicine; Children's Hospital Colorado, Aurora, CO, USA

Samuel Akech, MBChB, MMED

Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Nairobi, Kenya

David J. Albers, PhD

Departments of Biomedical Informatics, Bioengineering, Biostatistics and Informatics, University of Colorado School of Medicine; Aurora, CO, USA, Department of Biomedical Informatics, Columbia University, New York, NY, USA

Elizabeth R. Alpern, MD, MSCE

Division of Emergency Medicine, Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago and Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

Fran Balamuth MD, PhD, MSCE

Department of Pediatrics, University of Pennsylvania, Perelman School of Medicine and Division of Emergency Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

Melania Bembea, MD, MPH, PhD

Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Mohammad Jobayer Chisti, MBBS, MMed, PhD

Intensive Care Unit, Dhaka Hospital, Nutrition Research Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh

Idris Evans, MD, MSc

Department of Critical Care Medicine, University of Pittsburgh School of Medicine, and The Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Pittsburgh, Pennsylvania, USA

Christopher M. Horvat, MD, MHA

Department of Critical Care Medicine, University of Pittsburgh School of Medicine, and The Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Pittsburgh, Pennsylvania, USA

Juan Camilo Jaramillo-Bustamante, MD

Pediatric Intensive Care Unit, Hospital General de Medellín "Luz Castro de Gutiérrez" and Hospital Pablo Tobón Uribe, Medellín, Colombia  
Red Colaborativa Pediátrica de Latinoamérica (LARed Network)

Niranjan Kissoon, MD, MCCM

Department of Pediatrics, University of British Columbia, Vancouver, Canada

Kusum Menon, MD, MSc

Department of Pediatrics, Children's Hospital of Eastern Ontario and University of Ottawa, Ontario, Canada

Halden F. Scott, MD, MSCS

Department of Pediatrics (Pediatric Emergency Medicine), University of Colorado School of Medicine; Children's Hospital Colorado, Aurora, Colorado, USA

Scott L. Weiss, MD, FCCM

Division of Critical Care, Department of Pediatrics, Nemours Children's Health, Wilmington, Delaware and Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania, USA

Matthew O. Wiens, PharmD, PhD

Department of Anesthesiology, Pharmacology and Therapeutics, Faculty of Medicine, University of British Columbia, Vancouver, Canada; Institute for Global Health, BC Children's Hospital, Vancouver, Canada; Walimu, Uganda

Jerry J. Zimmerman, MD, PhD, MCCM  
Seattle Children's Hospital, Department of Pediatrics, University of Washington School of Medicine, Seattle, Washington, USA

Andrew C. Argent, MD, MBBCh, MMed\*\*\*  
Paediatrics and Child Health, University of Cape Town Faculty of Health Sciences, Cape Town, South Africa.

Lauren R. Sorce, PhD, RN, CPNP-AC/PC\*\*\*  
Department of Pediatrics, Northwestern University Feinberg School of Medicine; Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA

Luregn J. Schlapbach, MD, PhD\*\*\*  
Department of Intensive Care and Neonatology, and Children's Research Center, University Children's Hospital Zurich, University of Zurich, Zurich, Switzerland; and Child Health Research Centre, The University of Queensland, Brisbane, Australia

R Scott Watson, MD, MPH\*\*\*  
Department of Pediatrics, University of Washington; Seattle and Center for Child Health, Behavior, and Development and Pediatric Critical Care, Seattle Children's; Seattle, Washington, USA

*and the Society of Critical Care Medicine Pediatric Sepsis Definition Task Force (see Supplement)*

\*Contributed equally

\*\*Contributed equally

\*\*\*Contributed equally

**Word count:** 3794

**Date of revision:** December 31, 2023

**Key words:** child; children, criteria; critical illness; infection; informatics; organ dysfunction; sepsis; septic shock

**Correspondence:**

Tellen D. Bennett, MD, MS  
1890 N. Revere Court  
Mail Stop 600  
Aurora, CO 80045  
tellen.bennett@cuanschutz.edu  
Voice: 206-310-1374

## ABSTRACT

**Importance:** The Pediatric Sepsis Definition Task Force sought to develop and validate new clinical criteria for pediatric sepsis and septic shock using measures of organ dysfunction through a data-driven approach.

**Objective:** To derive and validate novel criteria for pediatric sepsis and septic shock across differently resourced settings.

**Design:** Multicenter, international, retrospective cohort study.

**Setting:** Ten health systems in the United States, Colombia, Bangladesh, China, and Kenya, three of which were used as external validation sites.

**Participants:** Emergency and inpatient encounters for children <18 years old from 2010-2019; 3,049,699 in the development (including derivation and internal validation) set and 581,317 in the external validation set.

**Exposure:** Stacked regression models to predict mortality in children with suspected infection were derived and validated using the best-performing organ dysfunction subscores from eight existing scores. The final model was then translated into an integer-based score used to establish binary criteria for sepsis and septic shock.

**Main Outcome and Measures:** The primary outcome for all analyses was in-hospital mortality. Model and integer-based score performance measures included area under the precision-recall curve (AUPRC, primary) and area under the receiver operator characteristic curve (AUROC, secondary). For binary criteria, primary performance measures were positive predictive value (PPV) and sensitivity.

**Results:** Among the 172,984 children with suspected infection in the first 24 hours (development set, 1.2% mortality), a 4-organ system model performed best. The integer version of that model – the Phoenix Sepsis Score– had AUPRCs of 0.23-0.38 (95% confidence intervals [CIs] range 0.20-0.39) and AUROCs of 0.71-0.92 (95% CIs 0.70-0.92) to predict mortality in the validation sets. Using a Phoenix Sepsis Score  $\geq 2$  points in children with suspected infection as criteria for sepsis and sepsis plus  $\geq 1$  cardiovascular point as criteria for septic shock resulted in a higher PPV and higher or similar sensitivity compared to the 2005 International Pediatric Sepsis Consensus Conference (IPSCC) criteria across differently resourced settings.

**Conclusions and Relevance:** The novel Phoenix sepsis criteria, which were derived and validated using data from higher and lower resource settings, had a higher sensitivity and PPV for the diagnosis of pediatric sepsis and septic shock than the existing IPSCC criteria.

**Abstract word count: 350**

## **KEY POINTS**

**Question:** What are the best-performing organ dysfunction-based criteria to implement the definition of sepsis and septic shock in children with suspected infection?

**Findings:** In this international, multicenter retrospective cohort study including over 3.6 million pediatric encounters, a novel score –the Phoenix Sepsis Score– was derived and validated to predict mortality in children with suspected or confirmed infection. The new criteria for pediatric sepsis and septic shock based on the score performed better than existing organ dysfunction scores and the International Pediatric Sepsis Consensus Conference criteria.

**Meaning:** The new data-driven criteria for pediatric sepsis and septic shock based on measures of organ dysfunction had improved performance compared with prior pediatric sepsis criteria.

## INTRODUCTION

Pediatric sepsis is a major public health problem that causes an estimated 3.3 million deaths annually worldwide.<sup>1</sup> However, the current criteria to diagnose pediatric sepsis, which were published in 2005 following the International Consensus Conference on Pediatric Sepsis, are outdated, have low specificity, do not allow for risk stratification in both lower and higher resource settings, and may be discordant with clinician-based diagnosis.<sup>2,3</sup> In 2016, the Sepsis-3 Task Force redefined adult sepsis as life-threatening organ dysfunction in the setting of infection and developed criteria using a large electronic health record (EHR) dataset and a data-driven approach.<sup>4,5</sup> In 2019, the Pediatric Sepsis Definition Task Force was convened to update the pediatric sepsis definition and criteria. The Task Force adopted the conceptual definition of pediatric sepsis as suspected infection with life-threatening organ dysfunction and sought to implement the definition using organ dysfunction criteria associated with higher risk of mortality. The goal was to develop criteria that would generalize across differently resourced settings [REF Consensus Criteria paper].

New pediatric sepsis criteria should maximize identification of true positive cases so that infected children with life-threatening organ dysfunction receive best practice sepsis care, are appropriately enrolled in clinical studies, and are correctly represented in epidemiological surveillance. Simultaneously, new criteria must minimize false positive cases so that infected children are not misdiagnosed with sepsis. This is important to reduce unnecessary antimicrobials and other treatments, optimize the efficiency of clinical studies, and avoid overcounting in surveillance. However, it is unclear which measures of organ dysfunction in children have an appropriate balance of sensitivity and positive predictive value (PPV) to achieve these goals and also generalize across differently resourced settings.

One challenge is that there is currently no large, centralized, multi-center, high-granularity database that includes pediatric emergency and inpatient care in differently resourced settings. Additionally, the

validation of the existing International Pediatric Sepsis Consensus Conference (IPSCC) criteria has been limited historically.<sup>2,3</sup> To address these gaps, a database was developed and used to derive and validate novel criteria for pediatric sepsis and septic shock based on measures of organ dysfunction in children with suspected infection.

## **METHODS**

### Overview

The existing organ dysfunction subscores for each organ system that best predicted mortality were first identified and then integrated into models to predict mortality in children with suspected infection. From the best-performing models, an integer-based score (the Phoenix Sepsis Score) was developed (eFigure 1). The binary (yes/no) Phoenix sepsis and septic shock criteria were then selected as thresholds of the Phoenix Sepsis Score.

### **Study Design, Setting, and Population**

A retrospective cohort study was performed using electronic health record (EHR) data from 10 hospital-based sites in 5 countries. The analysis plan was pre-specified in the funding application which supported this work. Six U.S. sites represent higher resource settings, 5 of which were in the development dataset (eFigure 2). Data from one U.S. site was held out for geographic external validation. Two international sites in Bangladesh and Colombia represent lower resource settings in the development dataset. Additionally, limited EHR and registry data from sites in China<sup>6</sup> and Kenya served as lower resource external validation sites. From each site, all emergency department (ED), inpatient, and intensive care unit (ICU) encounters of children <18 years old from 2010-2019 were included, with some sites providing shorter time windows (eTable 1). Data from newborns before discharge (birth hospitalizations) and children with post-conceptional age <37 weeks were excluded. Data harmonization, quality assurance, and all analyses were conducted as a

reproducible pipeline in a centralized, cloud-based environment (eFigure 2, eMethods). The study was approved with a waiver of consent and a central IRB at the University of Colorado, plus separate regulatory approvals at non-U.S. sites.

### **Outcomes, Definitions, and Main Measures**

The primary outcome for all analyses was in-hospital mortality, which was used to assess the likelihood that an organ dysfunction in the setting of an infection was life-threatening. The secondary outcome for all analyses was a composite of early death (within 72 hours of presentation to the hospital) or requirement of extracorporeal membrane oxygenation (ECMO) support. This secondary outcome was requested by the Task Force because early death and ECMO are more likely to be directly associated with sepsis in the first 24 hours of presentation than in-hospital mortality, which can occur later and be the result of complications during the hospitalization. Also, using ECMO to rescue children with sepsis-associated respiratory and/or cardiac failure could lead to survival of some children who would otherwise die. Suspected infection was defined in children who received systemic antimicrobials and had microbiological testing within the first 24 hours of the encounter. Comorbidities were defined based on the pediatric complex chronic conditions classification system,<sup>7</sup> and severe malnutrition was based on >3 standard deviations below the mean based on weight-for-age standards from the World Health Organization.<sup>8</sup> The systemic inflammatory response syndrome (SIRS) criteria were based on the IPSCC criteria.<sup>2,3</sup> Because dosing information necessary to calculate the vasoactive-inotropic score (VIS) was often missing at lower resource sites, the number of concurrent vasoactive agents was tested as a proxy. The area under the precision-recall curve (AUPRC) was used as the primary measure of organ dysfunction subscore, stacked regression sepsis model, and Phoenix Sepsis Score performance because it is more accurate than the area under the receiver-operator characteristic (AUROC) curve when analyzing imbalanced datasets (e.g., many more survivors than non-survivors). This is particularly important in children with infections given their lower baseline mortality compared to adults.<sup>9,10</sup>

The best way to interpret AUPRCs is to use the baseline rate as reference. If mortality is 1% (0.01) and the model AUPRC is 0.30, the model has 30-fold higher performance than a random model. Because the novel Phoenix sepsis and septic shock criteria represent single, binary thresholds, the primary performance measures used to evaluate them were sensitivity and PPV, which represent single points on the precision-recall curve. Missing data were imputed using a ‘last observation carried forward’ (LOCF) approach across physiologically appropriate time windows. Please see the eMethods for details.

### **Derivation and Validation of the Novel Criteria for Sepsis and Septic Shock**

The evaluation of which organ dysfunction subscores best predicted mortality involved all patients with and without suspected infection (eFigures 1-2). Then, stacked regression models<sup>11,12</sup> were derived and validated to predict mortality using the worst organ dysfunction subscores recorded in the first 24 hours of the encounter among children with suspected infection (eFigures 1-2). This approach was used to implement the concept of “an infection with life-threatening organ dysfunction”, which was adopted by the Pediatric Sepsis Task Force as the conceptual definition of sepsis.

The dataset was first divided into development (including derivation and internal validation) and external validation sets as described above and shown in eFigure 2. From each development site, 25% were held out for internal validation. The other three 25% portions of the development dataset were used to: (1) identify the best-performing criteria for each individual organ dysfunction based on the subscores of 8 existing and previously validated pediatric organ dysfunction criteria in all patients in the development datasets (including patients with suspected infection and those without) (eTable 2, eFigure 2),<sup>13-18</sup> (2) train and tune stacked regression models using a composite of the best performing individual organ dysfunction criteria in children with suspected

infection,<sup>11,12</sup> and (3) derive and internally validate the novel sepsis criteria based on the final stacked regression model. Finally, the novel criteria were validated in the external validation sets.

Stacked regression is a robust model-averaging approach that allows many models to be used simultaneously, leveraging the best predictive power of each model. The best-performing organ dysfunction subcomponent scores were used as input variables for stacked regression models that also predicted mortality. The stacked regression models took the organ dysfunction subscores as covariates and estimated the regression weights (or the relative contribution of each respective subcomponent's prediction to the overall prediction) in accordance with each subcomponent's predictive power, while maintaining a high degree of interpretability.<sup>12</sup> Additional information can be found in the eMethods.

Ridge, least absolute shrinkage and selection operator (LASSO), and elastic net regularized logistic regression were evaluated as the top-level stacked models. Ten-fold cross-validation was used to select the regularization parameter lambda in the stacked models that minimized deviance for each value of alpha (0 = ridge, 1 = LASSO). Please see the eMethods for additional information. The best-performing stacked regression models were identified using AUPRC. In the third step, the components of the final stacked regression model were translated into an integer-based score using a grid search and then its performance was compared to the final stacked model to ensure that the AUPRC remained stable. When measures and models had similar performance, the Task Force voted on which to choose based on parsimony, data collection burden, and face validity [Ref Consensus Criteria paper]. The Task Force then voted using a modified Delphi process on the thresholds of the score to define sepsis and septic shock and achieve the desired balance of sensitivity and PPV. In the final step, performance of the novel criteria was assessed across validation sets using sensitivity and PPV as primary metrics. Additional information is in the eMethods and eFigures 1-2.

## **Stratifications and Sensitivity Analyses**

During each step, pre-specified stratifications and sensitivity analyses were performed to ensure robustness. These included: (1) higher resource versus lower resource settings, where the higher resource sites were analyzed together given their overall similarity and lower resource sites were analyzed individually given their broader differences in underlying population, resources, and data quality; (2) no known prior comorbidities, to assess criteria performance in children without potential confounding by chronic and/or life-limiting conditions; (3) age groups, to ensure that performance remains appropriate across the pediatric spectrum; (4) ICU admission, given that many children with sepsis will receive ICU care; and (5) excluding patients who required operative care, to reduce confounding by mechanical ventilation or vasoactive medications related to receiving anesthesia or undergoing surgery.

## **RESULTS**

### **Cohort Demographic and Clinical Characteristics**

The development set included 3,049,699 ED, inpatient, and ICU encounters for children <18 years old, of which 172,984 (5.7%) had suspected infection in the first 24 hours (Table 1, eTables 3 and 4, eFigure 2). Of those, 2,065 (1.2%) died. The external validation set included 581,317 encounters, of which 45,855 (7.9%) had suspected infection in the first 24 hours. Of those, 540 (1.2%) died (Table 1 and eTable 5).

### **Best-Performing Individual Organ Dysfunction Criteria**

Organ dysfunction subscore input availability and missingness are shown in eFigures 3A-H. By 24 hours into an encounter, most patients in higher resource settings had SpO<sub>2</sub>, respiratory support, platelets, blood pressure, vasoactive agents, and Glasgow coma scale charted. Many also had FiO<sub>2</sub>,

lactate, and pupillary reactivity measured. Patients in lower resource settings were less likely to have available data on lactate, Glasgow Coma Scale, pupillary reactivity, and coagulation studies such as D-dimer and fibrinogen. The best-performing individual organ dysfunction criteria based on the primary measure of AUPRC and Task Force Delphi process when AUPRCs were similar included: cardiovascular (Pediatric Logistic Organ Dysfunction-2, PELOD-2, and vasoactive medication count), hematology/coagulation (Disseminated Intravascular Coagulation, DIC score), respiratory (pediatric Sequential Organ Failure Assessment, pSOFA), renal (pSOFA), hepatic (IPSCC), neurological (PELOD-2), immunologic (Pediatric Organ Dysfunction Information Mandate, PODIUM), and endocrine dysfunction (PODIUM), as shown in eFigure 4.

### **Derivation and Validation of the Stacked Models**

The best-performing stacked models included an 8-organ system ridge regression model and a 4-organ system LASSO model (eTable 6, eFigure 6). Overall, AUPRCs and AUROCs were similar between these 2 models (eFigure 7). The Task Force evaluated the two models and chose to advance the 4-organ system model because it had similar performance, but greater simplicity and lower dependence on laboratory measures. ([REF Consensus Criteria paper](#)). The Task Force acknowledged that the more comprehensive 8-organ system model may have utility in some circumstances (e.g., research). The 4-organ system model included criteria for respiratory (mechanical ventilation, PaO<sub>2</sub>/FiO<sub>2</sub>, and SpO<sub>2</sub>/FiO<sub>2</sub> ratios), cardiovascular (mean arterial pressure, lactate level, and vasoactive medications), coagulation (platelet count, INR, D-dimer, and fibrinogen), and neurologic dysfunction (Glasgow coma scale and pupillary reaction).

### **From the Stacked Model to the Phoenix Sepsis Score**

The 4-organ system model was translated into an integer-based score, the Phoenix Sepsis Score (Table 2). In doing so, the individual levels were re-weighted using a grid search and collapsed into a single level when performance was unaffected (e.g., the pSOFA respiratory subscores of 1 and 2 points were

collapsed into a single level). Mortality increased with higher score values in both higher and lower resource settings (Figure 1A and 1B and eFigure 5). The Phoenix Sepsis Score had AUPRCs of 0.23-0.38 (95% CIs range 0.20-0.39) and AUROCs of 0.71-0.92 (95% CIs range 0.70-0.92) to predict mortality in the internal and external validation sets, similar to the stacked sepsis model (Figure 2, eFigures 6-8). Compared to the existing IPSCC sepsis score as well as several organ dysfunction scores, the Phoenix Sepsis Score had the highest AUPRC to predict mortality at all validation sites combined, at all higher resource sites, and at three of the four lower resource sites (Figure 2). A notable limitation is that lower resource sites 2-4 did not record respiratory support, even when the patient received it, which limited the range of the score and likely resulted in lower performance at those sites. Additionally, lower resource site 2 had no recording of neurologic status, further limiting score range and performance at that site. However, the score at lower resource site 1 included data for all 4 organ systems. To enable capture of other organ dysfunctions for research or epidemiological purposes, an expanded score based on the 8-organ system model was also developed and named the 'Phoenix-8' score (eFigure 9).

### **From the Phoenix Sepsis Score to the Criteria for Pediatric Sepsis and Septic Shock**

The Task Force chose a Phoenix Sepsis Score  $\geq 2$  in patients with suspected infection as the new sepsis criteria, and sepsis with  $\geq 1$  cardiovascular point as criteria for septic shock. In the development set, children with sepsis in the first 24 hours had 7.1% mortality at the higher resource sites and 28.5% mortality at the lower resource sites. Children with sepsis in both higher and lower resource settings had a median Phoenix Sepsis Score of 3 points (interquartile range 2-4). Children with septic shock in the first 24 hours had 10.8% mortality at the higher resource sites and 33.5% mortality at the lower resource sites. The novel criteria had higher PPV and comparable or higher sensitivity than the IPSCC sepsis, severe sepsis, and septic shock criteria across all settings and using the secondary outcome of early death or ECMO (Figure 3, eFigure 10, and eTable 7). For example, for the primary outcome of death in the higher resource sites, the Phoenix sepsis criteria had a PPV of 5.3% to 7.1% (with a baseline mortality of 0.6 to 0.7%) and a sensitivity of 69.2% to 84.4% compared to the IPSCC severe sepsis criteria, which had a

PPV of 3.6% to 4.8% and a sensitivity of 58.7% to 70.7%, in the development and external validation sets, respectively. In the derivation and internal validation set of lower resource site 1, which had complete data for assessment of the criteria, the Phoenix sepsis criteria had a PPV of 22.2% (baseline mortality rate of 4.1%) and a sensitivity of 81.2% compared to the IPSCC severe sepsis criteria, which had a PPV of 12.7% and a sensitivity of 49.2%.

Per request of the Task Force, the concept of “organ dysfunction remote to the site of infection” was implemented by requiring that those with respiratory or neurologic dysfunction also had  $\geq 1$  point in a different organ system. Patients with sepsis who had remote organ dysfunction accounted for 85.2% of sepsis cases and had higher mortality than the whole sepsis cohort: 8% in the higher resource sites and 32.3% in lower resource sites (eFigure 11).

### **Sensitivity Analyses**

Performance of the pediatric sepsis criteria was consistent across age groups, with higher sepsis incidence and mortality in younger age groups, as expected (eTable 8). Similarly, the performance was consistent in patients with no known prior comorbidities, those admitted to the ICU, and after excluding patients who underwent surgery (eTable 8).

Clinical vignettes for children presenting with sepsis and septic shock and their corresponding Phoenix Sepsis Scores are provided in the eAppendix.

## **DISCUSSION**

New criteria for pediatric sepsis and septic shock were derived and validated by developing and curating a clinical database with >3.6 million pediatric hospital encounters at 10 sites in 5 countries. The development dataset was built using structured EHR data from an international cohort that was

geographically and racially diverse and had widely varying resources, a major strength of this study. A pre-specified data-driven approach was used to determine the best-performing organ dysfunction measures in children with suspected infection. An interpretable machine learning approach was used to develop a composite model that was the basis for the new Phoenix Sepsis Score and the new criteria. The new Phoenix criteria for pediatric sepsis and septic shock had higher PPV and comparable to or higher sensitivity than the IPSCC criteria for predicting mortality across differently resourced settings. These findings were consistent in multiple sensitivity analyses that included age, absence of prior comorbidities, ICU admission, and surgery.

#### Comparison **with the adult Sepsis-3 criteria**

The approach used in this study had both similarities with and differences from the derivation of the adult Sepsis-3 criteria.<sup>4</sup> Similar to Sepsis-3, the definition of sepsis was implemented as the combination of suspected infection with life-threatening organ dysfunction. Also, existing organ dysfunction scores and a large EHR database were used to develop the new criteria and in-hospital mortality was the primary outcome. However, there were also several important differences. First, instead of using existing complete organ dysfunction scores (e.g., the Sequential Organ Failure Assessment [SOFA] score) to derive the new criteria, the best-performing individual organ measures of existing scores were used to develop a novel composite score using stacked regression. Additionally, a database was built that included a geographically and demographically diverse population of children from both higher and lower resource settings to maximize generalizability. Furthermore, the performance of the individual organ dysfunction measures, the stacked models, and the Phoenix Sepsis Score was primarily evaluated using AUPRC, instead of AUROC, with the goal of maximizing the PPV and sensitivity of the final criteria. AUPRC is considered a better measure of classification performance for rare events (in this case, deaths) when compared to AUROC, which can have inflated performance when the proportions of events (deaths) and non-event (survivors) are imbalanced,<sup>10,19</sup> an issue that is particularly relevant in children

with infections given their lower mortality compared to adults. Finally, this analysis focused on diagnosis of sepsis within the first 24 hours of presentation to a hospital setting, when the majority of pediatric sepsis is diagnosed.<sup>20</sup>

### **Leveraging digital technology to develop and implement the Phoenix Sepsis Score**

This approach to the development of the Phoenix Sepsis Score and the criteria for sepsis and septic shock is a reflection of the growing digitization of healthcare around the world.<sup>21</sup> Most of the vital signs, laboratory tests, and interventions that are included in the Phoenix Sepsis Score are routinely collected in most lower resource settings and nearly all higher resource settings, according to the international survey conducted by the Pediatric Sepsis Definition Task Force.<sup>22</sup> Even in settings where not all variables are available, the Phoenix Sepsis Score was designed to accurately identify children with sepsis. The Score functions when not all variables are available because of its redundancy. Because the score has a possible range of 0 to 13 points, there are several ways to achieve the threshold of 2 points for sepsis diagnosis, as evidenced by the fact that patients with sepsis in both higher and lower resourced settings had a median Phoenix Sepsis Score of 3 points. This feature was primarily assessed in the datasets from lower resource settings. For example, although platelets were commonly measured at most sites, coagulation tests (e.g., D-Dimer and fibrinogen) were less frequently available. At lower resource site 1, where platelet count was routinely measured but coagulation factors such as D-Dimer and fibrinogen were not, the Phoenix Sepsis Score had excellent performance and the Phoenix sepsis criteria had higher sensitivity and PPV than the IPSCC sepsis and severe sepsis criteria. This makes the score and criteria readily translatable into EHR and other digital tools, such as web-based and mobile applications across differently resourced settings, even when some of the variables are not routinely collected.<sup>23</sup> Furthermore, digital implementation of the Phoenix Sepsis Score can enable longitudinal monitoring and provide clinicians and researchers with a tool to stratify severity of sepsis.

Additional considerations for the implementation and use of the Phoenix Sepsis Score and the novel criteria are discussed in the Special Communication article in this issue [[Ref Consensus Criteria paper](#)]

## **Limitations**

This study has several limitations. Retrospective data obtained from EHRs may have missing data and data entry errors. In this study, a robust quality assurance and harmonization process was developed and best practices were used to address outliers and missing data. However, not all errors or missing data can be reconciled. For example, at lower resource site 2 in the development dataset, which represents a lower-middle income country, respiratory support (e.g., mechanical ventilation, FiO<sub>2</sub>) and neurologic assessments (e.g., level of consciousness and pupillary reaction) are performed but not recorded in their clinical information systems. This reduces the ability to assess the score and criteria at that site. In contrast, score performance was excellent at lower resource site 1 and comparable to the higher resource sites. This demonstrates the potential for score performance in lower resource environments when these variables are recorded. Second, when deriving the stacked regression models, the Phoenix Sepsis Score, and the new criteria for sepsis and septic shock, a pragmatic approach was intentionally chosen, using the data as recorded during routine care as an indicator of how the criteria would perform in real-world implementations. However, it is acknowledged that some of the organ dysfunction measures used in the modeling process may not have reflected actual organ dysfunction, but rather were due to iatrogenic effects or clinician therapeutic choices, such as a lower GCS score in a patient receiving sedation or initiation of vasoactive medications in a patient with minimal cardiovascular dysfunction. Future work to determine the effects of these variables and clinician choices on the performance of the criteria is needed. Third, similar to the Sepsis-3 validation study, unique criteria for patients with chronic organ dysfunction were not developed.<sup>4</sup> Fourth, few databases from lower resource settings were available (a form of data poverty),<sup>24</sup> and the ones used may not be generalizable to every low resource environment. Fifth, the data from higher resource settings were exclusively from tertiary U.S. pediatric centers. Sixth, the datasets

from some of the sites included 10 years of data, possibly including changes in practice over that timeframe.

## **CONCLUSIONS**

The novel Phoenix sepsis criteria, which were derived and validated using a large international database of pediatric hospital encounters in higher and lower resource settings, had a higher sensitivity and positive predictive value for the diagnosis of pediatric sepsis and septic shock than existing International Pediatric Sepsis Consensus Conference Criteria.

### **Acknowledgements:**

The authors would like to thank the members of the Society of Critical Care Medicine Pediatric Sepsis Definition Task Force for their contributions to this work. In addition, Kathy Vermoch, Lori Harmon, and Lynn Retford at the Society of Critical Care Medicine provided invaluable assistance throughout this project.

### **Author contributions:**

*Concept and design:* The plan for this work was developed by TDB, LNS-P, and DJA. The project plan for the Pediatric Sepsis Definition Task Force was drafted by LJS, RSW, LRS, AA, NK, and JZ.

*Acquisition, analysis, or interpretation of data:* TDB and LNS-P led data acquisition and analysis including the building of the harmonized international database used to develop and validate the new criteria. FB, MB, TDB, MJC, IE, CH, JCJ-B, LNS-P, RSW, and SW curated data at contributing sites, performed data quality checks, and contributed to data harmonization. TDB and LNS-P led a team including DJA, PED, BM, MNR, and SR who conducted the harmonization and analysis of the data, with clinical and scientific contributions by RSW, LJS, HS, SW, FB, ERA, and KM. All Task Force members contributed to the interpretation of the data.

*Drafting of the manuscript:* TDB and LNS-P wrote the first draft of the manuscript.

*Critical review of the manuscript for important intellectual content:* All authors provided revisions to the manuscript and reviewed and approved the final version.

*Obtained Funding:* TDB and LNS-P obtained NIH funding for this work. JZ, NK, LJS, and RSW obtained funding for the Task Force through the Society of Critical Care Medicine.

*Administrative, technical, or material support:* LJS, RSW, AA, LRS served as co-Chairs and co-Vice Chairs of the Task Force and together with SCCM staff were responsible for the organization of the Task Force work. TDB, LNS-P, and MNR organized the work of the analytic team.

**Conflict of Interest Statement:**

Lauren R. Sorce is an elected member of the Executive Committee and serves as President-elect of the Society of Critical Care Medicine (SCCM) for 2023-2024 and President for 2024-2025. The research presented is her own work and does not represent SCCM. The rest of the authors declare no relevant conflicts of interest related to this work.

**Funding sources:**

This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) grant number R01HD105939 to TDB and LNS-P. LJS received support from the NOMIS foundation. The Society of Critical Care Medicine provided support to the Pediatric Sepsis Definition Task Force for travel of members, coordination of meetings, and other logistical support.

**Role of the Funders:**

The funding sources 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:** The data that support the findings of this study are not publicly available because the study sites retain data ownership. The data contain information that could compromise the privacy of research participants. We will share analytic code in free, publicly available repositories hosted by the gold-standard code-sharing site, GitHub (github.com).

**Pediatric Sepsis Definition Task Force Group Information:** See Supplement

## REFERENCES

1. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *Lancet*. 2020;395(10219):200-211.
2. Goldstein B, Giroir B, Randolph A, International Consensus Conference on Pediatric Sepsis. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med*. 2005;6(1):2-8.
3. Weiss SL, Fitzgerald JC, Maffei FA, et al. Discordant identification of pediatric severe sepsis by research and clinical definitions in the SPROUT international point prevalence study. *Crit Care*. 2015;19(1):325.
4. 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(8):762-774.
5. 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.
6. Zeng X, Yu G, Lu Y, et al. PIC, a paediatric-specific intensive care database. *Sci Data*. 2020;7(1):14.
7. 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.
8. World Health Organization. Nutrition for Health and Development. *WHO Child Growth Standards: Growth Velocity Based on Weight, Length and Head Circumference : Methods and Development*. World Health Organization; 2009.

9. Ozenne B, Subtil F, Maucourt-Boulch D. The precision--recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol*. 2015;68(8):855-859.
10. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS One*. 2015;10(3):e0118432.
11. Wolpert DH. Stacked generalization. *Neural Netw*. 1992;5(2):241-259.
12. Breiman L. Stacked regressions. *Mach Learn*. 1996;24(1):49-64.
13. Proulx F, Fayon M, Farrell CA, Lacroix J, Gauthier M. Epidemiology of sepsis and multiple organ dysfunction syndrome in children. *Chest*. 1996;109(4):1033-1037.
14. 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.
15. Leteurtre S, Duhamel A, Salleron J, et al. PELOD-2: an update of the PEdiatric logistic organ dysfunction score. *Crit Care Med*. 2013;41(7):1761-1773.
16. Rousseaux J, Grandbastien B, Dorkenoo A, Lampin ME, Leteurtre S, Leclerc F. Prognostic value of shock index in children with septic shock. *Pediatr Emerg Care*. 2013;29(10):1055-1059.
17. Khemani RG, Bart RD, Alonzo TA, Hatzakis G, Hallam D, Newth CJL. Disseminated intravascular coagulation score is associated with mortality for children with shock. *Intensive Care Med*. 2009;35(2):327-333.
18. Haque A, Siddiqui NR, Munir O, Saleem S, Mian A. Association between vasoactive-inotropic score and mortality in pediatric septic shock. *Indian Pediatr*. 2015;52(4):311-313.
19. Tharwat A. Classification assessment methods. *Applied Computing and Informatics*. 2020;17(1):168-192.
20. Scott HF, Brill R, Paul R, et al. Evaluating Pediatric Sepsis Definitions Designed for Electronic Health Record Extraction and Multicenter Quality Improvement. *Crit Care Med*. 2020;48(10):e916-e926.
21. Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA. Big data in global health: improving health in low- and middle-income countries. *Bull World Health Organ*. 2015;93(3):203-208.
22. Morin L, Hall M, de Souza D, et al. The Current and Future State of Pediatric Sepsis Definitions: An International Survey. *Pediatrics*. 2022;149(6). doi:10.1542/peds.2021-052565
23. Jimenez-Zambrano A, Ritger C, Rebull M, et al. Clinical decision support tools for paediatric sepsis in resource-poor settings: an international qualitative study. *BMJ Open*. 2023;13(10):e074458.
24. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit Health*. 2021;3(4):e260-e265.

### **Figure 1A. In-hospital mortality associated with the Phoenix Sepsis Score in patients at higher resource settings with suspected infection in the first 24 hours**

### **Figure 1B. In-hospital mortality associated with the Phoenix Sepsis Score in patients at lower resource settings with suspected infection in the first 24 hours**

Figure 1 shows the calibration of the Phoenix Sepsis Score in higher resource settings (sites with more technological resources, e.g. laboratory equipment, ventilators, and renal replacement therapy devices, to support organ dysfunction, panel A) and lower resource settings (sites with fewer technological resources to support organ dysfunction, panel B). For patients with suspected infection who have each possible integer value (lower x-axis) of the Phoenix Sepsis Score in the first 24 hours of the encounter, the y-axis shows mortality among those at the development (red), and internal validation sites (green) and the external validation sites (blue). Binomial confidence intervals for the mortality point estimate in each group are also shown. The middle of each panel shows cumulative mortality across Phoenix Sepsis Score categories. The number of encounters “at risk” and mortality counts in each group are shown across the bottom of that plot. At lower resource sites, some variables were rarely available (e.g., D-dimer and fibrinogen for coagulation dysfunction), even when other variables for the same organ systems were recorded (e.g., platelet count and INR), thus the maximum cumulative score achieved at lower resource sites was 9, instead of the maximum possible of 13.

### **Figure 2. Mortality prediction performance of the Phoenix Sepsis Score and organ dysfunction scores**

Figure 2 shows the performance of the Phoenix Sepsis Score (across the entire range from 0 to 13 points) across sites and in comparison to validated pediatric organ dysfunction scores and criteria to predict mortality in patients with suspected infection in the first 24 hours. Equivalent performance metrics for the secondary outcome, early death or ECMO, are shown in eFigure 7. All organ dysfunctions are evaluated across their respective full ranges, with higher scores indicating more organ dysfunction burden. The scores for IPSCC, Proulx, and PODIUM are based on the count of organ dysfunctions. More information about these scores is provided in the Methods and eTable 2. **Bolded** values indicate the best-performing score for the respective dataset and performance measure. The performance is presented as both quantitative AUPRC (top) and AUROC (bottom), with 95% confidence intervals (calculated using Logit transform and shown below each point estimate of performance), as well as visually using a color heatmap. Shading indicates highest (darkest) to lowest (lightest) in each row, blue for AUPRC and yellow for AUROC. AUPRC is the area under a curve drawn with sensitivity (also referred to as “recall”) and positive predictive value (also referred to as “precision”), across all potential thresholds for the points in the scores. AUPRC is a more reliable classifier performance metric than AUROC when the classes are imbalanced, for example when mortality is very low as in this study. AUROC is the area under a curve drawn with false positive rate on the x-axis and true positive rate on y-axis, again across all potential thresholds for the points in the scores. In this study, it is an indicator of how well a classifier can rank encounters with respect to mortality risk. *IPSCC*, International Pediatric Sepsis Consensus Conference, *PELOD-2*, Pediatric Logistic Organ Dysfunction, version 2; *pSOFA*, pediatric Sequential Organ Failure Assessment; *PODIUM*, Pediatric Organ Dysfunction Information Update Mandate.

### **Figure 3. Comparison of the sensitivity and positive predictive value of the novel Phoenix sepsis criteria with the current IPSCC sepsis and severe sepsis criteria across outcomes and patient subgroups in the internal validation sets**

Figure 3 shows the positive predictive value (PPV, or precision) and sensitivity for the Phoenix Sepsis Criteria compared to the 2005 International Pediatric Sepsis Consensus Conference (IPSCC) criteria for sepsis in children with suspected infection. The Phoenix Sepsis Criteria was based on achieving  $\geq 2$  points in the Phoenix Sepsis Score among patients with suspected infection in the first 24 hours of an encounter. The IPSCC sepsis and severe sepsis criteria were based on the systemic inflammatory response syndrome (SIRS) and IPSCC-based organ dysfunction among patients with suspected infection in the first 24 hours of an

encounter. The baseline rate of the outcome in each group (death or early death or extracorporeal membrane oxygenation [ECMO]) is shown as a horizontal dashed red line. Confidence intervals for each component (sensitivity, PPV) are shown as bands from each point in the plane representing that component (e.g., confidence intervals for PPV are parallel to the y-axis). When a confidence band is not visible, that means that it is narrow enough to be completely hidden by the point. These figures are similar to AUPRCs except at a single threshold for criteria that generate a binary response (e.g., yes/no sepsis criteria met) instead of across the entire range of possible points in the curve (e.g., 0-13 points of the Phoenix Sepsis Score, which is shown in Figure 2). Better performing criteria on these figures will be closer to the top right corner of the figure. A tradeoff exists between sensitivity and PPV for the different outcomes, with more sensitive criteria usually having lower PPV, and more specific criteria usually having higher PPV and lower sensitivity. Criteria that are close to the baseline outcome rate (horizontal dashed red line) have poor predictive value. This comparison is stratified by higher resource setting (HRS) sites 1-5 (the held-out 25% internal validation sets) with encounter mortality (A) and death in the first 72 hours or use of extracorporeal membrane oxygenation (ECMO) (B) as the outcomes (or prediction targets). Panel C and D shows the same comparisons for children in higher resource settings who have no known comorbidity and encounters including ICU stays, respectively. Panels E and F show the same comparison at lower resource setting (LRS) sites 1-2. \*At LRS sites 2 to 4, some of the Phoenix Sepsis Score and IPSCC data inputs (e.g., invasive mechanical ventilation, Glasgow Coma Scale score) are not recorded, even when they are performed, thus the assessment of the criteria performance at those sites is limited. LRS sites 1 and all HRS sites have inputs for all relevant organ systems in the criteria. The comparison of the sepsis criteria in the external validation sites is presented in eFigure 10 and shows similar results. The diagnostic performance measures for this comparison can also be found in eTable 7.

**Table 1. Cohort characteristics of encounters with suspected infection in the first 24 hours in the development dataset**

	<b>Derivation Cohort</b>	<b>Internal Validation Cohort</b>	<b>External Validation Cohort</b>
<b>Encounters, No.</b>	129,584	43,400	45,855
<b>Resource Setting, No. (%)</b>			
<b>Higher Resource Settings</b>	108,177 (83.5)	36,202 (83.4)	33,020 (72.0)
<b>Lower Resource Settings</b>	21,407 (16.5)	7,198 (16.6)	12,835 (28.0)
<b>Age in years, median (IQR)</b>	3.7 (0.9, 9.4)	3.7 (0.9, 9.3)	2.6 (0.6, 7.6)
<b>Sex, No. (%)</b>			
<b>Female</b>	62,868 (48.5)	21,041 (48.5)	22,295 (48.6)
<b>Male</b>	66,712 (51.5)	22,357 (51.5)	21,555 (47.0)
<b>Resource Setting, No. (%)</b>			
<b>Higher Resource Settings</b>	108,177 (83.5)	36,202 (83.4)	
<b>Lower Resource Settings</b>	21,407 (16.5)	7,198 (16.6)	
<b>Race, No. (%)*</b>			
<b>American Indian or Alaskan Native</b>	109 (0.1)	21 (0.0)	59 (0.1)
<b>Asian</b>	5,149 (4.0)	1,703 (3.9)	506 (1.1)
<b>Black</b>	22,709 (17.5)	7,512 (17.3)	7,476 (16.3)
<b>Multiple Races</b>	22,113 (17.1)	7,343 (16.9)	277 (0.6)
<b>Native Hawaiian or Other Pacific Islander</b>	105 (0.1)	31 (0.1)	70 (0.2)
<b>Other/Unknown</b>	22,095 (17.1)	7,309 (16.8)	14,051 (30.6)
<b>White</b>	57,518 (44.4)	19,533 (45.0)	23,545 (51.3)
<b>Ethnicity, No. (%)</b>			
<b>Hispanic or Latino</b>	33,698 (26.0)	11,457 (26.4)	55 (0.1)
<b>Age in years, median (IQR)</b>	3.74 (0.92, 9.38)	3.73 (0.93, 9.31)	
<b>Major comorbidity, No. (%)</b>			
<b>Technology Dependence</b>	18,951 (17.5)	6,011 (16.6)	5,677 (17.2)
<b>Severe malnutrition</b>	13,505 (10.4)	4,478 (10.3)	3,417 (7.5)
<b>Malignancy</b>	10,924 (10.1)	3,709 (10.2)	2,950 (8.9)

<b>Transplantation</b>	3,689 (3.4)	1,287 (3.6)	1,573 (4.8)
<b>Comorbidities per Pediatric Complex Chronic Condition (PCCC), No. (%)#</b>			
<b>No known prior comorbidity</b>	72,291 (66.8)	24,470 (67.6)	22,553 (68.3)
<b>1 PCCC</b>	9,406 (8.7)	3,150 (8.7)	2,580 (7.8)
<b>2 or more PCCCs</b>	26,480 (24.5)	8,582 (23.7)	7,887 (23.9)
<b>SIRS, No. (%)^</b>	56,711 (43.8)	18,848 (43.4)	21,436 (46.7)
<b>Locations visited during encounter, (not mutually exclusive), No. (%)</b>			
<b>Presented through the ED</b>	92,507 (71.6)	31,092 (71.9)	26,940 (61.6)
<b>Had 1 or more ICU stays</b>	23,128 (17.9)	7,840 (18.1)	10,702 (23.4)
<b>Had 1 or more OR visit(s)</b>	17,604 (13.6)	6,098 (14.1)	469 (1.1)
<b>Outcomes, No. (%)</b>			
<b>Death</b>	1,538 (1.2)	527 (1.2)	540 (1.2)
<b>Early Death or ECMO</b>	834 (0.6)	305 (0.7)	349 (0.8)

Table 1 shows site, demographic, care location, comorbidity, and outcome characteristics of those with suspected or confirmed infection in the first 24 hours of the encounter at the 7 development sites, stratified by the 75% derivation cohort versus the 25% internal validation cohort. *HRS*, higher resource settings; *LRS*, lower resource setting; *IQR*, interquartile range; For race categories\*, “Multiple Races” indicates that in the EHR data, the patient’s race was recorded as “multi-racial,” “multiple,” or “two or more races.” “Unknown/Other” indicates that the patient’s race was recorded in the EHR data as “other,” “unknown,” “not specified,” “information not recorded,” “patient declined,” “patient refused,” “refused,” or as a race category unique to a particular international country or region. For ethnicity categories#, PCCC is a system to classify pediatric chronic diseases using International Classification of Diseases (ICD) diagnosis and procedure codes and was only assessed in the higher resource sites, where the information was available (percentages for PCCC-related counts are based on higher resource setting encounters).<sup>7</sup> The major comorbidities of technology dependence (e.g. requiring a gastrostomy, a tracheostomy, a central line, etc.), malignancy, and transplantation were defined in the PCCC system. Severe malnutrition was defined as based on <3 standard deviations below the mean based on weight-for-age standards from the World Health Organization and assessed in all sites.<sup>8</sup> Early Death is defined as death in <72 hours from the beginning of the encounter. Systemic inflammatory response syndrome (SIRS)^ is calculated using temperature, white blood cell count, heart rate, and respiratory rate, with higher values reflecting more inflammation. SIRS criteria are met when two or more values are above the threshold for age, including at least temperature or white blood cell count. See Supplemental Methods for additional details. *IQR*, interquartile range; *ED*, emergency department; *OR*, operating room; *ICU*, intensive care unit; *ECMO*, extracorporeal membrane oxygenation.

**Table 2. The Phoenix Sepsis Score**

	<b>0 points</b>	<b>1 point</b>	<b>2 points</b>	<b>3 points</b>
<b>Respiratory</b> (0-3 points)	P/F $\geq$ 400 or S/F $\geq$ 292 <sup>1</sup>	P/F <400 on any respiratory support <sup>2</sup> or S/F <292 on any respiratory support <sup>2</sup>	P/F 100-200 and IMV or S/F 148-220 and IMV	P/F <100 and IMV or S/F <148 and IMV
<b>Cardiovascular</b> (0-6 points)  Age-based <sup>3</sup>  <1 month 1 to 11 months 1 to <2 years 2 to <5 years 5 to <12 years 12 to 17 years	<ul style="list-style-type: none"> <li>● No vasoactive medications<sup>4</sup></li> <li>● Lactate<sup>5</sup> &lt;5 mmol/L</li> <li>● MAP<sup>6</sup> (mmHg) &gt;30 &gt;38 &gt;43 &gt;44 &gt;48 &gt;51</li> </ul>	<u>1 point each (up to 3) for:</u> <ul style="list-style-type: none"> <li>● 1 vasoactive medication</li> <li>● Lactate 5-10.9 mmol/L</li> <li>● MAP (mmHg) 17-30 25-38 31-43 32-44 36-48 38-51</li> </ul>	<u>2 points each (up to 6) for:</u> <ul style="list-style-type: none"> <li>● <math>\geq</math>2 vasoactive medications</li> <li>● Lactate <math>\geq</math>11 mmol/L</li> <li>● MAP (mmHg) &lt;17 &lt;25 &lt;31 &lt;32 &lt;36 &lt;38</li> </ul>	
<b>Coagulation</b> <sup>7</sup> (0-2 points)	<ul style="list-style-type: none"> <li>● Platelets <math>\geq</math>100 K/<math>\mu</math>L</li> <li>● INR <math>\leq</math>1.3</li> <li>● D-Dimer <math>\leq</math>2 mg/L FEU</li> <li>● Fibrinogen <math>\geq</math>100 mg/dL</li> </ul>	<u>1 point each (max. 2 points) for:</u> <ul style="list-style-type: none"> <li>● Platelets &lt;100 K/<math>\mu</math>L</li> <li>● INR &gt;1.3</li> <li>● D-Dimer &gt;2 mg/L FEU</li> <li>● Fibrinogen &lt;100 mg/dL</li> </ul>		
<b>Neurologic</b> <sup>8</sup> (0-2 points)	<ul style="list-style-type: none"> <li>● GCS<sup>9</sup> &gt;10</li> <li>● Pupils reactive</li> </ul>	GCS $\leq$ 10	Fixed pupils bilaterally	

P/F, PaO<sub>2</sub>/FiO<sub>2</sub> ratio; S/F, SpO<sub>2</sub>/FiO<sub>2</sub> ratio (only SpO<sub>2</sub> of 97% or less); IMV, invasive mechanical ventilation; MAP, mean arterial pressure; INR, international normalized ratio of prothrombin time; GCS, Glasgow coma scale score.

**Notes for use:** The score may be calculated in the absence of some variables (e.g., even if lactate level is not measured and vasoactive medications are not used, a cardiovascular score can still be ascertained using blood pressure). It is expected that laboratory tests and other measurements will be obtained at the discretion of the medical team based on clinical judgment. Unmeasured variables contribute no points to the score.

<sup>1</sup>S/F ratio is only calculated if SpO<sub>2</sub> is 97% or less.

<sup>2</sup>The respiratory dysfunction of 1 point can be assessed in any patient on oxygen, high flow, non-invasive positive pressure, or IMV respiratory support, and includes P/F <200 and S/F <220 in children who are not on IMV.

<sup>3</sup>Ages are not adjusted for prematurity, and the criteria do not apply to birth hospitalizations, children with post-conceptional age <37 weeks, or those 18 years of age or older.

<sup>4</sup>Vasoactive medications include any dose of epinephrine, norepinephrine, dopamine, dobutamine, milrinone, and/or vasopressin (for shock).

<sup>5</sup>Lactate can be arterial or venous. Lactate reference range is 0.5-2.2 mmol/L.

<sup>6</sup>Use measured MAP preferentially (invasive arterial if available or non-invasive oscillometric), and if measured MAP is not available, a calculated MAP ( $1/3 \times \text{systolic} + 2/3 \times \text{diastolic}$ ) may be used as an alternative.

<sup>7</sup>The coagulation variables reference ranges are: platelets 150-450 K/ $\mu$ L; D-Dimer <0.5 mg/L FEU; Fibrinogen 180-410 mg/dL. The INR reference range is based on the local reference prothrombin time.

<sup>8</sup>The neurologic dysfunction subscore was pragmatically validated in both sedated and non-sedated patients, and those on and off IMV support.

<sup>9</sup>The GCS measures level of consciousness based on verbal, eye, and motor response and ranges from 3 to 15, with a higher score indicating better neurological function.