Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Additive Predictive Value of the Sepsis ImmunoScore ® : A Multicenter, Prospective Evaluation of AI-Enabled Prediction of Sepsis Risk and Patient Outcomes

Akhil Bhargava

akhil.bhargava2016@gmail.com

Prenosis (United States)

Carlos López-Espina

Prenosis (United States)

Lee Schmalz

Prenosis (United States)

Shah Khan

Prenosis (United States)

Gregory L. Watson

Prenosis (United States)

Robin Carver

Prenosis (United States)

Grace Staples

Prenosis (United States)

Lincoln Updike

Prenosis (United States)

Dennys Urdiales

Prenosis (United States)

Rashid Bashir

University of Illinois Urbana-Champaign

Bobby Reddy, Jr.

Prenosis (United States)

Nathan I. Shapiro

Beth Israel Deaconess Medical Center

Article

Keywords: sepsis, artificial intelligence, clinical decision support, biomarkers, emergency care, intensive care, triage

Posted Date: June 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10168107/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. AB, CLE, LS, SK, GLW, RC, GS, LU, DU, RB, BRJ, and NS disclose employment and/or financial interests in Prenosis, Inc.

Abstract

Background

Early identification of sepsis remains challenging in emergency and acute care settings, where nonspecific symptoms can complicate assessment of a patient's risk of deterioration. We evaluated whether the Sepsis ImmunoScore®, an FDA-authorized artificial intelligence/machine learning tool integrating biomarkers and clinical data, could provide additive value beyond provider judgment in assessing risk of sepsis and need for critical care utilization.

Methods

We conducted a multicenter, prospective observational study across three U.S. health systems, enrolling adults with suspected infection at the time of blood culture, from April 2024–August 2025. Clinicians provided survey-based assessments of infection, Sepsis-3 risk, and risk of sepsis-related organ support and were blinded to Sepsis ImmunoScore results. The Sepsis ImmunoScore was calculated retrospectively using patient data from the electronic health record (EHR) and remnant blood samples. The primary outcome was Sepsis-3 within 24 hours. Discrimination was assessed using area under the receiver operating characteristic curve (AUROC), with additive value tested using likelihood ratio tests (LRT) and continuous net reclassification improvement (NRI). Risk stratification was examined through cross-classification of provider and Sepsis ImmunoScore categories.

Results

Among 401 patients with valid Sepsis ImmunoScore results, 103 (25.7%) met Sepsis-3 criteria. For sepsis risk prediction, the combined Provider + Sepsis ImmunoScore model achieved an AUROC of 0.83, versus 0.68 for Providers Alone (ΔAUC = + 0.15; $p < 0.001$). Sepsis ImmunoScore added independent predictive value by LRT ($\chi^2 = 86.44$; $p < 0.001$) and significantly improved reclassification (NRI = 0.98). Similar additive value was observed for predicting ICU transfer, vasopressor use, critical care, and mortality. Within every provider-assessed risk stratum (Low, Medium, High/Very High), the Sepsis ImmunoScore revealed a monotonic gradient of risk; notably, 25.7% of patients assessed as Low risk by providers but High/Very High risk by Sepsis ImmunoScore had sepsis.

Conclusions

The Sepsis ImmunoScore consistently enhances clinician assessment by improving discrimination and uncovering occult risk within provider-defined categories. These findings support its potential as an EHR-integrated decision-support tool that may improve early sepsis identification and risk stratification in emergent and acute care settings.

Background

Sepsis remains a leading cause of morbidity, mortality, and healthcare utilization worldwide,(1) with outcomes highly dependent on timely recognition and appropriate early management.(2) Early clinical signs of evolving sepsis are often subtle or nonspecific, particularly at the point of initial evaluation in emergency departments (ED) and acute care settings.(3) Diagnostic uncertainty complicates decisions regarding antimicrobial initiation, level of monitoring, and intensive care unit (ICU) disposition—choices that carry substantial consequences for both patient outcomes and resource allocation.

In response to the challenges of early sepsis identification, numerous electronic health record (EHR)-based decision-support tools and automated alerts have been developed.(4, 5) However, most existing systems rely on rule-based triggers derived from vital signs or laboratory abnormalities.(6) These tools often generate high rates of false alerts without meaningfully improving accuracy, resulting in clinician override, alert fatigue, workflow burden, and limited impact on clinical outcomes.(7) While artificial intelligence (AI) and machine learning (ML) approaches have begun to emerge, few have demonstrated real-world clinical utility—namely, the ability to meaningfully enhance decision-making rather than simply increase alert volume.(8, 9)

This study (funded via BARDA Contract 75A50123C00019) was performed to evaluate the additive predictive value of the Sepsis ImmunoScore®, a pathogen-agnostic, FDA-authorized AI/ML tool intended to aid risk assessment for presence of Sepsis-3 or progression to it within 24 hours of the time of blood culture order among ED or hospitalized patients.(10, 11) We hypothesized that Sepsis ImmunoScore could aid clinicians in the assessment of the presence of sepsis and the need for critical care utilization and mortality. The purpose of this study was therefore to evaluate the additive predictive value of Sepsis ImmunoScore in real-world emergency and acute care practice by determining whether Sepsis ImmunoScore can meaningfully enhance risk assessment and decision-making beyond provider judgment alone.

Methods

Study Design and Setting

We conducted a multicenter, prospective observational study across three U.S. health systems (Carle Foundation Hospital in Urbana, IL; Corewell Health in Royal Oak, MI; and Mercy Hospitals East Community in St. Louis, MO) to evaluate the additive predictive value of the Sepsis ImmunoScore. The study included adults (≥ 18 years) with suspected infection, defined by a clinical decision to obtain a first blood culture order (LOINC 600-7) in the ED (all sites) or ICU (Mercy only). Availability of remnant lithium-heparin plasma from blood samples drawn for routine care was required for inclusion. None of these patients were used in the development or training of the Sepsis ImmunoScore algorithm.

Prenosis software was integrated with each site's Epic EHR via Fast Healthcare Interoperability Resources and Health Level Seven Version 2 (FHIR and HL7 V2) pathways to continuously screen for

eligibility and enroll patients. Clinical, laboratory, and biomarker data were ingested in real time, enabling standardized, site-agnostic data capture. Upon identification and enrollment of eligible adult patients, the software automatically placed a Sepsis ImmunoScore order in the EHR that was calculated in the background, and providers were blinded to the results. This study did not involve an intervention or influence patient care; as such, each hospital's institutional review board (IRB) waived the need for informed consent on behalf of enrolled patients. Clinician participation was optional. Those who agreed to take part signed an informed consent form in advance of completing any surveys. Enrollment took place April 2024–August 2025. Because surveys were optional and were therefore completed for a subset of eligible patients, we compared included patients with concurrently enrolled emergency department patients who had a valid Sepsis ImmunoScore result but no completed survey, to assess the representativeness of the surveyed cohort with respect to baseline characteristics and Sepsis ImmunoScore discrimination.

Sepsis Assessment

Sepsis risk was evaluated near the time of blood culture both by clinicians and by the Sepsis ImmunoScore platform. Providers were notified by study staff when a patient was enrolled and asked to complete a standardized six-question survey within 3.5 hours of the blood culture order, clinical responsibilities permitting. The survey captured three domains as both probabilities (0–100%) and ordinal Likert categories (Low/Unlikely, Medium/Possible, High/Likely, Very High/Very Likely):

1. **Infection**: Risk of infection as the cause of their illness
2. **Sepsis**: Sepsis-3 risk (an increase of 2 or more Sequential Organ Failure Assessment [SOFA] points due to infection) at present or within 24 hours (12)
3. **Organ Support**: Risk of requiring sepsis-related organ support (continuous positive airway pressure/bilevel positive airway pressure [CPAP/BIPAP]/mechanical ventilation or vasopressor use, or new renal replacement therapy) within 24 hours

Categorical responses were treated as ordinal variables; continuous responses were analyzed as reported. To prevent duplication, surveys from repeat visits were excluded.

The Sepsis ImmunoScore algorithm, described elsewhere,(10) is a locked (i.e., not site-specific) algorithm that synthesizes up to 22 clinical variables and biomarkers, including C-reactive protein (CRP) and procalcitonin (PCT), to produce a sepsis risk score from 1–100. Scores correspond to pre-defined risk categories (Low, Medium, High, Very High) based on validated thresholds. The model is highly transparent and uses SHAP (Shapley Additive exPlanations) values, which characterize the direction and magnitude of each feature's contribution to predicted risk.(13, 14) Because CRP and PCT are required to generate the Sepsis ImmunoScore but infrequently obtained in routine practice, remnant blood samples were retrieved for enrolled patients. These were used for retrospective calculation of the Sepsis ImmunoScore if collected within 24 hours before to 3.5 hours after the blood culture order. Hospital staff centrifuged remnant samples, aliquoted into 175 µL tubes, stored at − 80°C, and shipped samples to the

Prenosis central laboratory, where CRP and PCT were measured using the Luminex Assay Platform. Baseline characteristics and outcomes for patients with and without valid Sepsis ImmunoScore results were compared to assess potential exclusion bias.

Endpoints

The primary endpoint was presence of Sepsis-3 at evaluation or within 24 hours of blood culture order. Sepsis-3 classification was operationalized using automated EHR data extraction and defined as ≥ 4 qualifying antimicrobial days or confirmed infection via pathogen testing, and a SOFA increase ≥ 2 points from baseline.(15)

Secondary outcomes included ICU admission within 24 hours, vasopressor initiation within 24 hours, assisted ventilation (BiPAP, CPAP or mechanical ventilation) placement within 24 hours, critical care interventions (assisted ventilation or vasopressor administration) within 24 hours, and mortality within 28 days.

Statistical Analysis

All analyses were performed in R version 4.2.3 (R Foundation for Statistical Computing). All statistical tests were two-tailed with $\alpha = 0.05$. Exact p-values are reported except where $p < 0.001$. For descriptive population statistics, continuous variables were assessed via the Shapiro-Wilk test and categorical variables are reported as n (%). Between-group comparisons used Wilcoxon rank-sum tests for continuous variables and chi-squared or Fisher's exact tests for categorical variables.

Discriminative performance analysis. Discriminative performance was assessed using area under the receiver operating characteristic curve (AUROC, also abbreviated AUC) for three models per endpoint: Provider Alone, Sepsis ImmunoScore Alone, and Provider + Sepsis ImmunoScore combined, with both predictors (i.e., the survey-based Sepsis Risk Probability and the Sepsis Risk Score) entered as covariates in logistic regression. AUROCs with 95% confidence intervals were estimated via DeLong's method (pROC v1.19.0.1). The primary endpoint (Sepsis-3 within 24 hours) was paired with provider-assessed probability of Sepsis-3, while Organ Support probability was paired with the secondary endpoints (ICU transfer, vasopressor use, assisted ventilation, critical care, and 28-day mortality). No multiplicity adjustment was applied to these six pre-specified comparisons.

Additive value and reclassification testing. The additive predictive value of the Sepsis ImmunoScore was tested using the likelihood ratio test (LRT; lmtest package), comparing nested logistic regression models with and without Sepsis ImmunoScore continuous score as a main effect (degrees of freedom [df] = 1).(16) The base model comprised the Provider Sepsis Risk Probability for the Sepsis-3 endpoint and the Provider Organ Support Probability for the secondary endpoints; in each case the full model added the Sepsis ImmunoScore continuous score. Continuous net reclassification improvement (NRI) was calculated using custom bootstrap functions (boot v1.3.28.1; B = 1,000) to quantify the extent to

which the combined model appropriately shifts predicted probabilities upward for events and downward for non-events relative to provider assessment alone.(17) Importantly, this analysis was observational only and did not influence clinical care.

Risk stratification analysis. To visualize the additive stratification value of the Sepsis ImmunoScore within levels of provider assessment, Sepsis-3 event rates were calculated within subgroups defined by the cross-classification of provider assessment category and Sepsis ImmunoScore risk category. For this analysis, both the provider and Sepsis ImmunoScore four-level scales were collapsed into three levels (Low, Medium, and High/Very High) to ensure sufficient sample sizes within each cell. For the primary endpoint (Sepsis-3 within 24h), all nine combinations of Provider level × Sepsis ImmunoScore level were evaluated. For secondary endpoints, where lower event counts limited cell sizes, the Sepsis ImmunoScore was further collapsed into two levels (Low/Medium vs High/Very High) while maintaining three provider levels, yielding six subgroups per endpoint. Event rates are presented as n/N (%).

Statistical significance of this categorical stratification was assessed using Cochran–Mantel–Haenszel (CMH) tests.(18) For the primary endpoint (Sepsis-3), a generalized CMH test evaluated the association between the three-level Sepsis ImmunoScore category (Low, Medium, High/Very High) and Sepsis-3, stratified by the three-level Provider Sepsis Risk assessment. The general-association statistic and, treating the Sepsis ImmunoScore as ordinal, the Mantel correlation (linear-trend) statistic are reported. For the secondary endpoints, where lower event counts limited stratum cell sizes, the Sepsis ImmunoScore was dichotomized (Low/Medium vs High/Very High) and a 2×2×K Mantel–Haenszel test was applied, stratified by the three-level Provider Organ Support assessment; Mantel–Haenszel adjusted odds ratios with 95% confidence intervals are reported.

Sensitivity analysis. AUROCs were computed separately by site for the Sepsis ImmunoScore, provider assessment, and standard diagnostics (PCT, CRP, NEWS) as a pre-specified sensitivity analysis to evaluate consistency of Sepsis ImmunoScore performance across settings and to characterize sources of site-level variability in provider accuracy. Internal validation was performed via repeated 10-fold cross-validation (B = 1,000) to assess potential overfitting.

Results

The study enrolled 445 patients who had blood cultures ordered and remnant samples available from routine blood tests collected within the required timeframe. Of these, 401 (90.1%) generated valid Sepsis ImmunoScore results. Patients without a valid Sepsis ImmunoScore were more likely to have Sepsis-3 within 24 hours than those with valid results (40.9% vs. 25.7%, p = 0.048), and to have a longer length of stay (4.2 days [95% CI, 2.9–6.1] vs. 2.8 days [2.4–3.2], p = 0.036). They were also more likely to be female and to have higher provider risk assessment results, but had no other significant differences (see Table S1 for a full comparison of baseline demographics and outcomes).

Of the patients with valid Sepsis ImmunoScore results, baseline demographic characteristics were similar between sepsis and non-sepsis patients (Table S2). Sepsis patients had higher prevalence of

immunocompromise and congestive heart failure (Table S2) and experienced significantly higher rates of all outcomes of interest than non-sepsis patients, except for readmission and 1-year mortality (Table 1).

Table 1
Patient Outcomes and Risk Assessments by Sepsis-3 Status

Characteristic	Overall	Non-Sepsis	Sepsis	P-value
N	401	298 (74.3%)	103 (25.7%)	
Outcomes				
Antimicrobial Days, median (IQR)	3.0 (1.0–6.0)	2.0 (1.0–5.0)	5.0 (4.0–8.0)	< 0.001
Inpatient, n (%)	321 (80.0%)	225 (75.5%)	96 (93.2%)	< 0.001
ICU Transfer (24h)	81 (20.2%)	31 (10.4%)	50 (48.5%)	< 0.001
28-Day In-Hospital Mortality	23 (5.7%)	4 (1.3%)	19 (18.4%)	< 0.001
Organ Dysfunction (%)	187 (46.6%)	84 (28.2%)	103 (100.0%)	< 0.001
30-Day Readmission	64 (16.0%)	48 (16.1%)	16 (15.5%)	1.000
Assisted Ventilation (24h)	54 (13.5%)	22 (7.4%)	32 (31.1%)	< 0.001
Vasopressor Use (24h)	46 (11.5%)	10 (3.4%)	36 (35.0%)	< 0.001
Critical Care Intervention (24h)	77 (19.2%)	28 (9.4%)	49 (47.6%)	< 0.001
Length of Stay, days, mean (95% CI)	2.8 (2.4–3.2)	2.2 (1.9–2.5)	6.4 (5.3–7.8)	< 0.001
Sepsis-3 Within 24 Hours (%)	103 (25.7%)	0 (0.0%)	103 (100.0%)	< 0.001
Readmission or Mortality Within 1 year	192 (47.9%)	147 (49.3%)	45 (43.7%)	0.536
3-Hour Broad-Spectrum Abx	274 (68.3%)	185 (62.1%)	89 (86.4%)	< 0.001
Sepsis ImmunoScore				
Risk Score, median (IQR)	20 (10–60)	20 (10–40)	70 (40–80)	< 0.001
Low	134 (33.4%)	125 (41.9%)	9 (8.7%)	< 0.001
Medium	82 (20.4%)	70 (23.5%)	12 (11.7%)	
High	164 (40.9%)	101 (33.9%)	63 (61.2%)	
Very High	21 (5.2%)	2 (0.7%)	19 (18.4%)	
NEWS Score at Evaluation, median (IQR)	3.0 (1.0–5.0)	2.0 (1.0–4.0)	4.0 (2.0–6.0)	< 0.001
Provider Infection Assessment				
Probability, median (IQR)	80.0 (50.0–95.0)	80.0 (50.0–95.0)	85.0 (50.0–95.0)	0.374

Characteristic	Overall	Non-Sepsis	Sepsis	P-value
Low	31 (7.7%)	27 (9.1%)	4 (3.9%)	0.090
Medium	114 (28.4%)	84 (28.2%)	30 (29.1%)	
High	121 (30.2%)	95 (31.9%)	26 (25.2%)	
Very High	135 (33.7%)	92 (30.9%)	43 (41.7%)	
Provider Sepsis Risk Assessment				
Probability, median (IQR)	50.0 (20.0–80.0)	45.0 (15.0–70.0)	70.0 (40.0–95.0)	< 0.001
Low	114 (28.4%)	101 (33.9%)	13 (12.6%)	< 0.001
Medium	152 (37.9%)	115 (38.6%)	37 (35.9%)	
High	72 (18.0%)	49 (16.4%)	23 (22.3%)	
Very High	63 (15.7%)	33 (11.1%)	30 (29.1%)	
Provider Organ Support Risk Assessment				
Probability, median (IQR)	20.0 (10.0–40.0)	15.0 (5.0–27.5)	33.0 (15.0–75.0)	< 0.001
Low	253 (63.1%)	210 (70.5%)	43 (41.7%)	< 0.001
Medium	94 (23.4%)	68 (22.8%)	26 (25.2%)	
High	33 (8.2%)	15 (5.0%)	18 (17.5%)	
Very High	21 (5.2%)	5 (1.7%)	16 (15.5%)	

[a] *Critical Care Intervention (24h) = Assisted Ventilation OR Vasopressor within 24 hours.*

Abx, antibiotics; CI, Confidence interval; h, hours; ICU, Intensive care unit; IQR, Interquartile range; N, Number of patients; Sepsis-3, Third International Consensus Definitions for Sepsis and Septic Shock.

Discriminative performance and additive value. Site-specific Provider Alone AUROC for Sepsis Risk and Infection Risk prediction varied widely across sites (0.57 to 0.83 for sepsis and 0.46 to 0.67 for infection), and overall AUROCs for NEWS, PCT, and CRP were 0.66 (95% CI, 0.60–0.72), 0.72 (95% CI, 0.67–0.78), and 0.63 (95% CI 0.57–0.69), respectively (Table S3). However, the overall Provider + Sepsis ImmunoScore combined model achieved an AUROC of 0.83 (95% CI 0.78–0.88) vs 0.68 (95% CI 0.63–0.74) for prediction of sepsis by Provider Alone (Fig. 1).

For secondary endpoints, Provider + Sepsis ImmunoScore combined predictions also showed consistent positive changes in AUROC across all secondary endpoints (+ 0.03 to + 0.06) (Table 2).

Table 2
AUROC Comparison of Secondary Endpoints

Outcome	Survey Question	N Events	Provider Alone (95% CI)	Sepsis ImmunoScore Alone (95% CI)	Provider + Sepsis ImmunoScore (95% CI)	Delta AUC
Sepsis-3 (24h)	**Sepsis**	**103**	**0.68 (0.63–0.74)**	**0.81 (0.76–0.86)**	**0.83 (0.78–0.88)**	**+ 0.15**
ICU Transfer (24h)	Organ Support	81	0.72 (0.66–0.78)	0.71 (0.65–0.78)	0.76 (0.70–0.82)	+ 0.04
28-Day Mortality	Organ Support	23	0.83 (0.74–0.93)	0.85 (0.77–0.92)	0.89 (0.81–0.96)	+ 0.05
Vasopressor (24h)	Organ Support	46	0.72 (0.64–0.81)	0.73 (0.64–0.82)	0.78 (0.70–0.85)	+ 0.06
Assisted Ventilation (24h)	Organ Support	54	0.69 (0.61–0.77)	0.68 (0.60–0.77)	0.72 (0.64–0.80)	+ 0.03
Critical Care (24h)[a]	Organ Support	77	0.70 (0.63–0.77)	0.70 (0.63–0.78)	0.74 (0.68–0.81)	+ 0.05

[a] *Critical Care = Vasopressor or Assisted Ventilation within 24 hours.*

AUC/AUROC, area under the curve; CI, confidence interval; h, hours; ICU, intensive care unit.

Discriminative performance was internally validated using repeated 10-fold cross-validation with bootstrap resampling (B = 1,000), demonstrating stable AUROC estimates for the combined model and limited evidence of overfitting (Table S4).

Additive value and reclassification. LRT demonstrated that inclusion of the Sepsis ImmunoScore significantly improved model fit beyond Provider Alone prediction of Sepsis-3 within 24 hours ($\chi^2 = 86.44$, $p < 0.001$) (Table 3). Significant improvements ($p < 0.01$) in model fit were observed for all five secondary endpoints.

Consistent with these findings, continuous net reclassification improvement (NRI) analysis demonstrated that the combined Provider + Sepsis ImmunoScore model more accurately classified individual patients than provider assessment alone. NRI was positive and statistically significant for all endpoints, indicating appropriate upward reclassification of events and downward reclassification of non-events with addition of the Sepsis ImmunoScore (Table 3).

Table 3
Likelihood Ratio Test (LRT) and Net Reclassification Improvement (NRI) for Sepsis-3 and Sepsis-Related Adverse Outcomes

Endpoint	Survey Question	LRT χ^2	LRT p-value	Continuous NRI (95% CI)	NRI p-value
Sepsis-3 (24h)	**Sepsis**	**86.44**	**< 0.001**	**0.98 (0.77–1.20)**	**< 0.001**
ICU Transfer (24h)	Organ Support	17.78	< 0.001	0.59 (0.31–0.83)	< 0.001
28-Day Mortality	Organ Support	16.85	< 0.001	0.98 (0.61–1.40)	< 0.001
Vasopressor (24h)	Organ Support	15.28	< 0.001	0.64 (0.32–0.93)	< 0.001
Assisted Ventilation (24h)	Organ Support	8.46	0.004	0.43 (0.17–0.68)	< 0.001
Critical Care (24h)	Organ Support	17.6	< 0.001	0.49 (0.28–0.77)	< 0.001

h, hours; ICU, intensive care unit; LRT, likelihood ratio test; NRI, net reclassification improvement.

Risk stratification. We cross-classified each patient for each categorical survey question (Infection Risk, Sepsis Risk, and Organ Support Risk), assigning them by the provider's assessment vs the Sepsis ImmunoScore classification of risk as Low, Medium, High, or Very High. Due to limited sample size in some strata, the High Risk and Very High Risk categories were combined for further analysis and, for secondary outcomes, Low/Medium were combined as well. Full 4×4 cross-classification tables using all original categories are provided in the Supplement (Table S5).

Figure 2 visualizes the stratification across the consolidated sepsis risk categories. For all three survey questions, Sepsis ImmunoScore re-stratified risk within every level of provider assessment with a consistent monotonic gradient. Even among patients whom providers assessed as Low Risk for Sepsis-3, Sepsis ImmunoScore High/Very High patients had a Sepsis-3 rate of 25.7%, compared to 3.8% for those placed in Low Risk categories by both providers and Sepsis ImmunoScore.

Sepsis-3 event rate stratification by Provider Sepsis Risk Assessment × Sepsis ImmunoScore (Low, Medium, High/Very High). Bars show n/N (%) in each Sepsis ImmunoScore category, and dashed reference lines show the actual event rate in each provider-assessed category, without stratification by Sepsis ImmunoScore. Within every provider risk stratum, Sepsis ImmunoScore provides additional separation of Sepsis-3 event rates.

Steeper gradients indicate a larger relative change in clinical risk across patients in the same provider category when accounting for their Sepsis ImmunoScore category. Among patients rated as Low Risk, Medium Risk, and High/Very High Risk by providers, patients with a High/Very High Sepsis ImmunoScore class were 6.8, 6.9, and 4.2 times more likely to have sepsis, respectively, than those with a Low Sepsis

ImmunoScore rating in the same provider categories. For example, across the provider High/Very High risk group, sepsis was confirmed in 13.3% of those with a Low Sepsis ImmunoScore rating, 19.2% of those with a Medium Sepsis ImmunoScore, and 55.7% of those with a High/Very High Sepsis ImmunoScore. This represents a 4.2-fold gradient in sepsis risk and an absolute difference of 42.4% among Sepsis ImmunoScore-stratified patients within the same provider risk group. By Cochran–Mantel–Haenszel analysis, a High/Very High Sepsis ImmunoScore was associated with significantly higher odds of Sepsis-3 within strata of provider assessment (adjusted OR 6.45, [95% CI 3.74–11.12]; $p < 0.001$) (Table S6).

For secondary endpoint risk stratification, results were similar; Sepsis ImmunoScore enhanced discrimination among patients within provider risk strata for critical care, ICU transfer, and mortality (Fig. 3); the categorical stratification was significant for each by Cochran–Mantel–Haenszel analysis (adjusted OR: critical care 2.77 [95% CI 1.54–5.01]; ICU transfer 2.67 [1.50–4.74]; 28-day mortality 7.28 [1.53–34.6]; all $p < 0.01$) (Table S6). Sample sizes for patients requiring vasopressor use or assisted ventilation were too small to generate a reliable estimate for these outcomes. Notably, among patients the providers assessed as High/Very High Risk, those who also had a High/Very High Sepsis ImmunoScore rating had a 31.8% rate of 28-day mortality vs 0% (0/10) for those identified as Low/Medium Risk by the Sepsis ImmunoScore. Sepsis ImmunoScore High/Very High patients consistently showed higher event rates than Sepsis ImmunoScore Low/Medium patients within each provider stratum. Together, this gradient and sub-stratification pattern suggest the Sepsis ImmunoScore can identify distinct risk profiles within a group of patients who all received the same score from providers, which holds implications for diagnosis or reassessment, clinical workflow, and patient disposition.

Representativeness of the surveyed cohort. Because clinician surveys were optional, we compared the 401 patients with completed surveys with 1,137 concurrently enrolled emergency department patients who had a valid Sepsis ImmunoScore result but no completed survey (Table S7). Patients with completed surveys were similar in most baseline characteristics and in the rate of Sepsis-3 within 24 hours (25.7% vs 24.9%, $p = 0.79$), but had modestly higher acuity, with more frequent ICU transfer (20.2% vs 11.4%), assisted ventilation (13.5% vs 7.3%), and vasopressor use (11.5% vs 7.8%), and a higher NEWS at evaluation. Despite these differences, the Sepsis ImmunoScore discriminated Sepsis-3 comparably in both groups (AUROC 0.81 with a completed survey vs 0.80 without a survey), with similar consistency for ICU transfer, critical care, and 28-day mortality (Table S8). In the full Sepsis ImmunoScore-resulted ED population (n = 1,538), the Sepsis ImmunoScore achieved an AUROC of 0.80 (95% CI, 0.78–0.83) for Sepsis-3, matching its performance in the pivotal validation study.(10) Because providers were blinded to the Sepsis ImmunoScore, survey completion could not have been influenced by the score. The stability of discrimination across patients with and without completed surveys indicates that the Sepsis ImmunoScore's additive value reflects the broader ED population rather than a feature specific to patients with completed surveys.

Discussion

In this prospective, multicenter cohort, Sepsis ImmunoScore consistently provided additive discrimination and clinically meaningful risk stratification beyond provider judgment in the evaluation of Sepsis-3 risk, infection risk, and the risk of sepsis-related adverse outcomes. These findings suggest that the Sepsis ImmunoScore provides additive value as a decision-support tool for early sepsis risk assessment and escalation of care.

Discriminative performance analyses showed that Sepsis-3 prediction accuracy consistently and substantially improved when provider assessment was augmented with Sepsis ImmunoScore. The combined model achieved an AUROC of 0.83 compared with 0.68 for Providers Alone, with incremental value confirmed by likelihood ratio testing. Internal cross-validation with bootstrap resampling further supports the robustness of this improvement, yielding stable AUROC estimates and narrow confidence bounds across iterations. This level of discrimination is consistent with the pivotal validation study and other studies of the Sepsis ImmunoScore, in which it achieved a comparable AUROC for sepsis,(6, 10) supporting the external validity of these findings.

Cross-classification of provider assessment and Sepsis ImmunoScore categories revealed a consistent monotonic gradient in Sepsis-3 event rates within every level of provider-assessed risk. These findings indicate that the Sepsis ImmunoScore identifies clinically meaningful heterogeneity within groups of patients who appear similar based on provider assessment alone. Regardless of whether providers assessed patients as Low, Medium, or High/Very High Risk, the Sepsis ImmunoScore further separated patients into groups with markedly different outcome rates. When examining patients that providers assessed as Low Risk for Sepsis-3, Sepsis ImmunoScore cross-category analysis revealed an occult high-risk subset of patients, among whom 1 in 4 had sepsis—a rate 6.8 times higher than patients classified as Low Risk by both provider and Sepsis ImmunoScore. This within-stratum discrimination extended to downstream outcomes of clinical significance, including 28-day mortality, critical care utilization, and ICU transfer. This supports the premise that providers and Sepsis ImmunoScore capture complementary aspects of patient risk rather than overlapping signals. Sepsis ImmunoScore's recognition of subtle, patient-specific biological risk factors could help identify occult sepsis sooner than reliance on provider recognition alone.

ED and ICU clinicians routinely manage high volumes of patients with nonspecific presentations, subtle early physiologic deterioration, or misleadingly benign initial vital signs. Sepsis ImmunoScore's value lies in its ability to integrate multivariate physiologic and biomarker signals that may not be readily apparent to the evaluating clinician in time-pressured environments. LRT and NRI analyses suggest that the Sepsis ImmunoScore adds independent prognostic value beyond clinician judgment alone for almost every endpoint. That added value was greatest where clinicians struggled most: across the three sites, the Sepsis ImmunoScore held steady (AUROC 0.80–0.84), whereas unaided provider discrimination for Sepsis-3 ranged from 0.57 to 0.83 (Table S3). At the site where provider performance was weakest, the Sepsis ImmunoScore discriminated as well as at the strongest. A risk signal that performs consistently regardless of local expertise or staffing could help close the documented gaps in sepsis recognition

between community hospitals and academic centers,(19–21) standardizing early identification and reducing unwarranted variation in care.

Collectively, these findings indicate that the Sepsis ImmunoScore can enhance, rather than replace, clinician assessment of sepsis risk. Integrated into ED and ICU workflows through the EHR, it returns an early, quantitative, explainable risk estimate alongside routine laboratory results and flags patients whose biomarker-based risk diverges from their clinical appearance. Concordance with the clinician's assessment can reinforce a decision; discordance can prompt the clinician to reconsider sepsis risk, pursue further investigation, or adjust the care plan. This is the logic of biomarker-guided decision support: a rapid, non-binding signal clinicians may follow or override, which PCT-guided workflows have begun to validate in the ED.(22) Because the Sepsis ImmunoScore discriminated sepsis more accurately than PCT (AUROC 0.81 vs 0.72; Table S3), it is well suited to this role. Our study only establishes predictive value, not a direct effect on care; an interventional trial is the next step to test whether acting on the Sepsis ImmunoScore improves patient outcomes.

Limitations. The observational design of this study cannot establish whether access to Sepsis ImmunoScore results changes provider behavior or patient outcomes. This study demonstrates the independent added value of the Sepsis ImmunoScore to clinical decision-making using prospectively collected real-world data, but interventional studies are needed to quantify the impact on clinical actions and outcomes. Some secondary endpoints also had limited event counts (e.g., 28-day mortality), affecting power for certain analyses despite consistent directional findings. The Sepsis ImmunoScore requires 12 of 22 routinely collected inputs—six vital signs plus white blood cell count, blood urea nitrogen, creatinine, platelets, CRP, and PCT—to be measured to generate a Sepsis ImmunoScore result. When any was not ordered or had not returned in time, no score was produced, as observed for 9.9% of eligible patients. This missingness was not random: patients without a result more often had sepsis (40.9% vs 25.7%, p = 0.048) and stayed longer. Our estimates therefore describe a result-available population and may not extend to the sickest patients, in whom these inputs were more often delayed—a selection effect that could inflate or attenuate discrimination. Because the required inputs are routine in acute care, this gap reflects ordering and timing rather than unusual data demands; deployment studies should analyze every enrolled patient (intention-to-diagnose) and record which inputs were missing. Finally, clinician surveys were completed for a subset of enrolled patients; although surveyed patients were somewhat more acute than non-surveyed patients, the Sepsis ImmunoScore discriminated comparably in both groups and Sepsis-3 rates were similar, indicating the surveyed cohort was generally representative (Tables S7 and S8).

Conclusion

Across a multi-site cohort of patients with suspected infection, the Sepsis ImmunoScore demonstrated consistent additive value as an adjunct to clinician assessment, offering improved discrimination of high-risk and low-risk cases of sepsis and related adverse outcomes. Cross-class risk stratification highlights Sepsis ImmunoScore’s biology-based ability to surface occult sepsis risk or redirect clinical

resources, while also confirming the ability of the Sepsis ImmunoScore to enhance rather than replace clinical judgment. Its consistent performance across institutions and its capacity to add independent value to clinician assessments suggest that the Sepsis ImmunoScore could meaningfully support emergency and critical care decision-making and improve the precision of early sepsis diagnosis and management. Prospective interventional studies are warranted to determine the impact of real-time Sepsis ImmunoScore-guided decision support on clinical care and patient outcomes.

Declarations

Author contributions

AB conceptualized the study, contributed to study methodology, led the data analysis and interpretation, and drafted the manuscript. GLW, NS, CLE, BRJ, and RB assisted with manuscript drafting. NS, CLE, GLW, and LS contributed to study methodology, data acquisition, data analysis, and data interpretation. GLW and LU assisted in study design, data acquisition and interpretation. SK, RC, GS, LU, and DU contributed to study methodology and data acquisition. AB, GLW, NS, BRJ, and CLE contributed to study concept and design. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank the study coordinators, research staff, clinicians, and lab technicians for their contributions to this study. Their work was conducted as part of their job duties; they did not receive additional compensation. Writing support was provided by Katherine Brind'Amour of HealthWords, Ltd.

This study was funded by the Biomedical Advanced Research and Development Authority (BARDA), Contract75A50123C00019. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Competing interests

AB, CLE, LS, SK, GLW, RC, GS, LU, DU, RB, BRJ, and NS disclose employment and/or financial interests in Prenosis, Inc.

Data availability

Restrictions apply to the availability of this study's data, which were used under license for the current study and are not publicly available. Data are available from the authors upon reasonable request and with permission from Prenosis, Inc.

Code availability

The underlying code for this study is not publicly available for proprietary reasons.

References

1. Rudd, K.E., Johnson, S.C., Agesa, K.M., Shackelford, K.A., Tsoi, D., Kievlan, D.R. et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet*. 395(10219), 200–11 (2020).
2. Kim, H.I., Park, S. Sepsis: Early recognition and optimized treatment. *Tuberc Respir Dis* (Seoul). 82(1), 6–14; doi:10.4046/trd.2018.0041 (2019).
3. Evans, L. et al. Surviving sepsis campaign: international guidelines for management of sepsis and sepsis shock 2021. *Intensive Care Med*. 47(11), 1181–1247; doi:10.1007/s00134-021-06506-y (2021).
4. Williams, B. The National Early Warning Score: from concept to NHS implementation. *Clin Med* (Lond). 22(6):499–505; doi:10.7861/clinmed.2022-news-concept (2022).
5. Churpek, M.M. et al. Quick Sepsis-related Organ Failure Assessment, Systemic Inflammatory Response Syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. *Am J Respir Crit Care Med*. 195(7):906–911; doi:10.1164/rccm.201604-0854OC (2017).
6. Watson, G.L. et al. The Sepsis ImmunoScore predicts sepsis, mortality, and deterioration better than clinical scores and widely available biomarkers. *medRxiv*.10.01.25337117; doi:https://doi.org/10.1101/2025.10.01.25337117 (2025).
7. Hwang, M.I., Bond, W.F., Powell, E.S. Sepsis alerts in emergency departments: A systematic review of accuracy and quality measure impact. *West J Emerg Med.* 21(5), 1201–10 (2020).
8. Bignami, E.G. et al. Artificial Intelligence in Sepsis Management: An Overview for Clinicians. *J Clin Med*. 14(1), 286; doi:10.3390/jcm14010286 (2025).
9. Papareddy, P. et al. Transforming sepsis management: AI-driven innovations in early detection and tailored therapies. *Crit Care*. 29(1), 366; doi:10.1186/s13054-025-05588-0 (2025).
10. Bhargava, A. et al. FDA-Authorized AI/ML Tool for Sepsis Prediction: Development and Validation. *NEJM AI*. 1(12); AIoa2400867 (2024).
11. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. Sepsis ImmunoScore DEN230036 Marketing Authorization Letter. April 2, 2024. https://www.accessdata.fda.gov/cdrh_docs/pdf23/DEN230036.pdf. Accessed 11 Sep 2025.
12. Singer, M. et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 315(8):801–810; doi:10.1001/jama.2016.0287 (2016).
13. Watson, G.L. et al. Interpretability of an FDA-authorized AI/ML sepsis diagnostic tool improved by SHAP values. *JAMIA Open*. 9(1):ooag020; doi.org/10.1093/jamiaopen/ooag020 (2026).
14. Lundberg, S.M., Lee, S.I. A unified approach to interpreting model predictions. Advances in neural information processing systems. 31st Conference on Neural Information Processing Systems (NIPS 2017). 30; doi.org/10.48550/arXiv.1705.07874 (2017).

15. Rhee, C. et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009–2014. *JAMA.* 318(13):1241–9; doi:10.1001/jama.2017.13836 (2017).
16. Pepe, M.S., Kerr, K.F., Longton, G., Wang, Z. Testing for improvement in prediction model performance. *Stat Med*. 32(9):1467–1482; doi: 10.1002/sim.5727 (2013).
17. Pencina, M.J., D'Agostino, R.B. Sr., Steyerberg, E.W. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 30(1), 11–21; doi:10.1002/sim.4085 (2011).
18. Montaña, R., Flores, S.V., Nieves-Mariñez, A., Jara, M. Prueba Cochran-Mantel-Haenszel: herramienta clave en la investigación médica para el control de variables confusoras [Cochran-Mantel-Haenszel Test: A key tool in medical research for controlling confounding variables]. *Semergen.* 52(2):102665; doi:10.1016/j.semerg.2025.102665 (2026).
19. Draeger, L. et al. Barriers and facilitators to optimal sepsis care – a systematized review of healthcare professionals' perspectives. *BMC Health Serv Res*. 25, 591; doi.org/10.1186/s12913-025-12777-8 (2025).
20. Abdalhafith, O. et al. Critical care nurses' knowledge, confidence, and clinical reasoning in sepsis management: a systematic review. *BMC Nurs*. 24, 424; https://doi.org/10.1186/s12912-025-02986-1 (2025).
21. Rudd, K.E. et al. The global burden of sepsis: barriers and potential solutions. *Crit Care*. 22, 232; doi.org/10.1186/s13054-018-2157-z (2018).
22. Todd, S. et al. Procalcitonin testing combined with NEWS2 evaluation compared with usual care based on NEWS2 for identification of sepsis and antibiotic initiation in the emergency department in England and Wales (PRONTO): a multicentre, randomised, controlled, open-label, phase 3 trial. *Lancet Respir Med*. 14(5):417–431; doi:10.1016/S2213-2600(25)00433-3 (2026).

Figures

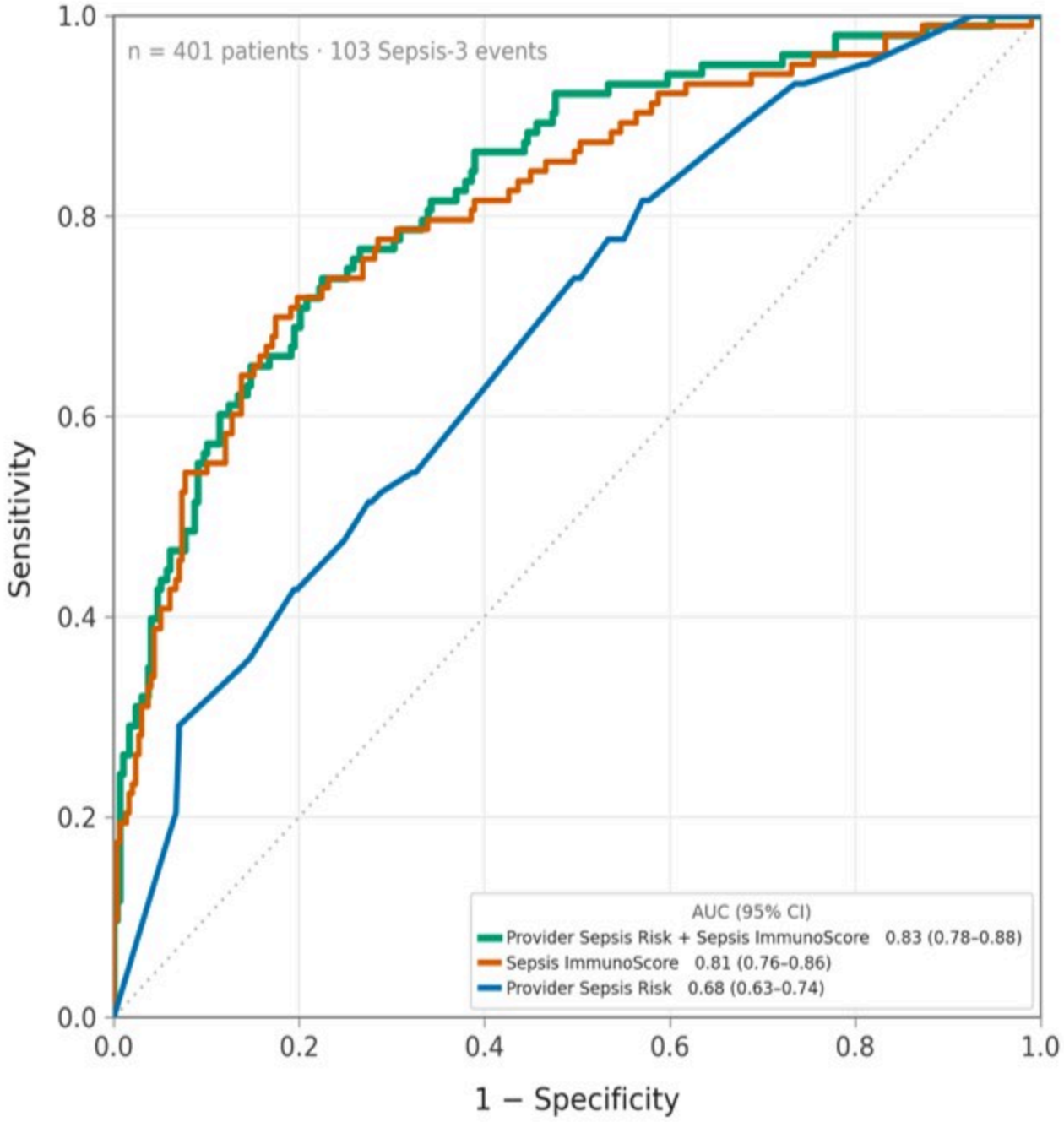


Figure 1

AUROC for Sepsis-3 Criteria within 24 Hours of Evaluation

AUC/AUROC, area under the curve; CI, confidence interval; SIS, Sepsis ImmunoScore.

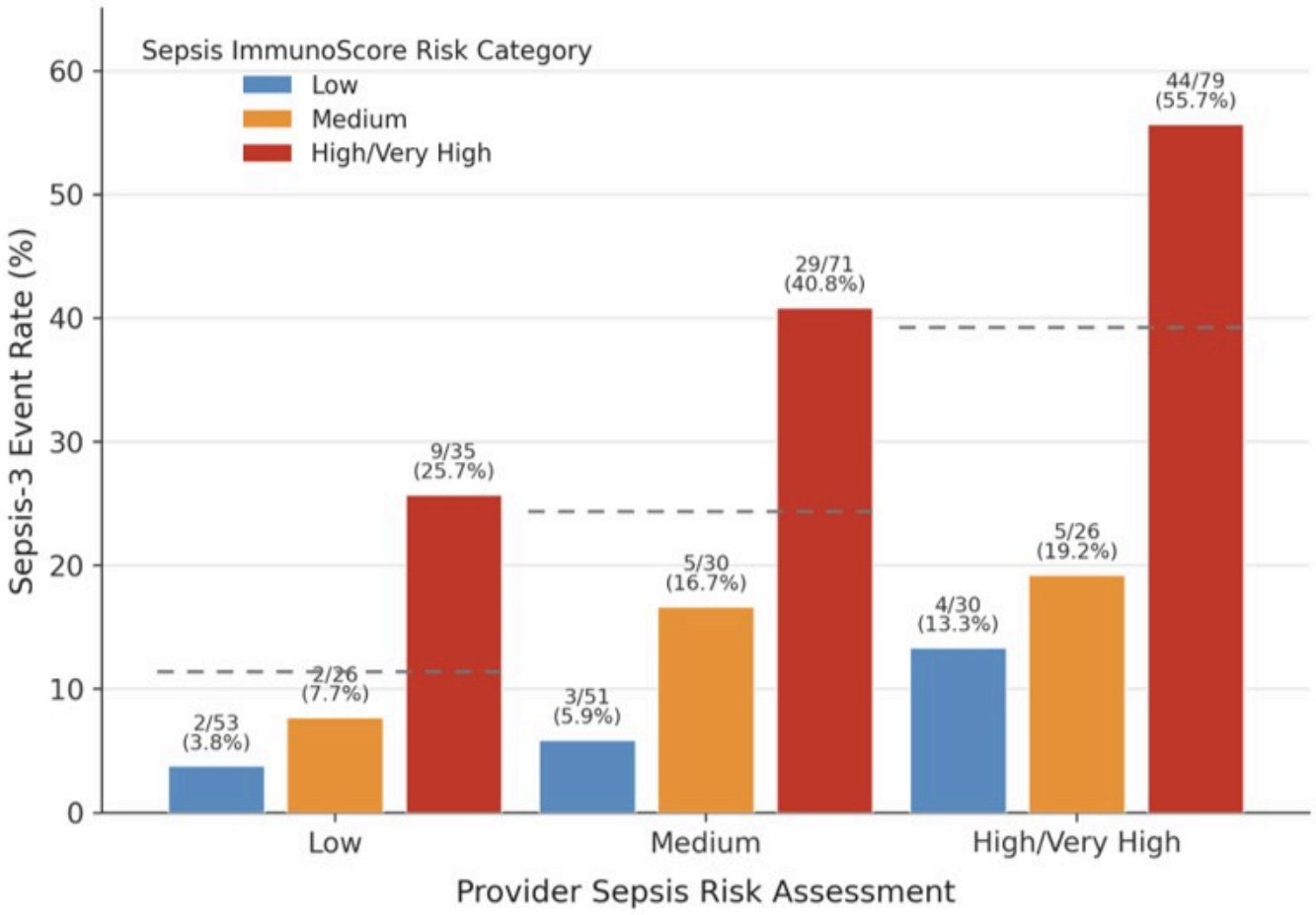


Figure 2

Sepsis-3 Event Rate by Provider Sepsis Risk Assessment, Stratified by Sepsis ImmunoScore Assessment.

Sepsis-3 event rate stratification by Provider Sepsis Risk Assessment × Sepsis ImmunoScore (Low, Medium, High/Very High). Bars show n/N (%) in each Sepsis ImmunoScore category, and dashed reference lines show the actual event rate in each provider-assessed category, without stratification by Sepsis ImmunoScore. Within every provider risk stratum, Sepsis ImmunoScore provides additional separation of Sepsis-3 event rates.

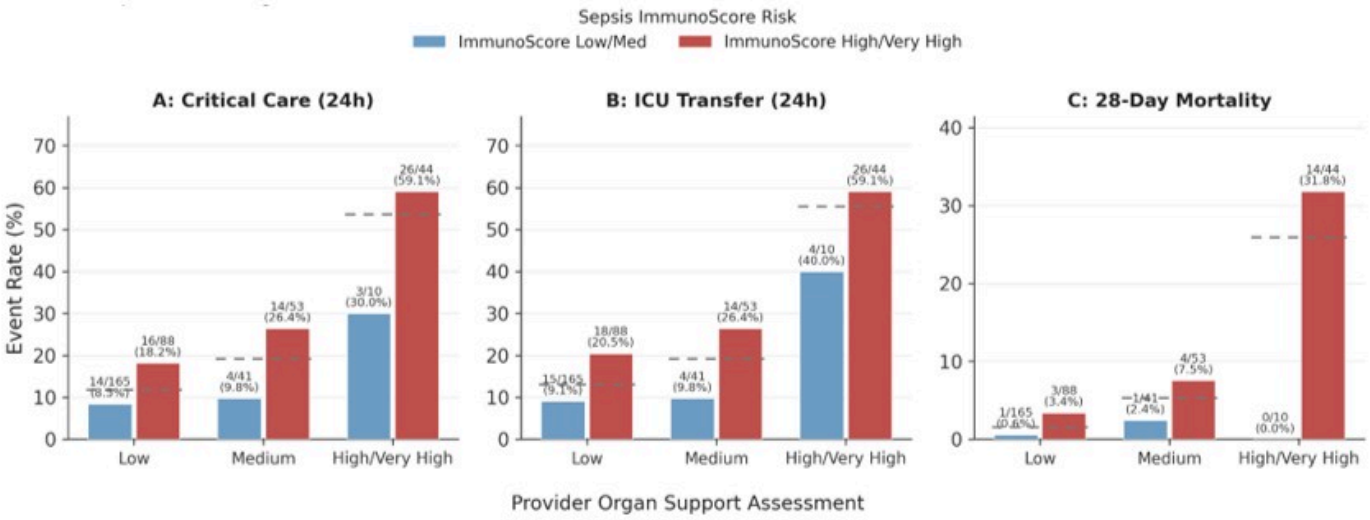


Figure 3

Secondary Endpoint Stratification by Provider Organ Support Risk Assessment, Stratified by Sepsis ImmunoScore (Low/Med vs High/Very High).

Secondary endpoint rate stratification bars have n/N (%) annotations, and dashed reference lines show actual event rate in each provider-assessed category, without stratification by Sepsis ImmunoScore. Within every provider risk stratum, Sepsis ImmunoScore provides additional separation of secondary outcome rates.

h, hours; ICU, intensive care unit; Med, Medium.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplement062326.docx