JAMA | **Original Investigation**

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

L. Nelson Sanchez-Pinto, MD, MBI; Tellen D. Bennett, MD, MS; Peter E. DeWitt, PhD; Seth Russell, MS; Margaret N. Rebull, MA; Blake Martin, MD; Samuel Akech, MBChB, MMED; David J. Albers, PhD; Elizabeth R. Alpern, MD, MSCE; Fran Balamuth, MD, PhD, MSCE; Melania Bembea, MD, MPH, PhD; Mohammod Jobayer Chisti, MBBS, MMed, PhD; Idris Evans, MD, MSc; Christopher M. Horvat, MD, MHA; Juan Camilo Jaramillo-Bustamante, MD; Niranjan Kissoon, MD; Kusum Menon, MD, MSc; Halden F. Scott, MD, MSCS; Scott L. Weiss, MD; Matthew O. Wiens, PharmD, PhD; Jerry J. Zimmerman, MD, PhD; Andrew C. Argent, MD, MBBCh, MMed; Lauren R. Sorce, PhD, RN, CPNP-AC/PC; Luregn J. Schlapbach, MD, PhD; R. Scott Watson, MD, MPH; and the Society of Critical Care Medicine Pediatric Sepsis Definition Task Force

IMPORTANCE The Society of Critical Care Medicine 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, SETTING, AND PARTICIPANTS Multicenter, international, retrospective cohort study in 10 health systems in the US, Colombia, Bangladesh, China, and Kenya, 3 of which were used as external validation sites. Data were collected from emergency and inpatient encounters for children (aged <18 years) from 2010 to 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 8 existing scores. The final model was then translated into an integer-based score used to establish binary criteria for sepsis and septic shock.

MAIN OUTCOMES AND MEASURES The primary outcome for all analyses was in-hospital mortality. Model- and integer-based score performance measures included the area under the precision recall curve (AUPRC; primary) and area under the receiver operating characteristic curve (AUROC; secondary). For binary criteria, primary performance measures were positive predictive value 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 to 0.38 (95% CI range, 0.20-0.39) and AUROCs of 0.71 to 0.92 (95% CI range, 0.70-0.92) to predict mortality in the validation sets. Using a Phoenix Sepsis Score of 2 points or higher in children with suspected infection as criteria for sepsis and sepsis plus 1 or more cardiovascular point as criteria for septic shock resulted in a higher positive predictive value and higher or similar sensitivity compared with 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 improved performance for the diagnosis of pediatric sepsis and septic shock compared with the existing IPSCC criteria.

- ➕ Editorial
- ➕ Related article
- ➕ Supplemental content
- ➕ CME Quiz at jamacmelookup.com

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

Group Information: The members of the Society of Critical Care Medicine Pediatric Sepsis Definition Task Force appear at the end of this article.

Corresponding Author: Tellen D. Bennett, MD, MS, Departments of Biomedical Informatics and Pediatrics (Critical Care Medicine), Children's Hospital Colorado, University of Colorado School of Medicine, 1890 N Revere Ct, Mail Stop 600, Aurora, CO 80045 (tell.bennett@cuanschutz. edu).

P ediatric sepsis is a major public health problem that causes an estimated 3.3 million deaths annually worldwide.[1] However, the current criteria to diagnose pediatric sepsis, which were published in 2005 following the International Pediatric Sepsis Consensus Conference (IPSCC), 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.[2,3] 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) data set and a data-driven approach.[4,5] In 2019, the Society of Critical Care Medicine 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.[6]

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 children are not misdiagnosed with sepsis. This is important to reduce unnecessary use of 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, multicenter, high-granularity database that includes pediatric emergency and inpatient care in differently resourced settings. Additionally, the validation of the existing IPSCC criteria has been limited historically.[2,3] 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 in Supplement 1). The binary 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 EHR data from 10 hospital-based sites in 5 countries. The analysis plan

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 more than 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.

was prespecified in the funding application that supported this work. Six US sites represent higher-resource settings, 5 of which were in the development data set (eFigure 2 in Supplement 1). Data from 1 US site was held out for geographic external validation. Two international sites in Bangladesh and Colombia represent lower-resource settings in the development data set. Additionally, limited EHR and registry data from sites in China[7] and Kenya served as lower-resource external validation sites. From each site, all emergency department, inpatient, and intensive care unit (ICU) encounters of children younger than 18 years from 2010-2019 were included, with some sites providing shorter time windows (eTable 1 in Supplement 1). Data from newborns before discharge (birth hospitalizations) and children with a postconceptional age of less than 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 and eAppendix 1 in Supplement 1). The study was approved with a waiver of consent by a central institutional review board at the University of Colorado, plus separate regulatory approvals at non-US sites.

Outcomes, Definitions, and Main Measures

The primary outcome for all analyses was in-hospital mortality, which was used to assess the likelihood that 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 as receipt of systemic antimicrobials and microbiological testing within the first 24 hours of the encounter. Comorbidities were defined based on the Pediatric Complex Chronic Conditions Classification System,[8] and severe malnutrition was

based on more than 3 SDs below the mean based on weight-for-age standards from the World Health Organization.[9] The systemic inflammatory response syndrome criteria were based on IPSCC criteria.[2,3] Because dosing information necessary to calculate the vasoactive-inotropic score 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 operating characteristic (AUROC) curve when analyzing imbalanced data sets (eg, many more survivors than nonsurvivors). This is particularly important in children with infections given their lower baseline mortality compared with adults.[10,11] 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 approach across physiologically appropriate time windows. See eAppendix 1 in Supplement 1 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 in Supplement 1). Then, stacked regression models[12,13] 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 in Supplement 1). This approach was used to implement the concept of "an infection with life-threatening organ dysfunction," which was adopted by the Pediatric Sepsis Definition Task Force as the conceptual definition of sepsis.

The data set was first divided into development (including derivation and internal validation) and external validation sets as described above and shown in eFigure 2 in Supplement 1. From each development site, 25% were held out for internal validation. The other three 25% portions of the development data set 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 data sets (including patients with suspected infection and those without) (eTable 2 and eFigure 2 in Supplement 1)[14-19]; (2) train and tune stacked regression models using a composite of the best-performing individual organ dysfunction criteria in children with suspected infection[12,13]; 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, leverag-

ing 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.[13] Additional information is available in eAppendix 1 in Supplement 1.

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) (see eAppendix 1 in Supplement 1 for additional information). The best-performing stacked regression models were identified using the AUPRC. In the third step, the components of the final stacked regression model were translated into an integer-based score using a grid search, then its performance was compared with 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.[6] 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 available in eAppendix 1 and eFigures 1-2 in Supplement 1.

Stratifications and Sensitivity Analyses

During each step, prespecified stratifications and sensitivity analyses were performed to ensure robustness. These included (1) higher-resource vs lower-resource settings, where the higher-resource sites were analyzed together given their overall similarity and the 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 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 emergency department, inpatient, and ICU encounters for children younger than 18 years, of which 172 984 (5.7%) had suspected infection in the first 24 hours (Table 1; eTables 3 and 4 and eFigure 2 in Supplement 1). Of those, 2065 (1.2%) died. The external

Table 1. Characteristics of Pediatric Patient Encounters With Suspected Infection in the First 24 Hours[a]

Characteristics	Derivation cohort	Internal validation cohort	External validation cohort
Encounters, No.	129 584	43 400	45 855
Resource setting, No. (%)			
Higher-resource settings	108 177 (83.5)	36 202 (83.4)	33 020 (72.0)
Lower-resource settings	21 407 (16.5)	7198 (16.6)	12 835 (28.0)
Age, median (IQR), y	3.7 (0.9-9.4)	3.7 (0.9-9.3)	2.6 (0.6-7.6)
Sex, No. (%)			
Female	62.868 (48.5)	21 041 (48.5)	22 295 (48.6)
Male	66 712 (51.5)	22 357 (51.5)	21 555 (47.0)
Race, No. (%)[b]			
American Indian or Alaska Native	109 (0.1)	21 (<0.1)	59 (0.1)
Asian	5149 (4.0)	1703 (3.9)	506 (1.1)
Black	22 709 (17.5)	7512 (17.3)	7476 (16.3)
Native Hawaiian or Other Pacific Islander	105 (0.1)	31 (0.1)	70 (0.2)
White	57 518 (44.4)	19 533 (45.0)	23 545 (51.3)
Multiple	22 113 (17.1)	7343 (16.9)	277 (0.6)
Other/unknown	22 095 (17.1)	7309 (16.8)	1.4 051 (30.6)
Hispanic or Latino ethnicity, No. (%)	33 698 (26.0)	11 457 (26.4)	55 (0.1)
Major comorbidities, No. (%)			
Technology dependence	18 951 (17.5)	6011 (16.6)	5677 (17.2)
Severe malnutrition	13 505 (10.4)	4478 (10.3)	3417 (7.5)
Malignancy	10 924 (10.1)	3709 (10.2)	2950 (8.9)
Transplant	3689 (3.4)	1287 (3.6)	1573 (4.8)
Comorbidities per PCCC, No. (%)[c]			
No known prior comorbidity	72 291 (66.8)	24 470 (67.6)	22 553 (68.3)
1 PCCC	9406 (8.7)	3150 (8.7)	2580 (7.8)
≥2 PCCCs	26 480 (24.5)	8582 (23.7)	7887 (23.9)
Systemic inflammatory response syndrome, No. (%)[d]	56 711 (43.8)	18 848 (43.4)	21 436 (46.7)
Locations visited during encounter (not mutually exclusive), No. (%)			
Presented to emergency department	92 507 (71.6)	31 092 (71.9)	26 940 (61.6)
≥1 Intensive care unit stays	23 128 (17.9)	7840 (18.1)	10 702 (23.4)
≥1 Operating room visits	17 604 (13.6)	6098 (14.1)	469 (1.1)
Outcomes, No. (%)			
Death	1538 (1.2)	527 (1.2)	540 (1.2)
Early death or extracorporeal membrane oxygenation	834 (0.6)	305 (0.7)	349 (0.8)

Abbreviation: PCCC, pediatric complex chronic condition.

[a] 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. Data from the 7 development sites are stratified by the 75% derivation cohort vs the 25% internal validation cohort.

[b] For race categories, "multiple" indicates that in the electronic health record, a patient's race was recorded as "multiracial," "multiple," or "2 or more races." "Other/unknown" indicates that a patient's race was recorded in the electronic health record 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.

[c] The PCCC system classifies pediatric chronic diseases using *International Classification of Diseases* diagnosis and procedure codes and was assessed

only at higher-resource sites, where the information was available (percentages for PCCC-related counts are based on higher-resource setting encounters).[8] The major comorbidities of technology dependence (eg, requiring gastrostomy, tracheostomy, central line), malignancy, and transplant were defined in the PCCC system. Severe malnutrition was defined as based on <3 SDs below the mean based on weight-for-age standards from the World Health Organization and assessed at all sites.[9] Early death is defined as death <72 hours after the beginning of the encounter.

[d] Systemic inflammatory response syndrome is assessed using temperature, white blood cell count, heart rate, and respiratory rate, with higher values reflecting more inflammation. Criteria are met when ≥2 values are outside the threshold for age, including at least temperature or white blood cell count. See eAppendix 1 in Supplement 1 for additional details.

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; eTable 5 in Supplement 1).

Best-Performing Individual Organ Dysfunction Criteria
Organ dysfunction subscore input availability and missingness are shown in eFigure 3, A-H, in Supplement 1. By 24 hours

into an encounter, most patients in higher-resource settings had information recorded for pulse oximetry oxygen saturation (SpO_2), respiratory support, platelet count, blood pressure, vasoactive agent use, and Glasgow Coma Scale score. Many also had fraction of inspired oxygen (FIO_2), 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 version 2 [PELOD-2] and vasoactive medication count), hematology/coagulation (Disseminated Intravascular Coagulation score), respiratory (pediatric Sequential Organ Failure Assessment [pSOFA]), renal (pSOFA), hepatic (IPSCC), neurologic (PELOD-2), immunologic (Pediatric Organ Dysfunction Information Update Mandate [PODIUM]), and endocrine dysfunction (PODIUM), as shown in eFigure 4 in Supplement 1.

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 and eFigure 6 in Supplement 1). Overall, AUPRCs and AUROCs were similar between these 2 models (eFigure 7 in Supplement 1). The task force evaluated the 2 models and chose to advance the 4-organ system model because it had similar performance but greater simplicity and lower dependence on laboratory measures. The task force acknowledged that the more comprehensive 8-organ system model may have utility in some circumstances (eg, research). The 4-organ system model included criteria for respiratory (mechanical ventilation, $PaO_2{:}FIO_2$, and $SpO_2{:}FIO_2$ ratios), cardiovascular (mean arterial pressure, lactate level, and vasoactive medications), coagulation (platelet count, international normalized ratio, D-dimer, and fibrinogen), and neurologic (Glasgow Coma Scale and pupillary reaction) dysfunction.

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 reweighted using a grid search and collapsed into a single level when performance was unaffected (eg, 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 1** and **Figure 2**; eFigure 5 in Supplement 1). The Phoenix Sepsis Score had AUPRCs of 0.23 to 0.38 (95% CI range, 0.20-0.39) and AUROCs of 0.71 to 0.92 (95% CI range, 0.70-0.92) to predict mortality in the internal and external validation sets, similar to the stacked sepsis model (**Figure 3**; eFigures 6-8 in Supplement 1). Compared with 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 3 of the 4 lower-resource sites (Figure 3). A notable limitation is that lower-resource sites 2-4 did not record respiratory support, even when a 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 in Supplement 1).

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

The task force chose a Phoenix Sepsis Score of 2 or greater in patients with suspected infection as the new sepsis criteria, and sepsis with 1 or more cardiovascular points 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 (IQR, 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 sensitivity that was comparable with or higher than the IPSCC sepsis, severe sepsis, and septic shock criteria across all settings and using the secondary outcome of early death or ECMO (**Figure 4**; eFigure 10 and eTable 7 in Supplement 1). For example, for the primary outcome of death at 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 with 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 the lower-resource site that 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 with 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 1 or more points 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 higher-resource sites and 32.3% in lower-resource sites (eFigure 11 in Supplement 1).

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 in Supplement 1). 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 in Supplement 1).

Clinical vignettes for children presenting with sepsis and septic shock and their corresponding Phoenix Sepsis Score data are provided in eAppendix 2 in Supplement 1.

Discussion

New criteria for pediatric sepsis and septic shock were derived and validated by developing and curating a clinical database with more than 3.6 million pediatric hospital encounters

Table 2. The Phoenix Sepsis Score[a]

	0 Points	1 Point	2 Points	3 Points
Respiratory (0-3 points)				
	$Pao_2:Fio_2$ ≥400 or $Spo_2:Fio_2$ ≥292[b]	$Pao_2:Fio_2$ <400 and any respiratory support[c] or $Spo_2:Fio_2$ <292 and any respiratory support[c]	$Pao_2:Fio_2$ 100-200 and IMV or $Spo_2:Fio_2$ 148-220 and IMV	$Pao_2:Fio_2$ <100 and IMV or $Spo_2:Fio_2$ <148 and IMV
Cardiovascular (0-6 points)				
		1 point each (up to 3) for:	**2 points each (up to 6) for:**	
	No vasoactive medications[d]	1 Vasoactive medication[d]	≥2 Vasoactive medications[d]	
	Lactate <5 mmol/L[e]	Lactate 5-10.9 mmol/L[e]	Lactate ≥11 mmol/L[e]	
Mean arterial pressure by age, mm Hg[f,g]				
<1 mo	>30	17-30	<17	
1 to 11 mo	>38	25-38	<25	
1 to <2 y	>43	31-43	<31	
2 to <5 y	>44	32-44	<32	
5 to <12 y	>48	36-48	<36	
12 to 17 y	>51	38-51	<38	
Coagulation (0-2 points)[h]				
		1 point each (maximum of 2 points) for:		
	Platelets ≥100 × 10^3/μL	Platelets <100 × 10^3/μL		
	International normalized ratio ≤1.3	International normalized ratio >1.3		
	D-dimer ≤2 mg/L FEU	D-dimer >2 mg/L FEU		
	Fibrinogen ≥100 mg/dL	Fibrinogen <100 mg/dL		
Neurologic (0-2 points)[i]				
	Glasgow Coma Scale score >10[j]; pupils reactive	Glasgow Coma Scale score ≤10[j]	Fixed pupils bilaterally	

Abbreviations: FEU, fibrinogen equivalent units; Fio_2, fraction of inspired oxygen; IMV, invasive mechanical ventilation; Spo_2, pulse oximetry oxygen saturation.

[a] The Phoenix Sepsis Score may be calculated in the absence of some variables (eg, 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 a medical team based on clinical judgment. Unmeasured variables contribute no points to the score.

[b] Calculated only when Spo_2 is ≤97%.

[c] Respiratory dysfunction of 1 point can be assessed in any patient receiving oxygen, high-flow, noninvasive positive pressure, or IMV respiratory support, and includes $Pao_2:Fio_2$ <200 and $Spo_2:Fio_2$ <220 in children who are not receiving IMV.

[d] Vasoactive medications include any dose of epinephrine, norepinephrine, dopamine, dobutamine, milrinone, and/or vasopressin (for shock).

[e] Lactate can be arterial or venous. Lactate reference range is 0.5-2.2 mmol/L.

[f] Use measured mean arterial pressure preferentially (invasive arterial if available or noninvasive oscillometric), and if measured mean arterial pressure is not available, a calculated mean arterial pressure (⅓ × systolic + ⅔ × diastolic) may be used as an alternative.

[g] Age is not adjusted for prematurity, and the criteria do not apply to birth hospitalizations, children with postconceptional age <37 weeks, or those aged ≥18 years.

[h] Coagulation variable reference ranges: platelets, 150-450 × 10^3/μL; D-dimer, <0.5 mg/L FEU; fibrinogen, 180-410 mg/dL. International normalized ratio reference range is based on local reference prothrombin time.
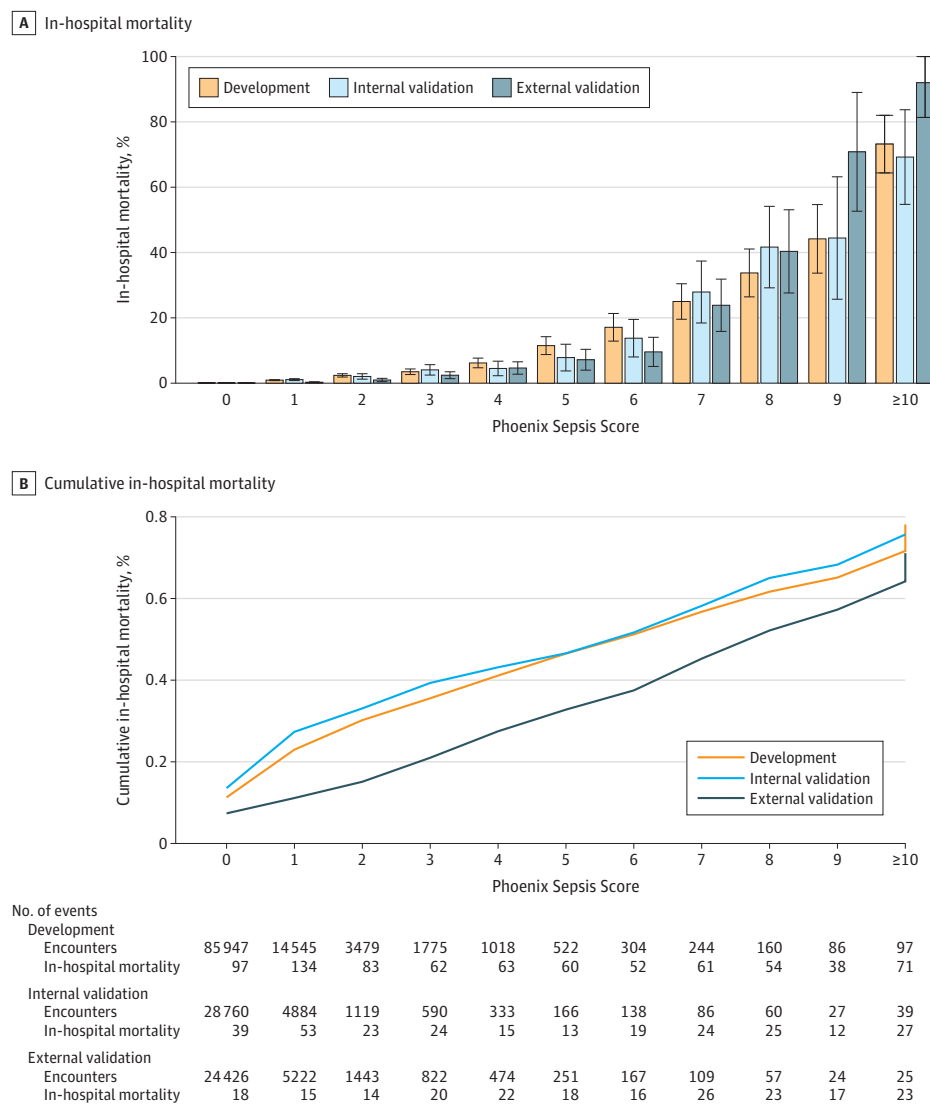
[i] The neurologic dysfunction subscore was pragmatically validated in both sedated and nonsedated patients and those with and without IMV support.

[j] The Glasgow Coma Scale score measures level of consciousness based on verbal, eye, and motor response and ranges from 3 to 15, with a higher score indicating better neurologic function.

at 10 sites in 5 countries. The development data set 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 prespecified 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 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.[4] 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 (eg, the 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

Figure 1. In-Hospital Mortality Associated With the Phoenix Sepsis Score in Patients in Higher-Resource Settings With Suspected Infection in the First 24 Hours
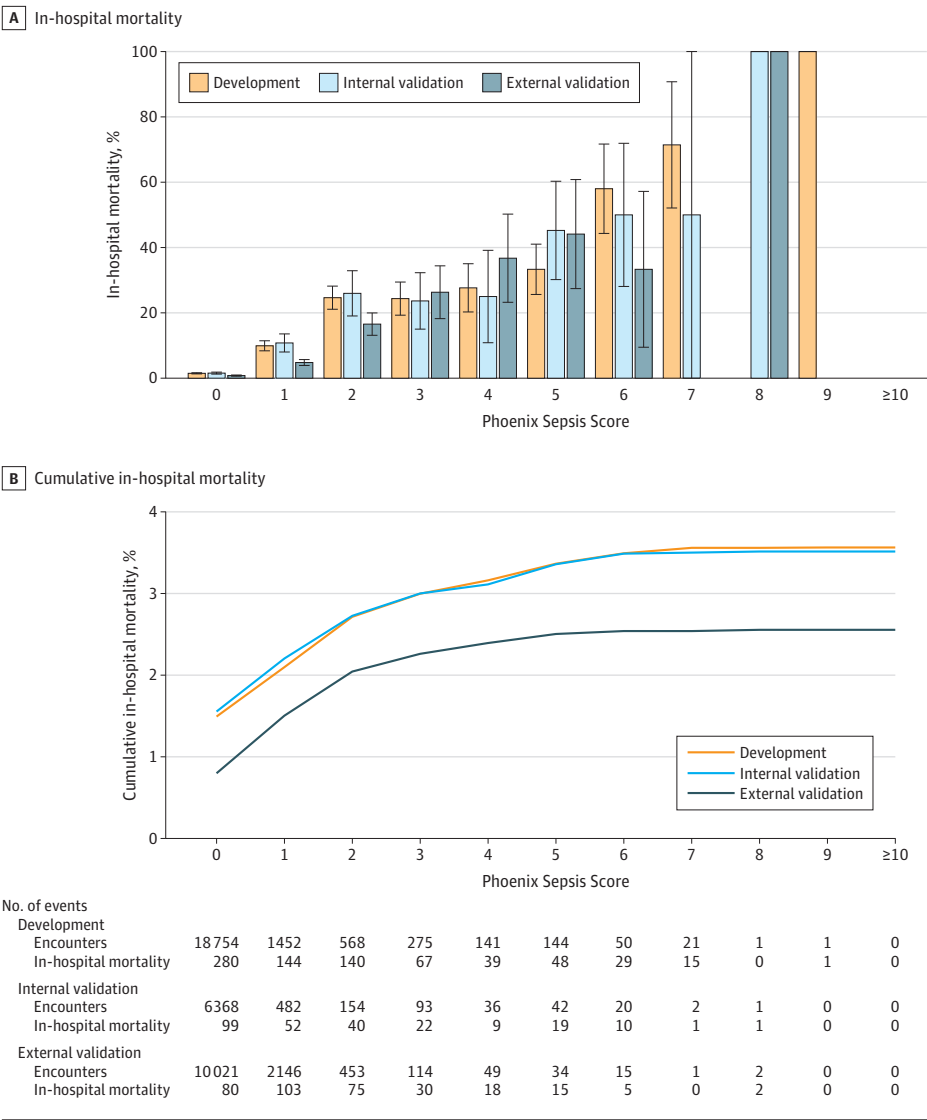
A In-hospital mortality

B Cumulative in-hospital mortality

No. of events											
Development											
Encounters	85 947	14 545	3479	1775	1018	522	304	244	160	86	97
In-hospital mortality	97	134	83	62	63	60	52	61	54	38	71
Internal validation											
Encounters	28 760	4884	1119	590	333	166	138	86	60	27	39
In-hospital mortality	39	53	23	24	15	13	19	24	25	12	27
External validation											
Encounters	24 426	5222	1443	822	474	251	167	109	57	24	25
In-hospital mortality	18	15	14	20	22	18	16	26	23	17	23

This figure shows calibration of the Phoenix Sepsis Score in higher-resource settings (sites with more technological resources, eg, laboratory equipment, ventilators, and kidney replacement therapy devices, to support organ dysfunction). For patients with suspected infection who have each possible integer value of the Phoenix Sepsis Score in the first 24 hours of the encounter, mortality among those at the development, internal validation, and external validation sites is shown. Binomial confidence intervals (whiskers) for the mortality point estimate in each group are also shown.

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 were primarily evaluated using the AUPRC, instead of the AUROC, with the goal of maximizing the PPV and sensitivity of the final criteria. The AUPRC is considered a better measure of classification performance for rare events (in this case, deaths) compared with the AUROC, which can have inflated performance when the proportions of events (deaths) and nonevents (survivors) are imbalanced,[11,20] an issue that is particularly relevant in children with infections given their lower mortality compared with 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.[21]

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 health care globally.[22] Most of the vital signs, laboratory tests, and interventions included in the Phoenix Sepsis Score are routinely collected in most lower-resource settings and in nearly all higher-resource settings, according to the Pediatric Sepsis Definition Task Force's international survey.[23] Even in settings where not all variables are available, the Phoenix Sepsis Score is 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

Figure 2. In-Hospital Mortality Associated With the Phoenix Sepsis Score in Patients in Lower-Resource Settings With Suspected Infection in the First 24 Hours

A In-hospital mortality

B Cumulative in-hospital mortality

No. of events											
Development											
Encounters	18 754	1452	568	275	141	144	50	21	1	1	0
In-hospital mortality	280	144	140	67	39	48	29	15	0	1	0
Internal validation											
Encounters	6368	482	154	93	36	42	20	2	1	0	0
In-hospital mortality	99	52	40	22	9	19	10	1	1	0	0
External validation											
Encounters	10 021	2146	453	114	49	34	15	1	2	0	0
In-hospital mortality	80	103	75	30	18	15	5	0	2	0	0

This figure shows the calibration of the Phoenix Sepsis Score in lower-resource settings (sites with fewer technological resources to support organ dysfunction). For patients with suspected infection who have each possible integer value of the Phoenix Sepsis Score in the first 24 hours of the encounter, mortality among those at the development, internal validation, and external validation sites is shown. Binomial confidence intervals (whiskers) for the mortality point estimate in each group are also shown. At lower-resource sites, some variables were rarely available (eg, D-dimer and fibrinogen for coagulation dysfunction), even when other variables for the same organ systems were recorded (eg, platelet count and international normalized ratio); thus, the maximum cumulative score achieved at lower-resource sites was 9, instead of the maximum possible of 13.

lower-resource settings had a median Phoenix Sepsis Score of 3 points. This feature was primarily assessed in the data sets from lower-resource settings. For example, although platelets were commonly measured at most sites, coagulation tests (eg, 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.[24] 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 accompanying consensus criteria article.[6]

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 data set, which represents a lower- to middle-income country, respiratory support (eg, mechanical ventilation, F_{IO_2}) and neurologic assessments (eg, level of consciousness and pupillary reaction) are performed but not recorded in the clinical information systems. This reduces the ability to assess the score and criteria at that site. In contrast, score performance was

Figure 3. Mortality Prediction Performance of the Phoenix Sepsis Score and Organ Dysfunction Scores

A Area under the precision recall curve (AUPRC)

	Internal validation set						
	Phoenix sepsis	IPSCC	Phoenix-8	PELOD-2	pSOFA	Proulx	PODIUM
Higher-resource sites 1-5	0.28 (0.27-0.28)	0.16 (0.16-0.16)	0.28 (0.27-0.28)	0.30 (0.30-0.31)	0.17 (0.17-0.18)	0.22 (0.22-0.22)	0.20 (0.20-0.21)
Lower-resource site 1	0.37 (0.35-0.39)	0.18 (0.16-0.20)	0.33 (0.31-0.35)	0.22 (0.20-0.24)	0.27 (0.25-0.29)	0.25 (0.23-0.27)	0.24 (0.23-0.26)
Lower-resource site 2	0.31 (0.30-0.32)	0.27 (0.26-0.28)	0.34 (0.32-0.35)	0.16 (0.15-0.17)	0.33 (0.32-0.34)	0.20 (0.19-0.22)	0.24 (0.22-0.25)
	External validation set						
Higher-resource site 6	0.38 (0.37-0.38)	0.16 (0.15-0.16)	0.37 (0.36-0.37)	0.35 (0.34-0.35)	0.20 (0.20-0.21)	0.21 (0.20-0.21)	0.22 (0.21-0.22)
Lower-resource site 3	0.26 (0.24-0.28)	0.15 (0.13-0.16)	0.23 (0.21- 0.24)	0.16 (0.14- 0.18)	0.14 (0.13-0.16)	0.16 (0.15-0.18)	0.13 (0.12-0.15)
Lower-resource site 4	0.23 (0.22-0.24)	0.13 (0.13-0.14)	0.20 (0.19-0.206)	0.11 (0.10-0.12)	0.18 (0.18-0.19)	0.15 (0.14-0.15)	0.11 (0.11-0.12)
All sites (internal and external validation sets)	0.21 (0.20-0.21)	0.12 (0.12-0.12)	0.20 (0.20-0.21)	0.18 (0.18- 0.18)	0.15 (0.15-0.15)	0.15 (0.15-0.15)	0.14 (0.14-0.14)

B Area under the receiver operating characteristic curve (AUROC)

	Internal validation set						
	Phoenix sepsis	IPSCC	Phoenix-8	PELOD-2	pSOFA	Proulx	PODIUM
Higher-resource sites 1-5	0.88 (0.88-0.88)	0.88 (0.88-0.88)	0.91 (0.90-0.91)	0.86 (0.86-0.87)	0.90 (0.89-0.90)	0.86 (0.85-0.86)	0.89 (0.89-0.90)
Lower-resource site 1	0.91 (0.90-0.92)	0.85 (0.83-0.86)	0.90 (0.89-0.91)	0.84 (0.83-0.86)	0.89 (0.87-0.90)	0.90 (0.89-0.91)	0.77 (0.75-0.79)
Lower-resource site 2	0.71 (0.70-0.72)	0.78 (0.77-0.80)	0.85 (0.84-0.86)	0.78 (0.77- 0.79)	0.83 (0.82-0.84)	0.72 (0.70-0.73)	0.78 (0.77-0.79)
	External validation set						
Higher-resource site 6	0.92 (0.92-0.92)	0.91 (0.91-0.92)	0.94 (0.94-0.94)	0.92 (0.92-0.92)	0.93 (0.93-0.93)	0.91 (0.91-0.91)	0.92 (0.91-0.92)
Lower-resource site 3	0.81 (0.80-0.83)	0.76 (0.74-0.78)	0.78 (0.76-0.79)	0.70 (0.67-0.71)	0.73 (0.71-0.75)	0.71 (0.69-0.73)	0.71 (0.69-0.73)
Lower-resource site 4	0.80 (0.79-0.81)	0.81 (0.80-0.81)	0.80 (0.79-0.80)	0.73 (0.72-0.74)	0.82 (0.81- 0.83)	0.74 (0.73-0.75)	0.75 (0.74-0.76)
All sites (internal and external validation sets)	0.82 (0.82-0.83)	0.83 (0.83-0.84)	0.87 (0.87-0.87)	0.80 (0.80-0.81)	0.86 (0.86-0.87)	0.81 (0.81-0.81)	0.84 (0.83-0.84)

IPSCC indicates International Pediatric Sepsis Consensus Conference; PELOD-2, Pediatric Logistic Organ Dysfunction version 2; PODIUM, Pediatric Organ Dysfunction Information Update Mandate; and pSOFA, pediatric Sequential Organ Failure Assessment. This figure compares the performance of the Phoenix Sepsis Score with 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 extracorporeal membrane oxygenation, are shown in eFigure 7 in Supplement 1. All types of organ dysfunction 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 counts of organ dysfunction (eAppendix 1 and eTable 2 in Supplement 1). Performance is presented as both quantitative with 95% CIs (calculated using logit transform), as well as visually using a color heat map. Shading indicates highest (darkest) to lowest (lightest) in each row. The 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. The AUPRC is a more reliable classifier performance metric than the AUROC when the classes are imbalanced, for example, when mortality is very low, as in this study. The AUROC is the area under a curve drawn with the false-positive rate on the x-axis and the true-positive rate on the y-axis. In this study, it is an indicator of how well a classifier can rank encounters with respect to mortality risk.

excellent at lower-resource site 1 and comparable with 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 Glasgow Coma Scale 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

Figure 4. Comparison of Sensitivity and PPV of Novel Phoenix Sepsis Criteria With Current IPSCC Sepsis and Severe Sepsis Criteria Across Outcomes and Patient Subgroups in the Internal Validation Sets
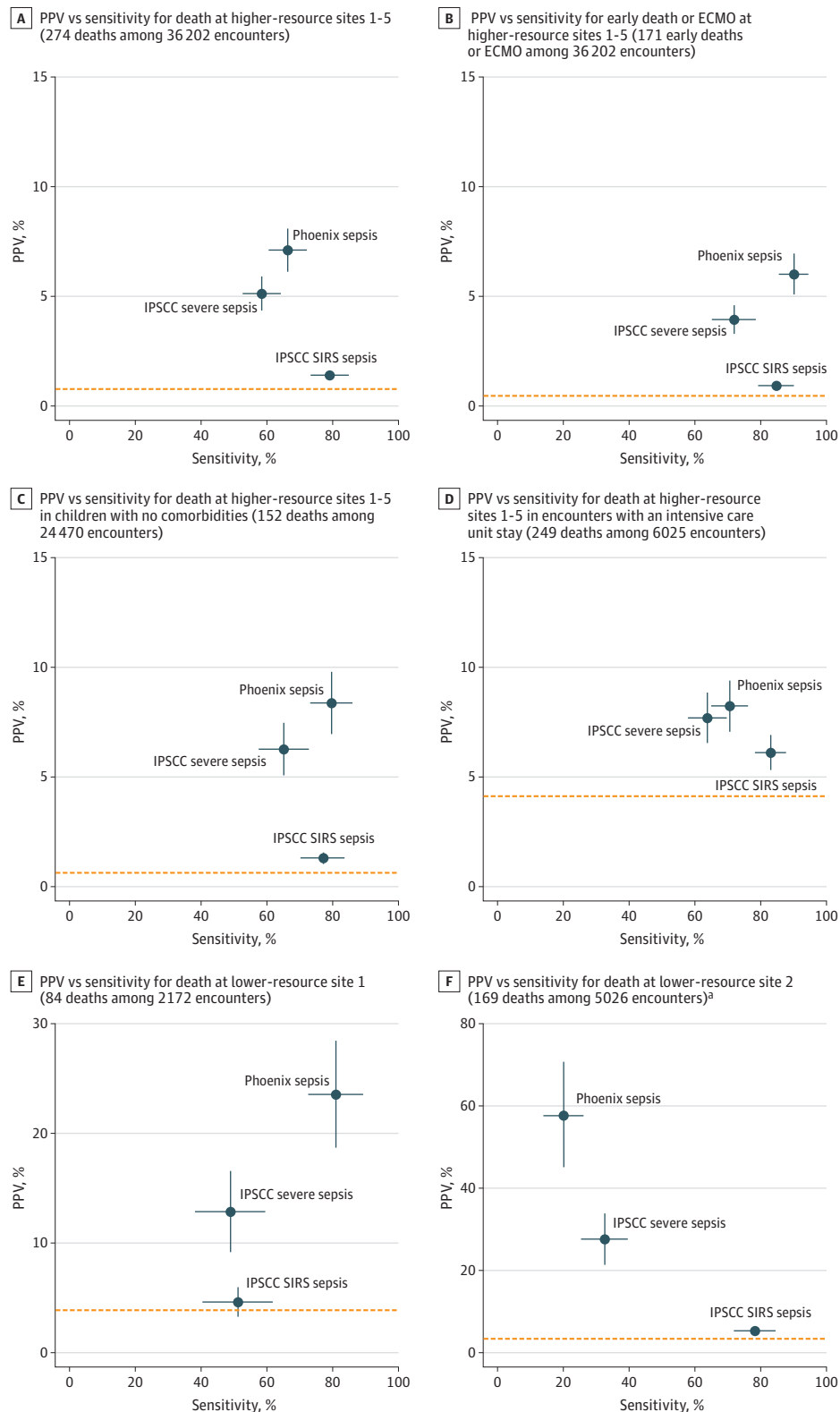
The positive predictive value (PPV, or precision) and sensitivity for the Phoenix vs 2005 International Pediatric Sepsis Consensus Conference (IPSCC) criteria for sepsis in children with suspected infection are shown. The Phoenix sepsis criteria are based on achieving ≥2 points in the Phoenix Sepsis Score among patients with suspected infection in the first 24 hours of an encounter. The IPSCC sepsis and severe sepsis criteria are based on systemic inflammatory response syndrome (SIRS) and IPSCC-based organ dysfunction among patients with suspected infection in the first 24 hours of an encounter. Baseline rates of the outcome in each group (death, or early death or extracorporeal membrane oxygenation [ECMO]) are shown as horizontal dashed lines. 95% CIs are shown as bands from each point in the plane representing that component (eg, CIs for PPV are parallel to the y-axis). Confidence bands that are not visible are narrow enough to be completely hidden by the point. These figures are similar to area under the precision recall curves except at a single threshold for criteria that generate a binary response (eg, yes/no sepsis criteria met) instead of across the range of possible points in the curve (eg, 0-13 points in the Phoenix Sepsis Score; see Figure 3). Better-performing criteria are closer to the top right corner. A trade-off exists between sensitivity and PPV, 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 have poor predictive value.

[a] At lower-resource site 2, some Phoenix Sepsis Score and IPSCC data inputs (eg, invasive mechanical ventilation, Glasgow Coma Scale score) are not recorded even when they are performed; thus, assessment of criteria performance is limited. Lower-resource site 1 and all higher-resource sites have inputs for all relevant organ systems in the criteria. Comparison of sepsis criteria in the external validation sites is shown in eFigure 10 in Supplement 1 with similar results. Diagnostic performance measures for this comparison are shown in eTable 7 in Supplement 1.

the criteria is needed. Third, similar to the Sepsis-3 validation study, unique criteria for patients with chronic organ dysfunction were not developed.[4] Fourth, few databases from lower-resource settings were available (a form of data poverty),[25] and the ones used may not be generalizable to every low-resource environment. Fifth, the data from higher-resource settings were exclusively from tertiary US pediatric centers. Sixth, the data sets from some of the sites included 10 years of data, possibly including changes in practice during that time frame.

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 improved performance for the diagnosis of pediatric sepsis and septic shock compared with the existing IPSCC criteria.

Author Affiliations: Departments of Pediatrics (Critical Care) and Preventive Medicine (Health and Biomedical Informatics), Northwestern University Feinberg School of Medicine, and Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois (Sanchez-Pinto); Departments of Biomedical Informatics and Pediatrics (Critical Care Medicine), University of Colorado School of Medicine, and Children's Hospital Colorado, Aurora (Bennett, Martin); Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora (DeWitt, Russell, Rebull); Kenya Medical Research Institute (KEMRI)–Wellcome Trust Research Programme, Nairobi, Kenya (Akech); Departments of Biomedical Informatics, Bioengineering, Biostatistics, and Informatics, University of Colorado School of Medicine, Aurora (Albers); Department of Biomedical Informatics, Columbia University, New York, New York (Albers); Division of Emergency Medicine, Department of Pediatrics, Ann and Robert H. Lurie Children's Hospital of Chicago, and Northwestern University Feinberg School of Medicine, Chicago, Illinois (Alpern); Department of Pediatrics, University of Pennsylvania, Perelman School of Medicine and Division of Emergency Medicine, Children's Hospital of Philadelphia, Philadelphia (Balamuth); Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland (Bembea); Intensive Care Unit, Dhaka Hospital, Nutrition Research Division, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh (Chisti); Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania (Evans, Horvat); Pediatric Intensive Care Unit, Hospital General de Medellín Luz Castro de Gutiérrez and Hospital Pablo Tobón Uribe, and Red Colaborativa Pediátrica de Latinoamérica (LARed Network), Medellín, Colombia (Jaramillo-Bustamante); Department of Pediatrics, University of British Columbia, Vancouver, Canada (Kissoon); Department of Pediatrics, Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa, Canada (Menon); Department of Pediatrics (Pediatric Emergency Medicine), University of Colorado School of Medicine, and Children's Hospital Colorado, Aurora (Scott); Division of Critical Care, Department of Pediatrics, Nemours Children's Health, Wilmington, Delaware (Weiss); Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania (Weiss); Department of Anesthesiology, Pharmacology, and Therapeutics, Faculty of Medicine, University of British Columbia, Vancouver, Canada (Wiens); Institute for Global Health, BC Children's Hospital, Vancouver, British Columbia, Canada (Wiens); Walimu, Kampala, Uganda (Wiens); Seattle Children's Hospital and Department of Pediatrics, University of Washington School of Medicine, Seattle (Zimmerman); Paediatrics and Child Health, University of Cape Town Faculty of Health Sciences, Cape Town, South Africa (Argent); Department of Pediatrics, Northwestern University Feinberg School of Medicine, and Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois (Sorce); Department of Intensive Care and Neonatology, Children's Research Center, University Children's Hospital Zurich, University of Zurich, Zurich, Switzerland (Schlapbach); Child Health Research Centre, The University of Queensland, Brisbane, Australia (Schlapbach); Department of Pediatrics, University of Washington, and Center for Child Health, Behavior, and Development and Pediatric Critical Care, Seattle Children's Hospital, Seattle (Watson).

The Society of Critical Care Medicine Pediatric Sepsis Definition Task Force Authors: Paolo Biban, MD; Enitan Carrol, MD, MBChB; Kathleen Chiotos, MD; Claudio Flauzino De Oliveira, MD, PhD; Mark W. Hall, MD; David Inwald, MB, MB BChir, PhD; Paul Ishimine, MD; Michael Levin, MD, PhD; Rakesh Lodha, MD; Simon Nadel, MBBS; Satoshi Nakagawa, MD; Mark J. Peters, PhD; Adrienne G. Randolph, MD, MS; Suchitra Ranjit, MD; Daniela Carla Souza, MD; Pierre Tissieres, MD, DSc; James L. Wynn, MD.

Affiliations of The Society of Critical Care Medicine Pediatric Sepsis Definition Task Force Authors: Verona University Hospital, Verona, Italy (Biban); University of Liverpool, Liverpool, England (Carrol); Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Chiotos); Associação de Medicina Intensiva Brasileira, São Paulo, Brazil (Flauzino De Oliveira); Nationwide Children's Hospital, Columbus, Ohio (Hall); Addenbrooke's Hospital, Cambridge University Hospital NHS Trust, Cambridge, England (Inwald); University of California, San Diego School of Medicine, La Jolla (Ishimine); Imperial College London, London, England (Levin); All India Institute of Medical Sciences, New Delhi, India (Lodha); St Mary's Hospital, London, England (Nadel); National Center for Child Health and Development, Tokyo, Japan (Nakagawa); University College London Great Ormond Street Institute of Child Health, London, England (Peters); Boston Children's Hospital, Boston, Massachusetts (Randolph); Apollo Children's Hospital, Chennai, India (Ranjit); University Hospital of the University of São Paulo, Sao Paulo, Brazil (Souza); Hôpital de Bicêtre, Paris, France (Tissieres); University of Florida, Gainesville (Wynn).

REFERENCES

1. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017. *Lancet*. 2020;395(10219):200-211. doi:10.1016/S0140-6736(19)32989-7

2. Goldstein B, Giroir B, Randolph A, et al. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med*. 2005;6(1):2-8. doi:10.1097/01.PCC.0000149131.72248.E6

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. doi:10.1186/s13054-015-1055-x

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. doi:10.1001/jama.2016.0288

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. doi:10.1001/jama.2016.0287

6. Schlapbach LJ, Watson RS, Sorce LR, et al; Society of Critical Care Medicine Pediatric Sepsis Definition Task Force. International consensus criteria for pediatric sepsis and septic shock. *JAMA*. Published online January 21, 2023. doi:10.1001/jama.2024.0179

7. Zeng X, Yu G, Lu Y, et al. PIC, a paediatric-specific intensive care database. *Sci Data*. 2020;7(1):14. doi:10.1038/s41597-020-0355-4

8. Feinstein JA, Russell S, DeWitt PE, et al. R package for pediatric complex chronic condition classification. *JAMA Pediatr*. 2018;172(6):596-598. doi:10.1001/jamapediatrics.2018.0256

9. 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.

10. Ozenne B, Subtil F, Maucort-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. doi:10.1016/j.jclinepi.2015.02.010

11. 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. doi:10.1371/journal.pone.0118432

12. Wolpert DH. Stacked generalization. *Neural Netw*. 1992;5(2):241-259. doi:10.1016/S0893-6080(05)80023-1

13. Breiman L. Stacked regressions. *Mach Learn*. 1996;24(1):49-64. doi:10.1007/BF00117832

14. Proulx F, Fayon M, Farrell CA, et al. Epidemiology of sepsis and multiple organ dysfunction syndrome in children. *Chest*. 1996;109(4):1033-1037. doi:10.1378/chest.109.4.1033

15. Matics TJ, Sanchez-Pinto LN. Adaptation and validation of a pediatric Sequential Organ Failure Assessment score and evaluation of the Sepsis-3 definitions in critically ill children. *JAMA Pediatr*. 2017;171(10):e172352. doi:10.1001/jamapediatrics.2017.2352

16. 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. doi:10.1097/CCM.0b013e31828a2bbd

17. Rousseaux J, Grandbastien B, Dorkenoo A, et al. Prognostic value of shock index in children with septic shock. *Pediatr Emerg Care*. 2013;29(10):1055-1059. doi:10.1097/PEC.0b013e3182a5c99c

18. Khemani RG, Bart RD, Alonzo TA, et al. Disseminated intravascular coagulation score is associated with mortality for children with shock. *Intensive Care Med*. 2009;35(2):327-333. doi:10.1007/s00134-008-1280-8

19. Haque A, Siddiqui NR, Munir O, et al. Association between vasoactive-inotropic score and mortality in pediatric septic shock. *Indian Pediatr*. 2015;52(4):311-313. doi:10.1007/s13312-015-0630-1

20. Tharwat A. Classification assessment methods. *Appl Comput Inform*. 2020;17(1):168-192. doi:10.1016/j.aci.2018.08.003

21. Scott HF, Brilli RJ, 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. doi:10.1097/CCM.0000000000004505

22. Wyber R, Vaillancourt S, Perry W, et al. Big data in global health. *Bull World Health Organ*. 2015;93(3):203-208. doi:10.2471/BLT.14.139022

23. Morin L, Hall M, de Souza D, et al. The current and future state of pediatric sepsis definitions. *Pediatrics*. 2022;149(6):e2021052565. doi:10.1542/peds.2021-052565

24. Jimenez-Zambrano A, Ritger C, Rebull M, et al. Clinical decision support tools for paediatric sepsis in resource-poor settings. *BMJ Open*. 2023;13(10):e074458. doi:10.1136/bmjopen-2023-074458

25. Ibrahim H, Liu X, Zariffa N, et al. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit Health*. 2021;3(4):e260-e265. doi:10.1016/S2589-7500(20)30317-4