Validation of a Machine Learning Diagnostic Tool for the Prediction of Sepsis and Critical Illness

Akhil Bhargava, M.S.1; Carlos Lopez-Espina, M.S.1; Lee Schmalz, B.S.,1; Shah Khan, PhD,1; Gregory L. Watson, PhD,1; Dennys Urdiales B.S.,1; Lincoln Updike B.S.,1; Niko Kurtzman, M.D.,2; Alon Dagan, M.D.,3; Amanda Doodlesack, M.D.,3; Bryan Stenson, M.D.,3; Deesha Sarma, M.D.,3; Eric Reseland, M.D.,3; John H. Lee, M.D. PhD3; Max Kravitz, M.D.,3; Peter S. Antkowiak, M.D., MPH3; Tatyana Shvilkina, D.O.,3; Aimee Espinosa, M.D.,5; Alexandra Halalau, M.D.,5; Carmen Demarco, M.D.,5; Francisco Davila, M.D.,5; Hugo Davila, M.D.,5; Matthew Sims, M.D., Ph.D.,5; Nicholas Maddens, M.D.,3; Ramona Berghea, M.D.,5; Scott Smith, M.D.,5; Ashok V. Palagiri, M.D.,5; Clinton Ezekiel, M.D.,7; Farid Sadaka, M.D.,7; Karthik Iyer, M.D.,7; Matthew Crisp, M.D., Ph.D.,7; Saleem Azad, D.O.,7; Vikram Oke, M.D.,7; Andrew Friederich, M.D.,8; Anwaruddin Syed, M.D.,8; Falgun Gosai, M.D.,8; Lavneet Chawla, M.D.,8; Neil Evans, M.D.,8; Kurian Thomas, M.D.,8; Roneil Malkani, M.D.,10; Roshni Patel, M.D.,10; Stockton Mayer, D.O.,10; Farhan Ali, M.D.,11; Lekshminarayan Raghavakurup, M.D.,11; Muleta Tafa, M.D., MPH11; Sahib Singh, M.D.,11; Samuel Raouf, M.D.,11; Sihai Dave Zhao, PhD12; Ruoqing Zhu PhD,12; Rashid Bashir, PhD13; Bobby Reddy, Jr. PhD1 and Nathan I. Shapiro M.D., MPH1,3,4

1 Prenosis, Inc., Chicago, Illinois, United States
2 Emory School of Medicine, Atlanta, Georgia, United States
3 Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, United States
4 Harvard Medical School, Boston, Massachusetts, United States
5 Beaumont Health System, Royal Oak, Michigan, United States
6 Mercy Hospital, St. Louis, Missouri, United States
7 Mercy Hospital, Jefferson, Missouri, United States
8 OSF Saint Francis Medical Center, Peoria, Illinois, United States
9 Davis School of Medicine, University of California, Sacramento, California, United States
10 Jesse Brown VA Medical Center, Chicago, Illinois, United States
11 Lifebridge Sinai Hospital, Baltimore, Maryland, United States
12 Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States
13 Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States

Corresponding author:
Nathan I. Shapiro, MD, MPH
Professor of Emergency Medicine
Department of Emergency Medicine
Beth Israel Deaconess Medical Center
Boston, MA
Nshapiro@bidmc.harvard.edu

Word Count: 2886
Key Points

Question: Is it feasible to develop an Artificial Intelligence/Machine Learning (AI/ML) learning model that accurately identifies patient risk for sepsis and sepsis-related critical illness?

Findings: The FDA approved AI/ML Sepsis ImmunoScore algorithm was created using a combination of 22 different demographic, clinical, and laboratory variables to predict risk of sepsis within 24 hours. The model was accurate, with an AUROC of 0.81 (0.77–0.86) in external validation. The algorithm was also predictive of secondary outcomes of sepsis-related critical illness.

Meaning: This Sepsis ImmunoScore algorithm identifies patients suspected of infection who are at high risk of having or developing sepsis and sepsis-related critical illness.

Abstract

Importance: Prompt and accurate diagnosis and risk assessment is a challenge with implications for clinical care of sepsis patients.

Objective: To describe the development of the Sepsis ImmunoScore Artificial Intelligence/Machine Learning (AI/ML) algorithm and assess its ability to identify patients with sepsis within 24 hours, and secondary endpoints of critical illness and mortality.

Design: Prospective study of adult (age 18 or older) patients from 5 US hospitals enrolled between April 2017 and July 2022.

Setting: Multi-center study from 5 hospitals

Participants: Inclusion criteria: suspected infection (indicated a blood culture order), emergency department or hospitalized patients, with a corresponding lithium-heparin plasma
sample available; exclusion criteria: none. Participants were enrolled into an algorithm development derivation cohort (n=2,366), an internal validation (n=393) cohort, or an external validation cohort (n=698).

**Main Outcomes and Measures:** The primary endpoint was the presence of sepsis (Sepsis-3) within 24 hours of test initiation. Secondary endpoints were clinically relevant metrics of critical illness: length of stay in the hospital, Intensive Care Unit (ICU) admission within 24 hours, use of mechanical ventilation within 24 hours, use of vasopressors within 24 hours, and in-hospital mortality.

**Results:** The overall diagnostic accuracy of the Sepsis ImmunoScore for predicting sepsis was high with an AUC of 0.85 (0.83–0.87) in the derivation cohort, 0.80 (0.74–0.86) in internal validation, and 0.81 (0.77–0.86) in external validation. The Sepsis ImmunoScore was divided into four risk categories with increasing likelihood ratios for sepsis: low 0.1 (0.1–0.2), medium 0.5 (0.3–0.8), high 2.1 (1.8–2.5), very high 8.3 (4.1–17.1). Risk categories also predicted in-hospital mortality rates: low: 0.0% (0.0%, 1.6%), medium: 1.9% (0.4%–5.5%), high: 8.7% (5.7%–12.7%), and very high: 18.2% (7.0%–35.5%) in the external validation cohort. Similar findings were observed for length of stay, ICU utilization, mechanical ventilation and vasopressor use.

**Conclusions and Relevance** The sepsis ImmunoScore, an AI/ML diagnostic tool, demonstrated high accuracy for predicting sepsis and critical illness that could enable prompt identification of patients at high risk of sepsis and adverse outcomes, which holds promise to inform medical decision making to improve care and outcomes in sepsis.
**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 an opportunity for risk assessment tools to assist clinicians in the quick and accurate identification of 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 an informative risk assessment tool in the hospital setting, we developed the Sepsis Immunoscore. The Sepsis ImmunoScore is a risk stratification tool that uses machine learning to aid in identifying patients likely to have or progress to sepsis within 24 hours of patient assessment. It was granted marketing authorization (De Novo pathway) by the United States Food and Drug Administration (FDA) in April 2024 as the first-ever AI diagnostic authorized for sepsis. The Sepsis ImmunoScore inputs up to 22 parameters derived from patient demographics, vital signs, routinely accepted general clinical laboratory tests, and sepsis specific biomarkers to generate a composite risk score. The risk score categorizes patients into one of four discrete risk groups based on the risk of sepsis. The Sepsis ImmunoScore embeds into a hospital EMR and functions as a diagnostic test, allowing healthcare providers to order and view the test results for a particular patient in the Electronic Health Record (EHR) system, similar to a laboratory test.

In this investigation, we describe the derivation and assess the performance of the Sepsis ImmunoScore functioning as a sepsis risk assessment tool. Accordingly, 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 (defined by Sepsis-3) within 24 hours,
and for secondary endpoints of in-hospital mortality, hospital length of stay, ICU admission, mechanical ventilator use, and vasopressor medication use.\textsuperscript{9}

---

**Methods**

**Study Design**

We conducted a prospective, observational, multi-center study to create a sepsis artificial intelligence/machine learning (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 (eFigure 1). Participants were enrolled at one of 5 participating hospitals. We obtained study approvals from the ethics boards of participating institutions under a waiver from informed consent, except OSF Saint Francis Medical Center, which required informed consent.

**Study Population**

Study inclusion criteria consisted of hospitalized adult patients (aged 18 or older) who had a suspected infection defined by the clinical decision to obtain a blood culture and who had a lithium-heparin (Li-Hep) plasma sample drawn within a 6-hour-window from the first blood culture order that was available for collection. There were no exclusion criteria. Subjects were enrolled between April 2017 and July 2022 from 5 hospital institution sites throughout the United States. The study participants were enrolled in three different cohorts: a *derivation cohort* (n=2,366) 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) that used a new set of participants from hospitals not involved in the algorithm derivation (additional details in supplemental appendix).
Study Outcomes

Endpoints

The primary endpoint was the presence of sepsis at presentation or within 24 hours of study inclusion using the Sepsis-3 criteria: suspected infection and Sequential Organ Failure Assessment (SOFA) score of 2 or greater from baseline. The derivation cohort used a sepsis-3 outcome derived from the medical record in an automated fashion, while the internal and external validation cohorts used expert clinical adjudication to apply the definitions and determine the sepsis-3 outcome. The secondary endpoints consisted of sepsis-related metrics of critical illness including: in-hospital mortality, hospital length of 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 linked to corresponding patient blood specimens. Data elements were abstracted from the EMR and included demographic information, coded ICD-10 diagnoses, medications, vital sign measurements, clinical laboratory test results (e.g., chemistry laboratory testing results, lactic acid), and sepsis-related laboratory measurements (C-reactive protein and procalcitonin – tested at external lab – see supplemental appendix for details), secondary outcomes metrics, and 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 (CCI) and were encoded based on ICD-10-CM encodings defined by the National Cancer Institute (NCI) Comorbidity Index/SEER. Immunocompromised patients were...
identified based on ICD-10-CM encodings defined by Agency for Healthcare Research and Quality (AHRQ).12,13

**Sepsis ImmunoScore**

**Algorithm Development**

The Sepsis ImmunoScore machine learning algorithm was created using a supervised, calibrated random forest that predicts the probability of a patient meeting Sepsis-3 criteria within 24 hours of study entry. A random forest was trained on the 2,366 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 3 repeats of 5-fold-cross-validation, and missing data were imputed using bagged trees. Predictions were calibrated to the probability of sepsis-3 to compute a sepsis risk score by regressing the outcome on the out-of-bag predictions of the random forest in the derivation cohort.14 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 online supplement)

**Risk Score and Risk Stratification 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 cell count, platelet count, creatinine, or blood urea nitrogen 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 between six hours prior to study entry and 3.5 hours after. Missing values for the remaining 12 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 likelihood ratios and predictive values along with 95% confidence intervals 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.\(^\text{15,16}\) We also estimated the area under the receiver operating characteristic curve (AUROC) of the sepsis risk score. Confidence intervals for the AUROC were estimated using a binormal approximation estimator for the standard error.\(^\text{17}\) Analyses were conducted using R statistical software version 4.2.1.

**Sample Size Calculation**

This study was powered based on the confidence interval of the AUROC for the sepsis endpoint.\(^\text{18}\) 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 age, racial backgrounds, ethnicities, and geographic location. 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 split the cohort into the current internal and external validation format.
Results

There were a total of 3,457 patient encounters included, with 2,366 encounters in the derivation set, 393 in the internal validation set, and 698 visits in the external validation set. The study enrolled participants with age, sex, race, and ethnicity, and co-morbidities typical of sepsis patients in the US (Table 1). The rate of sepsis was 32% in the derivation, 28% in internal validation, and 22% in the external validation cohorts (Table 1). Patients with sepsis had higher rates of severe illness and mortality compared to those without sepsis (Table 1).

Sepsis ImmunoScore

The Sepsis ImmunoScore algorithm uses up to 22 input parameters to generate the risk score and place patients in one of four discrete risk stratification categories. The 22 input parameters consist of demographic data (age), vital sign measurements, complete-metabolic-panel measurements, complete blood count panel measurements, lactate, and sepsis biomarkers PCT and CRP.

Primary Endpoint

The overall diagnostic accuracy for the Sepsis ImmunoScore was high for predicting sepsis with an AUC in the derivation set of 0.85 (95% confidence interval: 0.83–0.87) for the medical record derived sepsis outcome, and 0.80 (0.74–0.86) in the internal validation and 0.81 (0.77–0.86) in the external validation for the adjudicated sepsis outcome (eTable 1). The Sepsis ImmunoScore was divided into four risk categories with increasing risk of sepsis. (Figure 1, Table 2, eTable 2). Of note, in the external validation set, the likelihood ratios were: low 0.1 (0.1–0.2), medium 0.5 (0.3–0.8), high 2.1 (1.8–2.5), very high 8.3 (4.1–17.1). These are monotonically increasing and had no overlapping confidence intervals suggesting stepwise risk discrimination for sepsis.
Secondary Endpoints

We assessed the prognostic ability for the Sepsis ImmunoScore to predict the secondary outcomes of ICU admission within 24 hours, in-hospital mortality, 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 ranging from low, medium, high, and very high demonstrated good predictive ability based on both rate of outcome as well as the corresponding stratum specific likelihood ratios (Figure 2, Table 3, eTable 3). 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%, 1.6%), 1.9% (0.40%–5.5%), 8.7% (5.7%–12.7%), and 18.2% (7.0%–35.5%) respectively. Additionally, the observed median number of days for the composite length of stay endpoint in the low, medium, high, and very high-risk groups were: 4.0 (3.5–4.9), 5.7 (4.9–7.0), 7.7 (6.5–8.5), and 13.5 (7.1–19.1) respectively. The proportion of patients transferred to the ICU within 24 hours was 4.7% (2.4%–8.3%), 12.7% (8.0%–19.0%), 25.7% (20.7%–31.3%), and 54.6% (36.4%–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 predictive value and risk stratification category severity for each secondary endpoint (p-value < 0.01, Table 3, eTable 3). Risk stratification category severity was also associated with time to event for each secondary endpoint (eFigure 2).

Discussion

The Sepsis ImmunoScore is a comprehensive, multidimensional AI/ML tool that combines demographics, vital signs, clinical laboratory tests, and sepsis focused laboratory tests to assess risk of sepsis and risk of adverse outcomes. In this study, we developed the Sepsis
ImmunoScore and analyzed its ability to serve as a risk-stratification tool for patients with suspected infection, and its ability to predict the diagnosis of sepsis and prognosticate adverse clinical outcomes. We found the Sepsis ImmunoScore highly predictive of sepsis and secondary outcomes of in-hospital mortality, hospital length of stay, ICU admission, mechanical ventilation, and vasopressor administration within 24 hours.

There are a number of FDA-approved diagnostic tools available for patients with an infection; however, they are typically in the form of a single blood biomarker or sometimes multiple blood biomarkers. Procalcitonin 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. The IntelliSep Test is a blood test that measures leukocyte biophysical properties to create a score that identifies sepsis with organ dysfunction manifesting within the first three days after testing for adult patients with signs and symptoms of infection who present to the emergency department. Another test by Beckman, the Coulter Cellular Analysis System’s Early Sepsis Indicator measures Monocyte Distribution Width to identify sepsis risk. Other tests distinguish bacterial from non-bacterial infection in the ED or urgent care settings such as the FebriDx test which measures myxovirus resistance protein A and CRP from finger-stick blood. The MeMed BVTM measures blood concentrations of TRAIL, IP-10, and CRP to also distinguish patients with bacterial infections from those without. The Sepsis ImmunoScore has comparable or superior diagnostic accuracy to these other FDA approved tests. Moreover, the ImmunoScore uses multidimensional inputs across different domains (demographics, vital signs, laboratory tests etc.) plus sepsis biomarkers to create a comprehensive risk score for a given individual. The intent of the ImmunoScore is to embed in an EMR so that it can pull the different requisite inputs and display the score when it is ordered as a diagnostic test.

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. A large, external validation study of the widely deployed Epic Sepsis Model in 2021 reported an AUC of only 0.63, and recent reviews of validation studies of the Targeted Real-time Early Warning System (TREWS) score have raised concerns regarding the control group and false positives. Concerns of alert fatigue have also been raised for these systems, which may undermine their clinical utility.

The application of AI/ML to medicine has great potential, much of which is underdeveloped in medicine. The Sepsis ImmunoScore used clinically available data reflective of patient biologic state and machine learning to incorporate and 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 that underlie critical illness and are routinely collected or available in the setting of infection. 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 capitalizes on the synergy of thoughtfully applied AI/ML to expertly curated biologic data to better equip—not replace—clinicians in their challenging fight against sepsis.

Sepsis represents an ongoing diagnostic challenge to clinicians due to its often subtle and heterogeneous presentation. Determining the presence or likelihood of progression to sepsis, and the severity with associated clinical needs represents a continuing challenge to clinicians. The Sepsis ImmunoScore is unique in its approach due to its machine-learning based incorporation of 22 parameters to comprehensively assess a patient’s risk of being diagnosed with sepsis, plus its association with adverse outcomes. The ImmunoScore should serve as an adjunctive test to assist clinical decision making in the acute setting. Given its
strong predictive ability, the Sepsis ImmunoScore may help to improve patient outcomes by adequately 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 may also help to reduce over-triage by more accurately identifying patients at low risk for deterioration due to infection, for example allowing emergency department physicians to potentially treat these low-risk patients in the outpatient setting and promote antimicrobial stewardship.

**Limitations**

There are a number of limitations to our study. First, we used 5 hospitals in the study, it is possible that our findings may not generalize to specific populations that may differ from our hospitals. Second, we relied upon an EMR extraction, so it is possible that missingness or the use of ICD10 codes may have led to misclassification of certain elements such as comorbidities. Third, this was an observational study so we cannot assess the impact of the ImmunoScore on clinical decision-making and changes in therapeutic approaches. Fourth, the primary outcome of Sepsis-3 within 24 hours relied upon an automated calculation in the derivation set and adjudication for presence of infection in the internal and external validation; thus, misclassification of outcome may have occurred. Fifth, our inclusion criteria used the ordering of a blood culture as a surrogate indicator for a clinical suspicion of infection and patients where there was a clinical suspicion may not have had a blood culture ordered or other patients may have had a blood culture performed who had a very low (or no) suspicion of infection. Finally, covariate missingness may have affected algorithm performance.
Conclusions

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.
Acknowledgements

Authorship Contributions: Dr. Shapiro and Bhargava had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Bhargava, Lopez-Espina, Schmalz, Watson, Zhao, Zhu, Bashir, and Reddy Jr., and Shapiro


Drafting of the manuscript: Bhargava, Watson, and Shapiro

Critical revision of the manuscript for important intellectual content: All authors

Statistical analysis: Bhargava, Schmalz, and Watson

Obtained funding: Lopez-Espina, Watson, Bashir, and Reddy Jr.

Administrative, technical, or material support: Reddy Jr.

Study supervision: Bhargava, Lopez-Espina, Schmalz, Reddy Jr., and Shapiro

Funding/Support: This study was funded in part by the Defense Threat Reduction Agency, National Institutes of Health, Centers for Disease Control and Prevention, National Science Foundation, Biomedical Advanced Research and Development Authority, and Prenosis.

Role of the Funder/Sponsor: Prenosis was overall responsible for the design and conduct of the study, collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The other funding agencies had no role.

Conflict of Interest Disclosures: Zhao, Zhu, Shapiro and Bashir are consultants to Prenosis. Bhargava, Lopez-Espina, Schmalz, Khan, Watson, Urdiales, Updike, and Reddy. Jr are employed by Prenosis. Bashir and Shapiro have equity ownership in Prenosis, and Bashir has equity interest in VedaBio. Dr. Shapiro is a consultant for Luminos technologies, Cambridge Medical Technologies, and receives research support from Bluejay diagnostics and Inflammatix.

Acknowledgement Statement: 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.
References


Figure Legends

Figure 1. Sepsis ImmunoScore Stratification for Sepsis-3 in all Cohorts. Barplots are shown for the derivation, internal validation, and external validation datasets for the Sepsis-3 within 24 hours PVs for each Sepsis ImmunoScore risk stratification category. Dashed lines indicate the 95% CIs.

Figure 2. Sepsis ImmunoScore Risk Stratification for Morbidity and Mortality (External Validation). Barplots are shown for the derivation, internal validation, and external validation datasets for the secondary endpoints PVs (ICU 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. Dashed lines indicate the 95% CIs.
# Tables

## Table 1. Baseline Data and Adverse Outcomes for Derivation, Internal Validation and External Validation

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Patients Encounters, No. (%)</th>
<th>Derivation</th>
<th>Internal Validation</th>
<th>External Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N = 2366)</td>
<td>(N = 393)</td>
<td>(N = 698)</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical Site</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beth Israel Deaconess Medical Center – Boston, MA</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>356 (51.0)</td>
<td></td>
</tr>
<tr>
<td>OSF – Peoria, IL</td>
<td>712 (30.1)</td>
<td>87 (22.1)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Jesse Brown VA - Chicago, IL</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>65 (9.3)</td>
<td></td>
</tr>
<tr>
<td>Mercy Health - St. Louis, MO</td>
<td>1061 (44.8)</td>
<td>306 (77.9)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Beaumont - Royal Oak, MI</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>277 (39.7)</td>
<td></td>
</tr>
<tr>
<td>Carle Foundation Hospital – Urbana, IL</td>
<td>593 (25.1)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Age (mean (SD))</strong></td>
<td>64.20 (16.59)</td>
<td>64.06 (17.66)</td>
<td>62.80 (17.01)</td>
<td></td>
</tr>
<tr>
<td><strong>Male (%)</strong></td>
<td>1195 (50.5)</td>
<td>210 (53.4)</td>
<td>391 (56.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>1 (0.0)</td>
<td>0 (0.0)</td>
<td>2 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>12 (0.5)</td>
<td>2 (0.5)</td>
<td>14 (2.0)</td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>315 (13.3)</td>
<td>57 (14.5)</td>
<td>154 (22.1)</td>
<td></td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>1 (0.1)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>85 (3.6)</td>
<td>12 (3.1)</td>
<td>119 (17.0)</td>
<td></td>
</tr>
<tr>
<td><strong>White</strong></td>
<td>1953 (82.5)</td>
<td>322 (81.9)</td>
<td>408 (58.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>26 (1.1)</td>
<td>2 (0.5)</td>
<td>96 (13.8)</td>
<td></td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>1725 (72.9)</td>
<td>385 (98.0)</td>
<td>564 (80.8)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>615 (26.0)</td>
<td>6 (1.5)</td>
<td>38 (5.4)</td>
<td></td>
</tr>
<tr>
<td><strong>High-Risk Comorbidities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute Myocardial Infarction (%)</td>
<td>97 (4.1)</td>
<td>11 (2.8)</td>
<td>43 (6.2)</td>
<td></td>
</tr>
<tr>
<td>History of Myocardial Infarction (%)</td>
<td>101 (4.3)</td>
<td>21 (5.3)</td>
<td>54 (7.7)</td>
<td></td>
</tr>
<tr>
<td>Congestive Heart Failure (%)</td>
<td>583 (24.6)</td>
<td>103 (26.2)</td>
<td>170 (24.4)</td>
<td></td>
</tr>
<tr>
<td>Peripheral Vascular Disease (%)</td>
<td>225 (9.5)</td>
<td>49 (12.5)</td>
<td>72 (10.3)</td>
<td></td>
</tr>
<tr>
<td>Cerebrovascular Disease (%)</td>
<td>130 (5.5)</td>
<td>38 (9.7)</td>
<td>65 (9.3)</td>
<td></td>
</tr>
<tr>
<td>Chronic Obstructive Pulmonary Disease (%)</td>
<td>606 (25.6)</td>
<td>107 (27.2)</td>
<td>171 (24.5)</td>
<td></td>
</tr>
<tr>
<td>Dementia (%)</td>
<td>167 (7.1)</td>
<td>45 (11.5)</td>
<td>72 (10.3)</td>
<td></td>
</tr>
<tr>
<td>Paralysis (%)</td>
<td>68 (2.9)</td>
<td>10 (2.5)</td>
<td>20 (2.9)</td>
<td></td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>630 (26.6)</td>
<td>101 (25.7)</td>
<td>156 (22.3)</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2. Sepsis ImmunoScore Risk Stratification for Sepsis-3 Within 24 Hours

<table>
<thead>
<tr>
<th>Cohort</th>
<th>ImmunoScore Risk Category</th>
<th>Total Patients</th>
<th>Septic Patients</th>
<th>Sepsis PV [95% CI]</th>
<th>Sepsis Likelihood Ratio [95% CI]</th>
<th>Cochrán-Armitage Test (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>External Validation (N = 698)</td>
<td>Low</td>
<td>232</td>
<td>7</td>
<td>3.0% [1.2%, 6.1%]</td>
<td>0.1 [0.1, 0.2]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>157</td>
<td>20</td>
<td>12.7% [7.96%, 19.0%]</td>
<td>0.5 [0.3, 0.8]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>276</td>
<td>101</td>
<td>36.6% [30.1%, 42.6%]</td>
<td>2.1 [1.8, 2.5]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Very High</td>
<td>33</td>
<td>23</td>
<td>69.7% [51.3%, 84.4%]</td>
<td>8.3 [4.1, 17.1]</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 3. External Validation ImmunoScore Risk Stratification for Morbidity and Mortality

<table>
<thead>
<tr>
<th>Secondary Outcome</th>
<th>Sepsis Risk Category</th>
<th>Total Patients</th>
<th>Patients with Event</th>
<th>Predictive Value [95% CI]</th>
<th>Likelihood Ratio [95% CI]</th>
<th>Days [95% CI]</th>
<th>Cochran-Armitage Test (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICU Transfer within 24 Hours</td>
<td>Low</td>
<td>232</td>
<td>11</td>
<td>4.7% [2.4%, 8.3%]</td>
<td>0.2 [0.1, 0.4]</td>
<td>-</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>157</td>
<td>20</td>
<td>12.7% [8.0%, 19.0%]</td>
<td>0.7 [0.4, 1.1]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>276</td>
<td>71</td>
<td>25.7% [20.7%, 31.3%]</td>
<td>1.7 [1.3, 2.1]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Very High</td>
<td>33</td>
<td>18</td>
<td>54.6% [36.4%, 71.9%]</td>
<td>5.8 [3.0, 11.3]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>In-Hospital Mortality</td>
<td>Low</td>
<td>232</td>
<td>0</td>
<td>0.0% [0.0%, 1.6%]</td>
<td>0.0 [0.0, -]</td>
<td>-</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>157</td>
<td>3</td>
<td>1.9% [0.4%, 5.5%]</td>
<td>0.4 [0.1, 1.2]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>276</td>
<td>24</td>
<td>8.7% [5.7%, 12.7%]</td>
<td>1.9 [1.3, 2.9]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Very High</td>
<td>33</td>
<td>6</td>
<td>18.2% [7.0%, 35.5%]</td>
<td>4.5 [1.9, 10.7]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Very High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>--------</td>
<td>-------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mechanical Ventilation within 24 Hours</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>232</td>
<td>157</td>
<td>276</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td>6</td>
<td>6</td>
<td>18</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Very High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vasopressor within 24 Hours</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>232</td>
<td>157</td>
<td>276</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td>2</td>
<td>3</td>
<td>32</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Very High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Length of Stay</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>232</td>
<td>157</td>
<td>276</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Very High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Vasopressor within 24 Hours

- Low: 0.9% [0.1%, 3.1%] 0.1 [0.0, 0.5] < 0.001
- Medium: 1.9% [0.4%, 5.5%] 0.3 [0.1, 0.8] -
- High: 11.6% [8.1%, 16.0%] 1.7 [1.2, 2.4] -
- Very High: 39.4% [22.9%, 57.9%] 8.4 [4.2, 16.7] -

Length of Stay

- Low: 232 - 4.0 [3.5, 4.9]
- Medium: 157 - 5.7 [4.9, 7.0]
- High: 276 - 7.7 [6.5, 8.5]
- Very High: 33 - 13.5 [7.1, 19.1]
The image consists of three bar charts, each comparing Sepsis ImmunoScore risk categories across different validation types: Derivation, Internal Validation, and External Validation.

- **Derivation**: The chart shows a significant increase in the percentage of patients in the Very High risk category compared to Low, Medium, and High categories.
- **Internal Validation**: Similar to Derivation, a noticeable increase in Very High risk category is observed, though the overall distribution remains similar.
- **External Validation**: The chart indicates a slight decrease in Very High risk category compared to Internal Validation, with a more balanced distribution among the other categories.

The Y-axis represents the percentage of patients in each risk category, ranging from 0% to 100%.