

ORIGINAL ARTICLE

FDA-Authorized AI/ML Tool for Sepsis Prediction: Development and Validation

A. Bhargava ¹, C. López-Espina ¹, L. Schmalz ¹, S. Khan ¹, G.L. Watson ¹, D. Urdiales ¹, L. Updike ¹, N. Kurtzman ², A. Dagan ³, A. Doodlesack ³, B.A. Stenson ³, D. Sarma ³, E. Reseland ³, J.H. Lee ³, M.S. Kravitz ^{3,4}, P.S. Antkowiak ³, T. Shvilkina ³, A. Espinosa ⁵, A. Halalau ⁵, C. Demarco ⁵, F. Davila ⁵, H. Davila ⁵, M. Sims ⁵, N. Maddens ⁵, R. Berghea ⁵, S. Smith ⁵, A.V. Palagiri ⁶, C. Ezekiel ⁷, F. Sadaka ⁷, K. Iyer ⁷, M. Crisp ⁷, S. Azad ⁷, V. Oke ⁷, A. Friederich ⁸, A. Syed ⁸, F. Gosai ⁸, L. Chawla ⁸, N. Evans ⁹, K. Thomas ¹⁰, R. Malkani ¹⁰, R. Patel ¹⁰, S. Mayer ¹⁰, F. Ali ¹¹, L. Raghavakurup ¹¹, M. Tafa ¹¹, S. Singh ¹¹, S. Raouf ¹¹, S.D. Zhao ¹², R. Zhu ¹², R. Bashir ¹³, B. Reddy Jr. ¹ and N.I. Shapiro ^{1,3,4}

Received: August 30, 2024; Accepted: September 26, 2024; Published: November 27, 2024

Abstract

BACKGROUND Sepsis is a life-threatening condition that demands prompt treatment for improved patient outcomes. Its heterogenous presentation makes early detection challenging, highlighting the need for effective risk assessment tools. Artificial intelligence (AI) models could potentially identify patients with sepsis, but none have previously been authorized by the U.S. Food and Drug Administration (FDA) for commercial use. This study outlines the development and validation of the Sepsis ImmunoScore, the first FDA-authorized AI-based software designed to identify patients at risk of sepsis.

METHODS In this prospective study, we enrolled adult patients (18+ years of age) suspected of infection, as indicated by a blood culture order, from five U.S. institutions between April 2017 and July 2022. The participants were divided into an algorithm development cohort (n=2366), an internal validation cohort (n=393), and an external validation cohort (n=698). The primary end point was sepsis presence (as defined by Sepsis-3) within 24 hours of test initiation. Secondary end points included length of hospital stay, intensive care unit (ICU) admission within 24 hours, mechanical ventilation use within 24 hours, vasopressor use within 24 hours, and in-hospital mortality.

RESULTS The Sepsis ImmunoScore demonstrated high diagnostic accuracy, with an area under the curve of 0.85 (0.83 to 0.87) in the derivation cohort, 0.80 (0.74 to 0.86) in internal validation, and 0.81 (0.77 to 0.86) in external validation. The scores were categorized into four sepsis risk levels with corresponding likelihood ratios: low (0.1), medium (0.5), high (2.1), and very high (8.3). These risk categories also predicted in-hospital mortality: low (0.0%), medium (1.9%), high (8.7%), and very high (18.2%) in the external validation cohort. Similar trends were observed for other metrics, such as length of hospital stay, ICU utilization, mechanical ventilation, and vasopressor use.

The author affiliations are listed at the end of the article.

The authors' full names and academic degrees are listed in the [Appendix](#).

Dr. Shapiro can be contacted at nshapiro@bidmc.harvard.edu or at Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, MA.

CONCLUSIONS The Sepsis ImmunoScore demonstrated high accuracy for the identification and prediction of sepsis and critical illness metrics that could enable prompt identification of patients at high risk of sepsis and adverse outcomes, potentially improving clinical decision-making and patient outcomes. (Funded by the Defense Threat Reduction Agency and others.)

Introduction

Sepsis is a serious medical condition caused by a dysregulated immune response to infection, which can lead to organ dysfunction and significant morbidity and mortality.¹ Early treatment, particularly with antibiotics, can improve patient outcomes.²⁻⁷ However, heterogeneity in the presentation of sepsis makes early recognition difficult, leading to increased mortality.⁸ As a result, there is a need for risk assessment tools that help clinicians quickly and accurately identify patients at high risk of sepsis. Many previously proposed risk assessment tools exist, including clinical approaches, laboratory tests, and sepsis-specific biomarkers; however, none are universally accepted as routine in clinical practice.

To address the need for a diagnostic and risk assessment tool in the hospital setting, we developed the Sepsis ImmunoScore, intended for integration with an electronic medical record (EMR), which uses machine learning (ML) to identify patients likely to have or progress to sepsis within 24 hours of patient assessment. It was granted marketing authorization (de novo pathway) by the U.S. Food and Drug Administration (FDA) in April 2024 as the first-ever artificial intelligence (AI) diagnostic tool authorized for sepsis.

The objective of this investigation was to evaluate the performance of the Sepsis ImmunoScore and its ability to risk-stratify patients for the presence or development of sepsis (as defined by Sepsis-3) within 24 hours, and for secondary end points of in-hospital mortality, length of hospital stay, intensive care unit (ICU) admission, mechanical ventilation, and vasopressor medication use.⁹

Methods

STUDY DESIGN

We conducted a prospective, observational, multicenter study to create a sepsis AI/ML algorithm and assess its ability to identify the presence of sepsis within 24 hours,

and other secondary outcomes of critical illness morbidity and mortality (Fig. S1). Participants were enrolled at one of five participating hospitals throughout the United States between April 2017 and July 2022 (Table 1). We obtained study approvals from the ethics boards of participating institutions under a waiver of informed consent, except OSF Saint Francis Medical Center, which required informed consent.

STUDY POPULATION

Study inclusion criteria consisted of hospitalized adult patients (18 years of age or older) who had a suspected infection, as defined by the clinical decision to obtain a blood culture, and who had a lithium-heparin (Li-Hep) plasma sample drawn within 6 hours of the first blood culture order that was available for collection. There were no exclusion criteria. Participants were assigned to one of three cohorts: a *derivation* cohort (n=2366) where the algorithm was derived, an *internal validation* cohort (n=393) that assessed algorithm performance on a second set of participants from the same hospitals used in the derivation, and a final *external validation* cohort (n=698) of participants from hospitals not involved in the algorithm derivation (additional details in the Supplementary Appendix, Study Population section).

STUDY OUTCOMES

End Points

The primary end point was the presence of sepsis at presentation or within 24 hours of study inclusion using the Sepsis-3 criteria: suspected infection and a sequential organ failure assessment score of 2 or greater from baseline.⁹ The derivation cohort used a Sepsis-3 outcome derived from the medical record in an automated fashion,^{9,10} while the internal and external validation cohorts used expert clinical adjudication to apply the definitions and determine the Sepsis-3 outcome. The clinical adjudication occurred in a retrospective fashion and was carried out by clinicians who likely did not treat the patient but had access to the entire hospital chart and utilized information including laboratory testing, radiology testing, and clinical assessment and decision-making documentation. The secondary end points consisted of sepsis-related metrics of critical illness: in-hospital mortality, length of hospital stay, ICU admission, use of mechanical ventilator, and use of vasopressors.

DATA COLLECTION

Data were gathered directly through an offline EMR extraction and a transfer of de-identified data that were

Table 1. Baseline Data and Adverse Outcomes for the Derivation, Internal Validation, and External Validation Cohorts.*

Characteristic	Patient Encounters, No. (%)		
	Derivation Cohort (n=2366)	Internal Validation Cohort (n=393)	External Validation Cohort (n=698)
Clinical site			
Beth Israel Deaconess Medical Center — Boston, MA	0 (0.0)	0 (0.0)	356 (51.0)
OSF Saint Francis Medical Center — Peoria, IL	712 (30.1)	87 (22.1)	0 (0.0)
Jesse Brown VA Medical Center — Chicago, IL	0 (0.0)	0 (0.0)	65 (9.3)
Mercy Health — St. Louis, MO	1061 (44.8)	306 (77.9)	0 (0.0)
William Beaumont University Hospital — Royal Oak, MI	0 (0.0)	0 (0.0)	277 (39.7)
Carle Foundation Hospital — Urbana, IL	593 (25.1)	0 (0.0)	0 (0.0)
Age (mean [±SD])	64.20 (16.59)	64.06 (17.66)	62.80 (17.01)
Male (%)	1195 (50.5)	210 (53.4)	391 (56.0)
Female (%)	1171 (49.5)	183 (46.6)	307 (44.0)
Race (%)†			
Indigenous or Alaska Native	1 (0.0)	0 (0.0)	2 (0.3)
Asian	12 (0.5)	2 (0.5)	14 (2.0)
Black or African American	315 (13.3)	57 (14.5)	154 (22.1)
Native Hawaiian or other Pacific Islander	0 (0.0)	0 (0.0)	1 (0.1)
Unknown	85 (3.6)	12 (3.1)	119 (17.0)
White	1953 (82.5)	322 (81.9)	408 (58.5)
Ethnicity			
Hispanic or Latino	26 (1.1)	2 (0.5)	96 (13.8)
Not Hispanic or Latino	1725 (72.9)	385 (98.0)	564 (80.8)
Unknown	615 (26.0)	6 (1.5)	38 (5.4)
High-risk comorbidities			
Acute myocardial infarction (%)	97 (4.1)	11 (2.8)	43 (6.2)
History of myocardial infarction (%)	101 (4.3)	21 (5.3)	54 (7.7)
Congestive heart failure (%)	583 (24.6)	103 (26.2)	170 (24.4)
Peripheral vascular disease (%)	225 (9.5)	49 (12.5)	72 (10.3)
Cerebrovascular disease (%)	130 (5.5)	38 (9.7)	65 (9.3)
Chronic obstructive pulmonary disease (%)	606 (25.6)	107 (27.2)	171 (24.5)
Dementia (%)	167 (7.1)	45 (11.5)	72 (10.3)
Paralysis (%)	68 (2.9)	10 (2.5)	20 (2.9)
Diabetes (%)	630 (26.6)	101 (25.7)	156 (22.3)
Diabetes with complications (%)	423 (17.9)	93 (23.7)	155 (22.2)
Renal disease (%)	659 (27.9)	123 (31.3)	216 (30.9)
Mild liver disease (%)	118 (5.0)	19 (4.8)	94 (13.5)
Moderate and severe liver disease (%)	45 (1.9)	6 (1.5)	55 (7.9)
Peptic ulcer disease (%)	45 (1.9)	8 (2.0)	13 (1.9)
Rheumatologic disease (%)	105 (4.4)	17 (4.3)	33 (4.7)
AIDS (%)	17 (0.7)	2 (0.5)	6 (0.9)
Immunocompromised (%)	470 (19.9)	117 (29.8)	190 (27.2)
Covid-19 (%)	189 (8.0)	28 (7.1)	73 (10.5)
Adverse outcomes			
Sepsis-3 within 24 hours (%)	763 (32.2)	108 (27.5)	151 (21.6)
In-hospital mortality (%)	147 (6.2)	33 (8.4)	33 (4.7)
ICU transfer (%)	491 (27.7)	144 (36.6)	151 (21.6)
Placement of mechanical ventilation (%)	191 (8.1)	44 (11.2)	51 (7.3)
Administration of vasopressors (%)	223 (9.4)	53 (13.5)	77 (11.0)
Length of stay (median [IQR])	4.7 [2.6, 8.5]	4.98 [2.8, 10.8]	5.94 [3.3, 10.4]

* AIDS denotes acquired immunodeficiency syndrome; Covid-19, coronavirus 2019; ICU, intensive care unit; IQR, interquartile range; SD, standard deviation; and VA, Department of Veterans Affairs.

† Race was reported by the participants.

linked to corresponding patient blood specimens. Data elements were abstracted from the EMR, including demographic information, coded International Classification of Diseases 10th Revision (ICD-10) diagnoses, medications, vital sign measurements, clinical laboratory test results (e.g., chemistry laboratory testing results, lactic acid measurement), sepsis-related laboratory measurements (C-reactive protein [CRP] and procalcitonin [PCT] were tested at an external lab; see the Supplementary Appendix, C-Reactive Protein and Procalcitonin Measurements section, for details), secondary outcomes metrics, relevant data to conduct adjudication (e.g., microbiology results), and relevant orders (e.g., antibiotic administration). Comorbidities were based on the components of the Charlson Comorbidity Index and were encoded based on ICD-10–Clinical Modification (CM) encodings defined by the National Cancer Institute Comorbidity Index and Surveillance, Epidemiology, and End Results program.¹¹ Immunocompromised patients were identified based on ICD-10-CM encodings defined by the Agency for Healthcare Research and Quality.¹²

SEPSIS IMMUNOSCORE

Algorithm Development

The Sepsis ImmunoScore ML algorithm was developed using a supervised, calibrated random forest model that predicts the probability of a patient meeting Sepsis-3 criteria within 24 hours of study entry. The random forest model was trained on the 2366 patient encounters in the derivation cohort using 22 patient-specific features comprising demographics, vital signs, and laboratory tests measured close to study entry. Model parameters were optimized using three repeats of fivefold cross-validation, and missing data were imputed using bagged trees.

Predictions were calibrated to the probability of Sepsis-3 by regressing the outcome on the out-of-bag predictions of the random forest in the derivation cohort to compute a sepsis risk score.¹³ Sepsis risk scores were divided into four risk stratification categories by thresholds identified during the development process using out-of-bag predictions in the derivation cohort (see Supplementary Appendix, Risk Category Development section). Interventional SHapley Additive exPlanation (SHAP) values were calculated for each patient using the out-of-bag measurements to assess feature importance in the random forest model.¹⁴⁻¹⁷

RISK SCORE AND RISK CATEGORY GENERATION

To assess performance, the Sepsis ImmunoScore was calculated for patients in the internal and external validation

cohorts. Calibrated out-of-bag scores were used for the derivation cohort to reduce bias from overfitting in performance estimation. No result was generated for patients lacking a measurement for PCT, CRP, white blood cells, platelets, creatinine, or blood urea nitrogen in the time between 24 hours prior to study entry (blood culture order) and 3.5 hours after. Similarly, no result was generated for patients without a measurement for systolic blood pressure, diastolic blood pressure, inspired oxygen percentage, heart rate, or respiratory rate in the time between 6 hours prior to study entry and 3.5 hours after. Missing values for the remaining 10 input parameters were imputed by the Sepsis ImmunoScore to produce a sepsis risk score.

STATISTICAL ANALYSIS

Diagnostic accuracy was assessed by determining the ability of the Sepsis ImmunoScore and its corresponding risk stratification category (low, medium, high, or very high), to identify patients with the primary outcome of sepsis (Sepsis-3 within 24 hours of study entry) and secondary outcomes. We estimated stratum-specific likelihood ratios (SSLRs) and predictive values (PVs) for each of the risk categories and assessed for a monotonic increasing relationship between risk category severity and outcomes using a one-sided Cochran–Armitage hypothesis test.¹⁸⁻²⁰ We also estimated the area under the receiver operating characteristic curve (AUROC) of the sepsis risk score. We report all uncertainty intervals using 95% confidence intervals. Confidence intervals for the AUROC were estimated using a binormal approximation estimator for the standard error.²¹ Analyses were conducted using R statistical software version 4.2.1.

SEPSIS-3 DIAGNOSTIC AND PROGNOSTIC SENSITIVITY ANALYSIS

Our a priori primary outcome was the presence of sepsis within 24 hours of study inclusion using the Sepsis-3 criteria as described above. This included both patients who met the criteria at initial evaluation and those who developed sepsis over the subsequent 24 hours. This decision (made in consultation with the FDA) was driven by the rationale that treatment is similar for both groups (e.g., antibiotics and supportive care). We also performed a sensitivity analysis to assess the ability of the Sepsis ImmunoScore to discriminate each of these groups from those patients who never had nor developed sepsis within 24 hours. We assessed Sepsis ImmunoScore performance separately for *diagnosis* of Sepsis, defined as Sepsis-3 criteria present at initial evaluation, and *prognosis* of sepsis, defined as the development

of Sepsis-3 criteria within 24 hours in those who did not have sepsis at presentation. The area under the curve (AUC) and SSLRs were reported for these analyses.

SAMPLE SIZE CALCULATION

This study was powered based on the confidence interval of the AUROC for the sepsis end point.²¹ The calculation assumed a sepsis prevalence of 32%, an estimated AUROC of 0.75, a maximum allowable difference between the true AUC and its estimate of 0.023, and a significance level of 0.05, resulting in an estimated sample size of 735 subjects. Additional participants were enrolled beyond these calculations to include participants of varying ages, racial backgrounds, ethnicities, and geographic locations. The initial study design used a single validation cohort partially enrolled from hospitals included in the derivation set. However, based on direction from the FDA, we restructured the cohort into the current internal and external validation format.

Results

There were a total of 3457 patient encounters included with valid Sepsis ImmunoScore results, with 2366 encounters in the derivation cohort, 393 in the internal validation cohort, and 698 in the external validation cohort (Fig. S1). The study enrolled participants that matched the age, sex, race, ethnicity, and comorbidities typical of sepsis patients in the United States (Table 1). The rate of sepsis was 32% in the derivation cohort, 28% in the internal validation cohort, and 22% in the external validation cohort (Table 1). Patients with sepsis had higher rates of severe illness and death than those without sepsis (Table 1).

SEPSIS IMMUNOSCORE ALGORITHM DEVELOPMENT

The Sepsis ImmunoScore algorithm analyzes up to 22 input parameters to generate a risk score and place patients in one of four discrete risk stratification categories (Table 2). The parameters consist of demographic data, vital sign

Parameter	Data Source	Measurement Validity†	Multiple Measurement Selection‡	Mandatory Parameter
Age	—	1 year	Most recent	
Systolic blood pressure	Vitals	6 hours	Min	Yes
Diastolic blood pressure	Vitals	6 hours	Min	Yes
Temperature	Vitals	6 hours	Max	Yes
Respiratory rate	Vitals	6 hours	Max	Yes
Heart rate	Vitals	6 hours	Max	Yes
Blood oxygen saturation	Vitals	6 hours	Min	Yes
White blood cell count	CBC panel	24 hours	Most recent	Yes
Lymphocyte count	CBC panel	24 hours	Most recent	
Neutrophil count	CBC panel	24 hours	Most recent	
Platelet count	CBC panel	24 hours	Most recent	Yes
Blood urea nitrogen	BMP or CMP	24 hours	Most recent	Yes
Creatinine	BMP or CMP	24 hours	Most recent	Yes
Potassium	BMP or CMP	24 hours	Most recent	
Chloride	BMP or CMP	24 hours	Most recent	
Total carbon dioxide	BMP or CMP	24 hours	Most recent	
Sodium	BMP or CMP	24 hours	Most recent	
Albumin	CMP	24 hours	Most recent	
Bilirubin	CMP	24 hours	Most recent	
Procalcitonin	Stand-alone test	24 hours	Most recent	Yes
C-reactive protein	Stand-alone test	24 hours	Most recent	Yes
Lactate	Stand-alone test	24 hours	Most recent	

* BMP denotes basic metabolic panel; CBC, complete blood count; CMP, comprehensive metabolic panel; max, maximum; and min, minimum.

† Corresponds to the look-back period prior to the Sepsis ImmunoScore result time when a measurement is valid.

‡ Details on how a single measurement was selected if multiple valid measurements were identified.

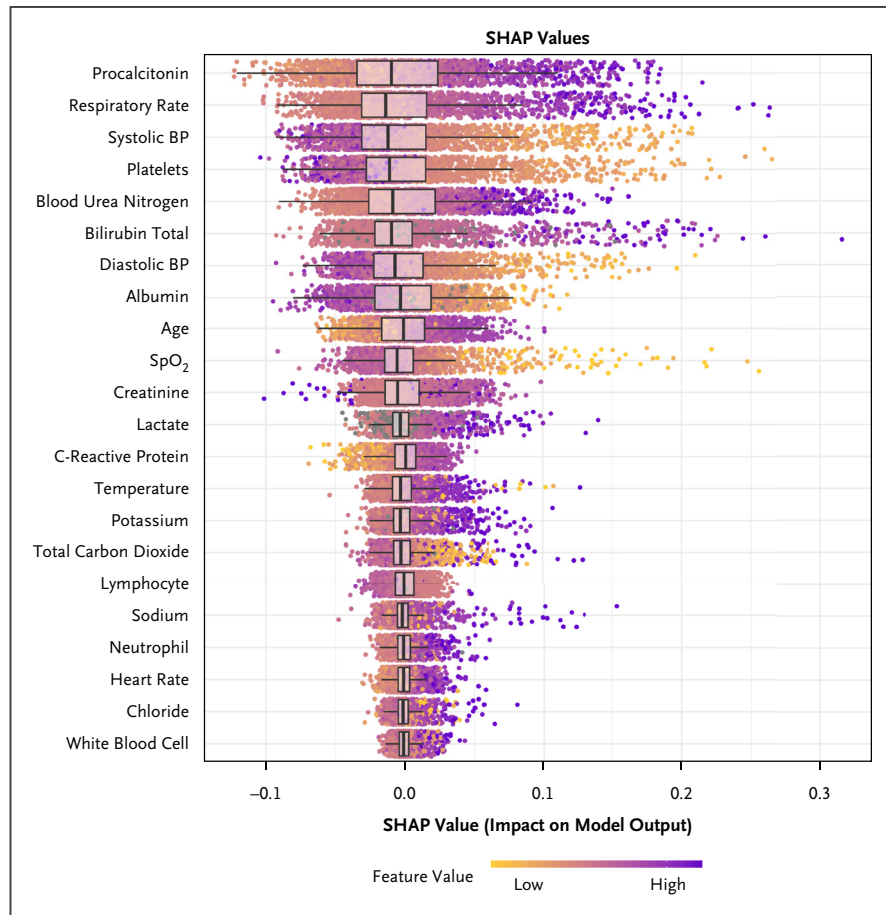


Figure 1. Sepsis ImmunoScore Feature Importance.

Scatter plots of the interventional SHapley Additive exPlanations (SHAP) values (a measure of feature contribution to the algorithm’s output) for each input feature across all patients in the derivation cohort. The features are listed on the y-axis in descending order of importance with the features contributing the most at the top of the list, as determined by the mean absolute interventional SHAP value across patients. Each data point is colored according to its observed standardized measurement value: purple indicates elevated measurements, while yellow indicates lower measurements. The individual values colored in purple offer the most impactful contributions (e.g., high procalcitonin and respiratory rate values, and low systolic blood pressure and platelet values all offered substantial contributions). Gray data points represent parameters that were imputed during the generation of the Sepsis ImmunoScore result. BP denotes blood pressure; SHAP, Shapley Additive exPlanations; and SpO₂, blood oxygen saturation.

measurements, comprehensive metabolic panel measurements, complete blood count panel measurements, lactate levels, and sepsis biomarkers PCT and CRP. The interventional SHAP values indicated that the three most influential parameters of the model were PCT, respiratory rate, and systolic blood pressure (Fig. 1). The AUC in the derivation set was 0.85 (95% confidence interval: 0.83 to 0.87) for the medical record-derived sepsis outcome (Table S1). Additionally, the Sepsis ImmunoScore risk categories were associated with increasing risk of sepsis in the derivation set (Fig. 1, Table S3).

PRIMARY END POINT

The Sepsis ImmunoScore demonstrated high overall diagnostic accuracy for predicting sepsis, with an AUC in the derivation set of 0.85 (95% confidence interval: 0.83 to 0.87) for the medical record-derived sepsis outcome, and 0.80 (0.74 to 0.86) in the internal validation and 0.81 (0.77 to 0.86) in the external validation for the adjudicated sepsis outcome (Table S1). The Sepsis ImmunoScore risk categories were associated with increasing risk of sepsis in both validation sets (Fig. 2, Table 3, Table S2). Of note, in the external validation set, the likelihood ratios were low,

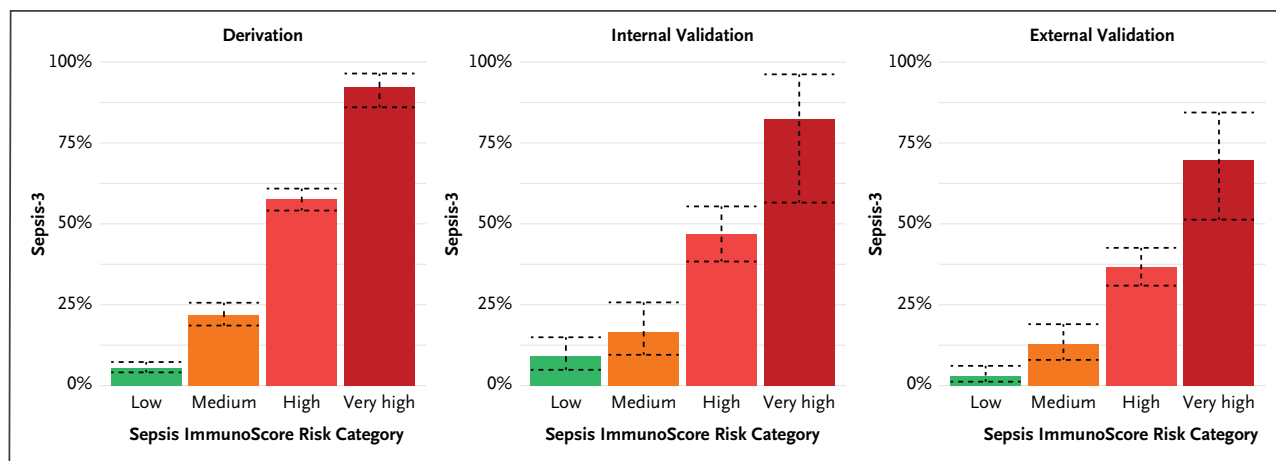


Figure 2. Sepsis ImmunoScore Stratification for Sepsis-3 in All Cohorts.

Bar charts are shown for the derivation, internal validation, and external validation cohort datasets for the Sepsis-3 criteria within 24 hours predictive values for each Sepsis ImmunoScore risk stratification category. Dashed lines indicate the 95% confidence intervals.

0.1 (0.1 to 0.2); medium, 0.5 (0.3 to 0.8); high, 2.1 (1.8 to 2.5); and very high, 8.3 (4.1 to 17.1) for Sepsis-3. These ratios are monotonically increasing with no overlapping confidence intervals, suggesting stepwise risk discrimination for sepsis.

SECONDARY END POINTS

We assessed the ability of the Sepsis ImmunoScore to predict the secondary outcomes of in-hospital mortality, length of stay, ICU admission within 24 hours, use of mechanical ventilation within 24 hours, and use of vasopressors within 24 hours. The Sepsis ImmunoScore was highly predictive of these outcomes. The Sepsis ImmunoScore categories low, medium, high, and very high demonstrated good predictive ability based on both rate of outcome as well as the corresponding SSLRs (Fig. 3, Table 4, Table S3). In the external validation cohort, the observed in-hospital

mortality rates in the low, medium, high, and very high risk groups were 0.0% (0.0 to 1.6%), 1.9% (0.40 to 5.5%), 8.7% (5.7 to 12.7%), and 18.2% (7.0 to 35.5%), respectively. Additionally, the observed median number of days for the composite length of stay end point in the low, medium, high, and very high risk groups were 4.0 (3.5 to 4.9), 5.7 (4.9 to 7.0), 7.7 (6.5 to 8.5), and 13.5 (7.1 to 19.1), respectively. The proportion of patients transferred to the ICU within 24 hours was 4.7% (2.4 to 8.3%), 12.7% (8.0 to 19.0%), 25.7% (20.7 to 31.3%), and 54.6% (36.4 to 71.9%), respectively. Similar trends were observed for mechanical ventilation and vasopressor usage. Cochran–Armitage hypothesis tests indicated statistically significant monotonic increasing relationships between outcome PV and risk stratification category severity for each secondary end point ($P < 0.01$, Table 4, Table S3). Risk stratification category severity was also associated with time to event for each secondary end point (Fig. S2).

Cohort	Patients (no.)	Patients with Sepsis (no.)	Sepsis PV (% [95% CI])	Sepsis Likelihood Ratio [95% CI]	Cochran–Armitage Test (P Value)
External validation (n=698)					
Low	232	7	3.0% [1.2%, 6.1%]	0.1 [0.1, 0.2]	<0.001
Medium	157	20	12.7% [7.96%, 19.0%]	0.5 [0.3, 0.8]	
High	276	101	36.6% [30.1%, 42.6%]	2.1 [1.8, 2.5]	
Very high	33	23	69.7% [51.3%, 84.4%]	8.3 [4.1, 17.1]	

*CI denotes confidence interval, and PV, predictive value.

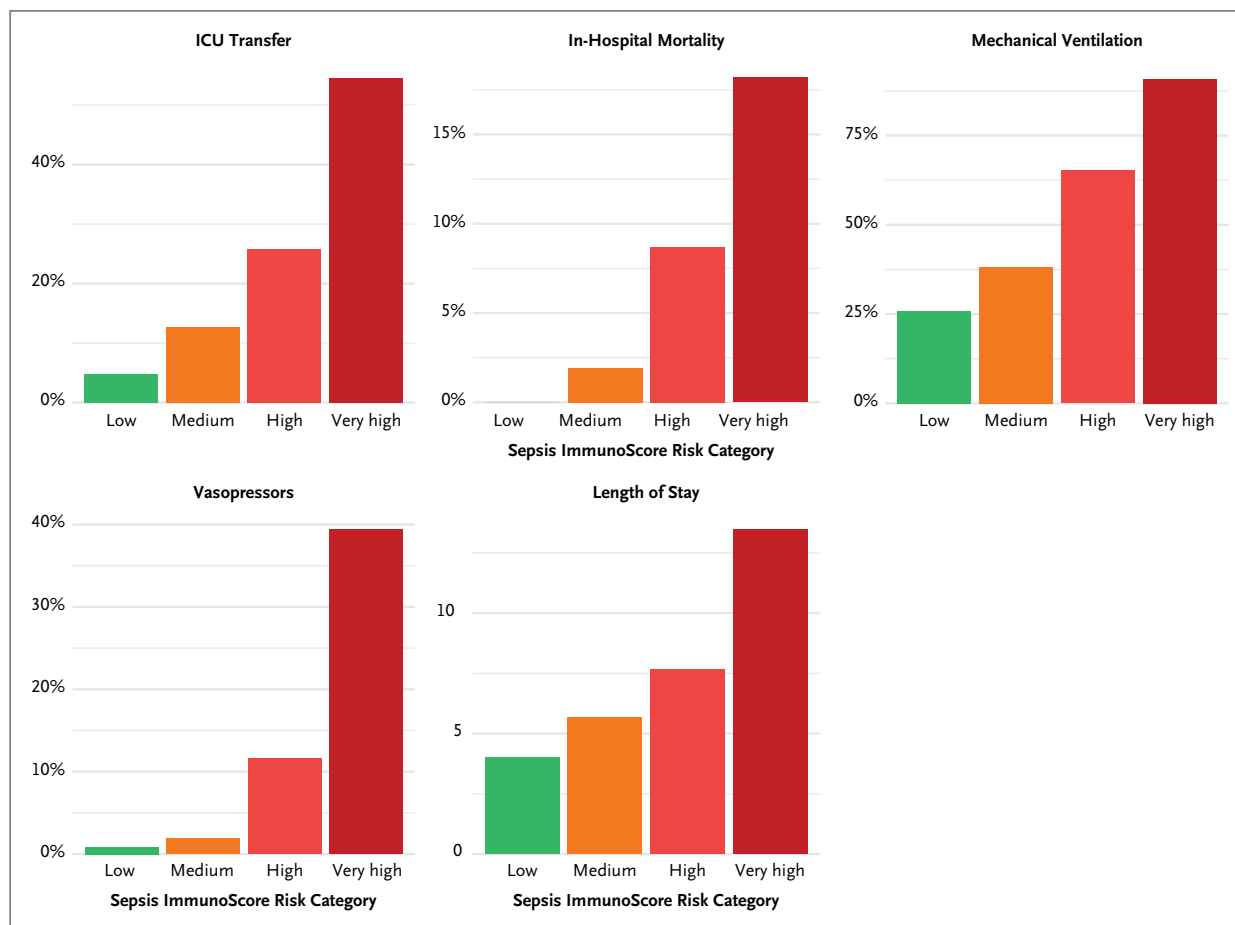


Figure 3. Sepsis ImmunoScore Risk Stratification for Morbidity and Mortality (External Validation).

Bar charts are shown for the external validation dataset for the secondary end point predictive values (intensive care unit transfer within 24-hours, in-hospital mortality, mechanical ventilation within 24-hours, vasopressor administration within 24-hours, and length of stay from inclusion time) for each Sepsis ImmunoScore risk stratification category. ICU denotes intensive care unit.

SEPSIS-3 DIAGNOSTIC AND PROGNOSTIC SENSITIVITY ANALYSIS

As a secondary sensitivity analysis, we also analyzed the performance of the Sepsis ImmunoScore using alternative definitions of the sepsis outcome based on the timing of sepsis. Namely, to assess *diagnostic performance*, we analyzed the Sepsis ImmunoScore's ability to discriminate patients who had sepsis at initial evaluation (n=99; 15.3%) from those who did not have sepsis within 24 hours (n=547, 84.7%). We also assessed *prognostic performance* by analyzing the Sepsis ImmunoScore's ability to discriminate patients without sepsis at initial evaluation who later developed sepsis within 24 hours (n=52; 7.4%) from those who never had sepsis (n=547, 92.6%). The AUC for sepsis diagnostic performance in the external validation set was 0.84 (0.78 to 0.89), and for the prognostic performance,

it was 0.76 (0.68 to 0.84). The Sepsis ImmunoScore risk categories demonstrated similar results to the original combined analysis for both the diagnosis (Table S4) and prognosis (Table S5) of sepsis. These data support good diagnostic and prognostic performance of the Sepsis ImmunoScore in this secondary analysis.

Discussion

The Sepsis ImmunoScore is a comprehensive, multidimensional AI/ML tool that combines demographics, vital signs, clinical laboratory tests, and sepsis-related laboratory tests to assess risk of sepsis and risk of adverse outcomes. In this study, we developed the Sepsis ImmunoScore and evaluated its ability to serve as a risk-stratification tool for

patients with suspected infection and to predict the diagnosis of sepsis and adverse clinical outcomes. We found the Sepsis ImmunoScore highly predictive of sepsis and the secondary outcomes of in-hospital mortality, length of hospital stay, ICU admission, mechanical ventilation, and vasopressor administration within 24 hours. We also found that the Sepsis ImmunoScore was predictive in diagnostic (predicting if a patient has sepsis at initial evaluation) and prognostic (predicting if a patient without sepsis at initial evaluation will develop sepsis within 24 hours) approaches.

Several FDA-approved tools for infection diagnosis are available, but they are often limited to detecting one or more blood biomarkers. PCT is a biomarker that evaluates the risk of progression to severe sepsis and septic shock in critically ill patients upon their first day in the ICU.²²⁻²⁶ The IntelliSep test, developed by Cytovale, is a blood test that measures leukocyte biophysical properties to create a score that identifies sepsis with organ dysfunction manifesting within the first 3 days after testing for adult

patients with signs and symptoms of infection who present to the emergency department.²⁷⁻²⁹ Another test, the Early Sepsis Indicator, developed by Beckman Coulter, measures monocyte distribution width to identify sepsis risk.³⁰⁻³² Other tests distinguish bacterial from nonbacterial infection in emergency department or urgent care settings, such as FebriDx, which measures myxovirus resistance protein A and CRP from finger-stick blood.³³⁻³⁶ The MeMed BV measures blood concentrations of TNF-related apoptosis-inducing ligand, interferon gamma-induced protein 10, and CRP to also distinguish patients with bacterial infections from those without.³⁷⁻⁴⁰ In contrast, the Sepsis ImmunoScore uses multidimensional inputs across several domains (demographics, vital signs, laboratory tests, etc.) plus sepsis biomarkers to create a comprehensive risk score for a given individual.

The Sepsis ImmunoScore is intended for EMR integration, enabling it to pull the requisite inputs and display the risk score when ordered as a diagnostic test. It is intended to

Table 4. External Validation Sepsis ImmunoScore Risk Stratification for Morbidity and Mortality.*

Secondary Outcome Sepsis Risk Category	Total Patients	Patients with Event	Predictive Value [95% CI]	Likelihood Ratio [95% CI]	Days (no.) [95% CI]	Cochran-Armitage Test (P Value)
ICU transfer within 24 hours						
Low	232	11	4.7% [2.4%, 8.3%]	0.2 [0.1, 0.4]	—	<0.001
Medium	157	20	12.7% [8.0%, 19.0%]	0.7 [0.4, 1.1]	—	
High	276	71	25.7% [20.7%, 31.3%]	1.7 [1.3, 2.1]	—	
Very high	33	18	54.6% [36.4%, 71.9%]	5.8 [3.0, 11.3]	—	
In-hospital mortality						
Low	232	0	0.0% [0.0%, 1.6%]	0.0 [0.0, 0.0]	—	<0.001
Medium	157	3	1.9% [0.4%, 5.5%]	0.4 [0.1, 1.2]	—	
High	276	24	8.7% [5.7%, 12.7%]	1.9 [1.3, 2.9]	—	
Very high	33	6	18.2% [7.0%, 35.5%]	4.5 [1.9, 10.7]	—	
Mechanical ventilation within 24 hours						
Low	232	6	2.6% [1.0%, 5.5%]	0.5 [0.2, 1.2]	—	0.008
Medium	157	6	3.8% [1.4%, 8.1%]	0.8 [0.4, 1.8]	—	
High	276	18	6.5% [3.9%, 10.1%]	1.4 [0.9, 2.2]	—	
Very high	33	3	9.1% [1.9%, 24.3%]	2.0 [0.6, 6.6]	—	
Vasopressor within 24 hours						
Low	232	2	0.9% [0.1%, 3.1%]	0.1 [0.0, 0.5]	—	<0.001
Medium	157	3	1.9% [0.4%, 5.5%]	0.3 [0.1, 0.8]	—	
High	276	32	11.6% [8.1%, 16.0%]	1.7 [1.2, 2.4]	—	
Very high	33	13	39.4% [22.9%, 57.9%]	8.4 [4.2, 16.7]	—	
Length of stay						
Low	232	232	—	—	4.0 [3.5, 4.9]	
Medium	157	157	—	—	5.7 [4.9, 7.0]	
High	276	276	—	—	7.7 [6.5, 8.5]	
Very high	33	33	—	—	13.5 [7.1, 19.1]	

*CI denotes confidence interval, and ICU, intensive care unit.

serve as an adjunct to clinical decision-making, utilizing a Bayesian approach that allows clinicians to combine the Sepsis ImmunoScore result with clinical assessment and traditional testing to inform their decisions. Rather than relying on a single cutoff to classify results as “normal” or “abnormal,” the Sepsis ImmunoScore reports four risk bands to capture a more similar test performance over a narrower range of values. For instance, approximately one third of patients test in the low-risk band, which has a likelihood ratio of 0.1, a sepsis prevalence of 3%, and a sensitivity of 95%. This low-risk band can help “rule out” sepsis in a patient with low to moderate clinical risk, assisting decisions for outpatient management. Conversely, for patients with moderate to high clinical risk, a high or very high Sepsis ImmunoScore is helpful from a Bayesian approach as a “rule-in” test. We also note that the Sepsis ImmunoScore bands are associated with the use of critical interventions such as mechanical ventilation and vasopressor use, as well as mortality. Thus, this may also inform clinicians when making clinical decisions such as patient disposition. When the clinical assessment is disparate from the Sepsis ImmunoScore, additional observations, assessments, and testing may be warranted.

While no other AI/ML tools are FDA-authorized for sepsis, many have been developed and clinically deployed, especially early detection tools that passively monitor patient data and alert clinicians when sepsis is suspected. The reported performance of these tools varies widely, and recent validation studies have raised concerns about their use.⁴¹⁻⁴⁴ A large external validation study of the widely deployed Epic Sepsis Model in 2021 reported an AUC of only 0.63,⁴¹ and recent reviews of validation studies of the Targeted Real-Time Early Warning System score have raised concerns regarding the control group and false positives.⁴² Concerns of alert fatigue have also been raised for these systems, which may undermine their clinical utility.⁴⁵⁻⁴⁷ The Sepsis ImmunoScore differs from early warning systems in that it is intended to be coupled with a clinical suspicion of infection (e.g., ordering of a blood culture) as opposed to implementation as a screening tool without specific context. However, misclassification in the clinical setting is still possible.

The application of AI/ML in medicine holds significant potential, much of which remains underdeveloped. The Sepsis ImmunoScore uses clinically available data reflective of patient biologic state and ML to identify objective patient assessments that are causally related to sepsis and associated adverse outcomes. Input features were carefully curated to select for measures of patient biology and

pathophysiology underlying critical illness, all of which are routinely collected or available in the setting of infection.⁴⁸ We did not include as eligible covariates subjective determinations or interventions that could be heavily influenced by site-specific protocols, clinician-specific perspectives, or other peculiarities of care. In addition to accurately diagnosing sepsis in an external validation set, we attribute the simultaneous association of the Sepsis ImmunoScore with other adverse outcomes in part to an explicit focus on patient host response biology. The result of this careful synthesis is a diagnostic tool that leverages the synergy between thoughtfully applied AI/ML and expertly curated biologic data to better equip — not replace — clinicians in their fight against sepsis.

Furthermore, while the current FDA authorization does not allow local calibration or model adjustment, future efforts may consider leveraging AI/ML algorithms’ ability to locally calibrate, pending changes to FDA authorization, to improve performance.

Sepsis represents an ongoing diagnostic challenge to clinicians due to its often subtle and heterogeneous presentation. Assessing the presence or risk of progression to sepsis and severity with associated clinical needs represents a continuing challenge for clinicians. The Sepsis ImmunoScore is unique in its approach due to its ML-based incorporation of 22 parameters to comprehensively assess a patient’s risk of being diagnosed with sepsis, plus its association with adverse outcomes. The Sepsis ImmunoScore could serve as an adjunctive test to assist clinical decision-making in the acute setting. Given its strong predictive ability, the Sepsis ImmunoScore has the potential to improve patient outcomes by informing physician decisions for patients potentially requiring sepsis-related care, such as the rapid administration of broad-spectrum antimicrobials, escalation of care, and administration of fluid or vasopressor medications. It also has the potential to reduce over-triage by more accurately identifying patients at low risk for deterioration due to infection, potentially enabling emergency department physicians to treat these low-risk patients in the outpatient setting.

LIMITATIONS

There are several limitations to our study. First, because we used five hospitals in the study, it is possible that our findings may not generalize to specific populations that differ from these hospitals. Second, we relied upon an EMR extraction, so it is possible that missing data or the use of ICD-10 codes may have led to misclassification of certain elements such

as comorbidities. Third, this was an observational study and did not assess the impact of the Sepsis ImmunoScore on clinical decision-making or changes in therapeutic approaches. Fourth, the primary outcome of Sepsis-3 within 24 hours relied on an automated calculation in the derivation set and adjudication for presence of infection in the internal and external validation; thus, misclassification of outcomes may have occurred. Fifth, our inclusion criteria used the ordering of a blood culture as a surrogate indicator for a clinical suspicion of infection. However, some patients with suspected infections may not have had a blood culture ordered, while others may have had a blood culture performed despite having little or no suspicion of infection. Sixth, we note that approximately 7.2% of the potentially eligible patients were excluded due to not having lab results or vital signs available to calculate the Sepsis ImmunoScore. Since patients without the requisite inputs had a higher prevalence of being from the in-hospital or ICU setting (Table S6), it is possible this could mildly (given the low prevalence) affect the generalizability of the score. Finally, absent covariate data may have affected algorithm performance.

The Sepsis ImmunoScore has demonstrated robust risk assessment performance in derivation, internal, and external validation. Future work is warranted to further establish its generalizability to other settings. Finally, additional studies are warranted to assess the impact of the Sepsis ImmunoScore on clinical decision-making, sepsis care, and associated resource utilization and costs. These investigations are ongoing.

Disclosures

Author disclosures are available at ai.nejm.org.

Supported in part by funding from the Defense Threat Reduction Agency, the National Institutes of Health, the Centers for Disease Control and Prevention, the National Science Foundation, the Biomedical Advanced Research and Development Authority, and Prenosis.

We are indebted to the study coordinators, research staff, and lab technicians who participated in the study. These contributions were part of these individuals' jobs, and they did not receive additional compensation.

Author Affiliations

¹Prenosis, Chicago

²Emory School of Medicine, Atlanta

³Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston

⁴Harvard Medical School, Harvard University, Boston

⁵William Beaumont University Hospital, Corewell Health, Royal Oak, MI

⁶Mercy Hospital, St. Louis

⁷Mercy Hospital Jefferson, Festus, Missouri

⁸OSF Saint Francis Medical Center, Peoria, Illinois

⁹Davis School of Medicine, University of California, Sacramento

¹⁰Jesse Brown Department of Veterans Affairs Medical Center, Chicago

¹¹Sinai Hospital of Baltimore, LifeBridge Health, Baltimore

¹²Department of Statistics, University of Illinois at Urbana-Champaign, Urbana

¹³Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana

The authors' full names and academic degrees are as follows: Akhil Bhargava, M.S., Carlos López-Espina, M.S., Lee Schmalz, B.S., Shah Khan, Ph.D., Gregory L. Watson, Ph.D., Denny Urdiales, B.S., Lincoln Updike, B.S., Niko Kurtzman, M.D., Alon Dagan, M.D., Amanda Doodlesack, M.D., Bryan A. Stenson, M.D., Deesha Sarma, M.D., Eric Reseland, M.D., John H. Lee, M.D., Ph.D., Max S. Kravitz, M.D., Peter S. Antkowiak, M.D., M.P.H., Tatyana Shvilkina, D.O., Aimee Espinosa, M.D., Alexandra Halalau, M.D., Carmen Demarco, M.D., Francisco Davila, M.D., Hugo Davila, M.D., Matthew Sims, M.D., Ph.D., Nicholas Maddens, M.D., Ramona Berghea, M.D., Scott Smith, M.D., Ashok V. Palagiri, M.D., Clinton Ezekiel, M.D., Farid Sadaka, M.D., Karthik Iyer, M.D., Matthew Crisp, M.D., Ph.D., Saleem Azad, D.O., Vikram Oke, M.D., Andrew Friederich, M.D., Anwaruddin Syed, M.D., Falgun Gosai, M.D., Lavneet Chawla, M.D., Neil Evans, M.D., Kurian Thomas, M.D., Roneil Malkani, M.D., Roshni Patel, M.D., Stockton Mayer, D.O., Farhan Ali, M.D., Lekshminarayan Raghavakurup, M.D., Muleta Tafa, M.D., M.P.H., Sahib Singh, M.D., Samuel Raouf, M.D., Sihai Dave Zhao, Ph.D., Ruoqing Zhu, Ph.D., Rashid Bashir, Ph.D., Bobby Reddy Jr, Ph.D., and Nathan I. Shapiro, M.D., M.P.H.

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;395:200–211. DOI: [10.1016/S0140-6736\(19\)32989-7](https://doi.org/10.1016/S0140-6736(19)32989-7).
2. Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. *Crit Care Med* 2014;42:1749–1755. DOI: [10.1097/CCM.0000000000000330](https://doi.org/10.1097/CCM.0000000000000330).
3. Gaieski DF, Mikkelsen ME, Band RA, et al. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med* 2010;38:1045–1053. DOI: [10.1097/CCM.0b013e3181cc4824](https://doi.org/10.1097/CCM.0b013e3181cc4824).
4. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006;34:1589–1596. DOI: [10.1097/01.CCM.0000217961.75225.E9](https://doi.org/10.1097/01.CCM.0000217961.75225.E9).
5. Puskarich MA, Trzeciak S, Shapiro NI, et al. Association between timing of antibiotic administration and mortality from septic shock in patients treated with a quantitative resuscitation protocol. *Crit Care Med* 2011;39:2066–2071. DOI: [10.1097/CCM.0b013e31821e87ab](https://doi.org/10.1097/CCM.0b013e31821e87ab).

6. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* 2021;47:1181-1247. DOI: [10.1007/s00134-021-06506-y](https://doi.org/10.1007/s00134-021-06506-y).
7. Liu VX, Fielding-Singh V, Greene JD, et al. The timing of early antibiotics and hospital mortality in sepsis. *Am J Respir Crit Care Med* 2017;196:856-863. DOI: [10.1164/rccm.201609-1848OC](https://doi.org/10.1164/rccm.201609-1848OC).
8. Filbin MR, Lynch J, Gillingham TD, et al. Presenting symptoms independently predict mortality in septic shock: importance of a previously unmeasured confounder. *Crit Care Med* 2018;46:1592-1599. DOI: [10.1097/CCM.0000000000003260](https://doi.org/10.1097/CCM.0000000000003260).
9. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016;315:801. DOI: [10.1001/jama.2016.0287](https://doi.org/10.1001/jama.2016.0287).
10. Rhee C, Dantes R, Epstein L, et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. *JAMA* 2017;318:1241. DOI: [10.1001/jama.2017.13836](https://doi.org/10.1001/jama.2017.13836).
11. National Cancer Institute. Table describing changes in codes for NCI Comorbidity Index. March 27, 2024 (<https://healthcaredelivery.cancer.gov/seermedicare/considerations/comorbidity-table.html>).
12. Agency for Healthcare Research and Quality. Prevention quality indicators — Appendix C: Immunocompromised state diagnosis and procedure codes. Rockville, MD: Agency for Healthcare Research and Quality, July 2016. August 31, 2023 (https://qualityindicators.ahrq.gov/Downloads/Modules/PQI/V60-ICD10/TechSpecs/PQI_Appendix_C.pdf).
13. Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv Large Margin Classif* 1999;10:61-74.
14. Shapley LS. A value for n-person games. Santa Monica, CA: RAND Corporation, 1952.
15. Janzing D, Minorics L, Bloebaum P. Feature relevance quantification in explainable AI: a causal problem. *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. PMLR; 2020:2907-2916. February 11, 2024 (<https://proceedings.mlr.press/v108/janzing20a.html>).
16. Sundararajan M, Najmi A. The many Shapley values for model explanation. *Proceedings of the 37th International Conference on Machine Learning*. PMLR, 2020:9269-9278. February 11, 2024 (<https://proceedings.mlr.press/v119/sundararajan20b.html>).
17. Chen H, Janizek JD, Lundberg S, Lee SI. True to the model or true to the data? June 29, 2020 (<http://arxiv.org/abs/2006.16234>). Preprint.
18. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005;365:1500-1505. DOI: [10.1016/S0140-6736\(05\)66422-7](https://doi.org/10.1016/S0140-6736(05)66422-7).
19. Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955;11:375-386. DOI: [10.2307/3001775](https://doi.org/10.2307/3001775).
20. Cochran WG. Some methods for strengthening the common χ^2 tests. *Biometrics* 1954;10:417-451. DOI: [10.2307/3001616](https://doi.org/10.2307/3001616).
21. Obuchowski NA. Sample size calculations in studies of test accuracy. *Stat Methods Med Res* 1998;7:371-392. DOI: [10.1177/096228029800700405](https://doi.org/10.1177/096228029800700405).
22. U.S. Food and Drug Administration Center for Devices and Radiological Health. 510(k) substantial equivalence determination decision summary, Roche Diagnostics K173927. April 17, 2024 (https://www.accessdata.fda.gov/cdrh_docs/reviews/K173927.pdf).
23. Bouadma L, Luyt C-E, Tubach F, et al. Use of procalcitonin to reduce patients' exposure to antibiotics in intensive care units (PRORATA trial): a multicentre randomised controlled trial. *Lancet* 2010;375:463-474. DOI: [10.1016/S0140-6736\(09\)61879-1](https://doi.org/10.1016/S0140-6736(09)61879-1).
24. Nobre V, Harbarth S, Graf JD, Rohner P, Pugin J. Use of procalcitonin to shorten antibiotic treatment duration in septic patients: a randomized trial. *Am J Respir Crit Care Med* 2008;177:498-505. DOI: [10.1164/rccm.200708-1238OC](https://doi.org/10.1164/rccm.200708-1238OC).
25. Hochreiter M, Köhler T, Schweiger AM, et al. Procalcitonin to guide duration of antibiotic therapy in intensive care patients: a randomized prospective controlled trial. *Crit Care* 2009;13:1-7. DOI: [10.1186/cc7903](https://doi.org/10.1186/cc7903).
26. Karlsson S, Heikkinen M, Pettilä V, et al. Predictive value of procalcitonin decrease in patients with severe sepsis: a prospective observational study. *Crit Care* 2010;14:1-10. DOI: [10.1186/cc9327.21078153](https://doi.org/10.1186/cc9327.21078153)
27. O'Neal HR Jr, Sheybani R, Janz DR, et al. Validation of a novel, rapid sepsis diagnostic for emergency department use. *Crit Care Explor* 2024;6:e1026. DOI: [10.1097/CCE.0000000000001026](https://doi.org/10.1097/CCE.0000000000001026).
28. U.S. Food and Drug Administration Center for Devices and Radiological Health. 510(k) substantial equivalence determination decision summary, Cytovale Inc. K220991. April 17, 2024 (https://www.accessdata.fda.gov/cdrh_docs/reviews/K220991.pdf).
29. Guillou L, Sheybani R, Jensen AE, et al. Development and validation of a cellular host response test as an early diagnostic for sepsis. *PLoS One* 2021;16:e0246980. DOI: [10.1371/journal.pone.0246980](https://doi.org/10.1371/journal.pone.0246980).
30. Crouser ED, Parrillo JE, Seymour CW, et al. Monocyte distribution width: a novel indicator of sepsis-2 and sepsis-3 in high-risk emergency department patients. *Crit Care Med* 2019;47:1018-1025. DOI: [10.1097/CCM.0000000000003799](https://doi.org/10.1097/CCM.0000000000003799).
31. Jo SJ, Kim SW, Choi JH, Choi SP, Lee J, Lim J. Monocyte distribution width (MDW) as a useful indicator for early screening of sepsis and discriminating false positive blood cultures. *PLoS One* 2022;17:e0279374.
32. U.S. Food and Drug Administration Center for Devices and Radiological Health. 510(k) substantial equivalence determination decision summary, Beckman Coulter, Inc. K181599. April 18, 2024 (https://www.accessdata.fda.gov/cdrh_docs/reviews/K181599.pdf).

33. U.S. Food and Drug Administration Center for Devices and Radiological Health. 510(k) substantial equivalence determination decision summary, FebriDx K230917. April 18, 2024 (https://www.accessdata.fda.gov/cdrh_docs/reviews/K230917.pdf).
34. Shapiro NI, Self WH, Rosen J, et al. A prospective, multi-centre US clinical trial to determine accuracy of FebriDx point-of-care testing for acute upper respiratory infections with and without a confirmed fever. *Ann Med* 2018;50:420-429. DOI: [10.1080/07853890.2018.1474002](https://doi.org/10.1080/07853890.2018.1474002).
35. Tong-Minh K, Daenen K, Endeman H, et al. Performance of the FebriDx rapid point-of-care test for differentiating bacterial and viral respiratory tract infections in patients with a suspected respiratory tract infection in the emergency department. *J Clin Med* 2023;13:163. DOI: [10.3390/jcm13010163](https://doi.org/10.3390/jcm13010163).
36. Shirley M. FebriDx®: a rapid diagnostic test for differentiating bacterial and viral aetiologies in acute respiratory infections. *Mol Diagn Ther* 2019;23:803-809. DOI: [10.1007/s40291-019-00433-x](https://doi.org/10.1007/s40291-019-00433-x).
37. U.S. Food and Drug Administration Center for Devices and Radiological Health. 510(k) substantial equivalence determination decision summary, MeMed Diagnostics Ltd. K210254. April 18, 2024 (https://www.accessdata.fda.gov/cdrh_docs/reviews/K210254.pdf).
38. Oved K, Cohen A, Boico O, et al. A novel host-proteome signature for distinguishing between acute bacterial and viral infections. *PLoS One* 2015;10:e0120012. DOI: [10.1371/journal.pone.0120012](https://doi.org/10.1371/journal.pone.0120012).
39. Srugo I, Klein A, Stein M, et al. Validation of a novel assay to distinguish bacterial and viral infections. *Pediatrics* 2017;140:e20163453. DOI: [10.1542/peds.2016-3453](https://doi.org/10.1542/peds.2016-3453).
40. Papan C, Argentiero A, Porwoll M, et al. A host signature based on TRAIL, IP-10, and CRP for reducing antibiotic overuse in children by differentiating bacterial from viral infections: a prospective, multicentre cohort study. *Clin Microbiol Infect* 2022;28:723-730. DOI: [10.1016/j.cmi.2021.10.019](https://doi.org/10.1016/j.cmi.2021.10.019).
41. Wong A, Otlés E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* 2021;181:1065-1070. DOI: [10.1001/jamainternmed.2021.2626](https://doi.org/10.1001/jamainternmed.2021.2626).
42. Schinkel M, Van Der Poll T, Wiersinga WJ. Artificial intelligence for early sepsis detection: a word of caution. *Am J Respir Crit Care Med* 2023;207:853-854. DOI: [10.1164/rccm.202212-2284VP](https://doi.org/10.1164/rccm.202212-2284VP).
43. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020;46:383-400. DOI: [10.1007/s00134-019-05872-y](https://doi.org/10.1007/s00134-019-05872-y).
44. Patton MJ, Liu VX, Might M. Temporal analysis of biomarker, severity-of-illness score, and machine learning model performance for sepsis prediction. *42nd International Symposium on Intensive Care & Emergency Medicine, ISICEM 2023*, Brussels, March 21-24, 2023.
45. The Lancet Respiratory Medicine. Crying wolf: the growing fatigue around sepsis alerts. *Lancet Respir Med* 2018;6:161. DOI: [10.1016/S2213-2600\(18\)30072-9](https://doi.org/10.1016/S2213-2600(18)30072-9).
46. Makam AN, Nguyen OK, Auerbach AD. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: a systematic review. *J Hosp Med* 2015;10:396-402. DOI: [10.1002/jhm.2347](https://doi.org/10.1002/jhm.2347).
47. Hwang MI, Bond WF, Powell ES. Sepsis alerts in emergency departments: a systematic review of accuracy and quality measure impact. *West J Emerg Med* 2020;21:1201. DOI: [10.5811/westjem.2020.5.46010](https://doi.org/10.5811/westjem.2020.5.46010).
48. Maslove DM, Tang B, Shankar-Hari M, et al. Redefining critical illness. *Nat Med* 2022;28:1141-1148. DOI: [10.1038/s41591-022-01843-x](https://doi.org/10.1038/s41591-022-01843-x).