Early Physician Gestalt Versus Usual Screening Tools for the Prediction of Sepsis in Critically Ill Emergency Patients

Sarah K. S. Knack, MD; Nathaniel Scott, MD, MHA; Brian E. Driver, MD; Matthew E. Prekker, MD; MPH; Lauren Page Black, MD; Charlotte Hopson, MS; Ellen Maruggi, BS; Olivia Kaus, BS; Walker Tordsen, BS; Michael A. Puskarich, MD, MS*

*Corresponding Author. E-mail: mike.puskarich@hcmed.org.

Study objective: Compare physician gestalt to existing screening tools for identifying sepsis in the initial minutes of presentation when time-sensitive treatments must be initiated.

Methods: This prospective observational study conducted with consecutive encounter sampling took place in the emergency department (ED) of an academic, urban, safety net hospital between September 2020 and May 2022. The study population included ED patients who were critically ill, excluding traumas, transfers, and self-evident diagnoses. Emergency physician gestalt was measured using a visual analog scale (VAS) from 0 to 100 at 15 and 60 minutes after patient arrival. The primary outcome was an explicit sepsis hospital discharge diagnosis. Clinical data were recorded for up to 3 hours to compare Systemic Inflammatory Response Syndrome (SIRS), Sequential Organ Failure Assessment (SOFA), quick SOFA (qSOFA), Modified Early Warning Score (MEWS), and a logistic regression machine learning model using Least Absolute Shrinkage and Selection Operator (LASSO) for variable selection. The screening tools were compared using receiver operating characteristic analysis and area under the curve calculation (AUC).

Results: A total of 2,484 patient-physician encounters involving 59 attending physicians were analyzed. Two hundred seventy-five patients (11%) received an explicit sepsis discharge diagnosis. When limited to available data at 15 minutes, initial VAS (AUC 0.90; 95% confidence interval [CI] 0.88, 0.92) outperformed all tools including LASSO (0.84; 95% CI 0.82 to 0.87), qSOFA (0.67; 95% CI 0.64 to 0.71), SIRS (0.67; 95% 0.64 to 0.70), SOFA (0.67; 95% CI 0.63 to 0.70), and MEWS (0.66; 95% CI 0.64 to 0.69). Expanding to data available at 60 minutes did not meaningfully change results.

Conclusion: Among adults presenting to an ED with an undifferentiated critical illness, physician gestalt in the first 15 minutes of the encounter outperformed other screening methods in identifying sepsis. [Ann Emerg Med. 2024;■:1-13.]

Please see page XX for the Editor's Capsule Summary of this article.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2024.02.009

SEE EDITORIAL, P. XX.

INTRODUCTION

Sepsis represents an increasingly common syndrome, responsible for approximately 11 million deaths annually in the United States and up to 20% of all deaths globally.[1-3] In aggregate, sepsis represents one-third of all inhospital deaths, making it the leading cause of inhospital mortality.[4] The cost of sepsis care exceeds $17 billion annually, outpacing most acute care conditions.[5]

Timely recognition and management are critical to improve patient outcomes in sepsis.[6-9] In an attempt to improve these outcomes, the Centers for Medicare and Medicaid Services (CMS) introduced the "Severe Sepsis and Septic Shock Early Management Bundle" (commonly referred to as SEP-1) in 2015 for the evaluation of care for patients with severe sepsis and septic shock. Although compliance is publicly reported and may be tied to hospital compensation, adherence remains low.[10,11] Efforts to improve adherence have been hampered by delays in recognition and exacerbated by variable, syndromic definitions. Sepsis-1 guidelines defined sepsis as suspicion of infection plus 2 or more Systemic Inflammatory Response Syndrome (SIRS) criteria, whereas current Sepsis-3 consensus guidelines define sepsis as life-threatening organ dysfunction caused by a dysregulated host response to infection (defined operationally as suspicion for infection

Editor's Capsule Summary

What is already known on this topic

Existing screening tools for the identification of sepsis among undifferentiated patients early in the emergency department (ED) care are not robust.

What question this study addressed

How does physician gestalt compare to existing screening tools in the initial minutes of ED presentation for the suspicion of sepsis?

What this study adds to our knowledge

In this approximately 2,500-patient prospective study, physician gestalt at both 15 and 60 minutes after patient arrival outperformed existing screening tools in identifying sepsis.

How this is relevant to clinical practice

Using physician gestalt as a sepsis identification method could allow for more appropriate use of guideline-recommended time-sensitive therapies.

plus change in sequential organ failure assessment [SOFA] score of 2 or more).[12,13] Both of these definitions require laboratory tests, such as white blood cell count, platelet count, creatinine, or total bilirubin concentrations that can take up to an hour to result. To address this limitation, a screening tool, quick SOFA (qSOFA), was developed based on examination and vital signs alone to identify patients most likely to require intensive care or die, although its diagnostic accuracy has been questioned.[14] The Modified Early Warning Score is a tool that has been validated to identify patients at high risk of clinical deterioration that may require ICU and have a higher risk of mortality.[15] Although most of these screening tools were designed for prediction of mortality or adverse outcomes, their use has been extrapolated to screen for patients with sepsis in need of immediate care.

In response to the perception that misdiagnosis represents the primary cause of delays in treatment, many hospital systems have implemented manual or electronic health record screening tools to identify potential patients with sepsis and initiate the SEP-1 bundle of care. Some hospitals have used published measures such as SIRS, qSOFA, SOFA, and Modified Early Warning Score (MEWS), whereas others use complex machine learning and artificial intelligence algorithms.[16,17] The performance of these tools, similar to many other clinical decision aids, has not been compared with physician gestalt.[18] This is potentially problematic because subtle observations and deductions by physicians are not captured in the electronic medical record. To address this knowledge gap, we aimed to measure the accuracy of standardized screening tools and a machine learning model to predict a hospital discharge diagnosis of sepsis, compared with physician gestalt in the hyperacute period immediately after patient presentation among undifferentiated patients with critical illness in the emergency department (ED).

MATERIALS AND METHODS

Study Setting

This single-center prospective study recorded the treating physician's gestalt for sepsis in patients who were critically ill presenting with a medical condition. Screening tools, including SIRS, qSOFA, SOFA, and MEWS, were calculated retrospectively. The study was conducted from September 2020 through May 2022 at an academic urban safety net hospital in Minneapolis, MN, with more than 100,000 annual ED visits, approximately 7,000 of which present to the area under study. Our hospital has a specialized 4-bed resuscitation area to care for all patients requiring emergency interventions. Patients are triaged to this area when they are deemed critically ill based on clinical presentation, vital signs, or mental status. In the resuscitation area, patients have 2 dedicated ED nurses, a health care assistant, and 2 physicians who take reports from paramedics or triage staff and immediately initiate care. Providers are regularly trained regarding sepsis bundle elements, including use of a standardized order set. The study was approved by the institutional review board (Hennepin Healthcare Research Institute; IRB# 20_4848) before initiation. As this was a study of physician decision making, physicians were informed of the study through multiple modalities, given a study information form as well as the opportunity to opt out, but proceeded without documentation of informed consent. Observational data collection and review of medical records were considered to be of minimal risk and could not practically be completed without alteration of consent. This study was approved to proceed with the waiver of informed consent consistent with 45 CFR 46.116(f). This study adhered to the EQUATOR guidelines for strengthening the reporting of observational studies in epidemiology (see Figure E1, available at http://www.annemergmed.com).

Patient Population

The patient population included adult (aged 18 years or older) critically ill, undifferentiated medical patients

presenting directly to the ED resuscitation area as described above. Patients are triaged to this area when they are deemed critically ill by EMS or triage staff based on clinical presentation or unstable vital signs. Patients with trauma and those with obvious causes of illness defined as presentation for cardiac arrest, ST-elevation myocardial infarction, suspected cerebrovascular accident, and active labor were excluded. Patients were excluded if they were transferred from another hospital or were initially evaluated in another part of the ED. Patients evaluated in a clinic and sent to the hospital were eligible for inclusion.

Data Collection

Faculty emergency physicians completed a visual analog scale (VAS) indicating their suspicion of sepsis at 15 minutes or less and 60 (+/- 15) minutes from patient presentation. If a faculty physician's VAS score was unavailable, the senior resident (postgraduate year 2 or above) score was used. Trained independent observers presented the responsible physician with an iPad with a REDCap survey consisting of a single question, "What is the likelihood that this patient has sepsis?" and a slider bar spanning 0 (no infection) to 100 (infection) to record the physician's response. Due to the high acuity of cases with rapid assessment and intervention, observers collected real-time vital signs and interventional data on eligible patients from the time of their arrival in the resuscitation area. Vital signs included pulse rate, respiratory rate, blood pressure, and peripheral oxygen saturation, recorded at the time of arrival and 60 minutes into the encounter. The time that the following interventions occurred (if applicable) were recorded prospectively: oral and rectal temperature obtained, laboratory tests performed, bedside ultrasound completed, radiography completed, urine collected, antibiotics initiated, vasopressor initiated, and intravenous fluid bolus started (see Figure E2 for data collection form, available at http://www.annemergmed.com). Physicians also filled out a short information sheet on the presence of tender abdomen, skin ulcer, and any skin findings.

Trained chart abstractors (research coordinators) were blinded to the study hypothesis and reviewed the medical record and collected demographics, laboratory studies obtained in the ED, first Glasgow Coma Score, admission location, and International Classification of Diseases (ICD-10) diagnostic codes. Laboratory studies were abstracted to calculate standard sepsis screening tools and included the result time and value of white blood cell count, platelet count, hemoglobin, creatinine, bilirubin, and lactate concentrations. When unavailable or missing during bedside data collection, vital signs and intervention times were abstracted from the medical record. Tracheal intubation and mortality were collected via an automated abstraction for each subject out of the medical record and matched by encounter number (see Figure E3 for data collection form, available at http://www.annemergmed.com). Abstractors were blinded to VAS values. Study data were collected and managed using REDCap electronic data capture tools.[19,20]

Outcomes

The primary outcome of the study was an ICD-10 code for sepsis, as defined by CMS, at hospital discharge that was marked as present on arrival.[21] A preplanned sensitivity analysis used an alternative sepsis diagnostic criterion implicit sepsis where 49 combinations of ICD-10 diagnostic codes for infection and 28 codes to define organ failure were used to establish the presence of severe sepsis or septic shock per Angus et al.[22]

Statistical Analysis

Descriptive statistics were used to characterize the patients with and without a sepsis diagnosis. The number of positive qSOFA, SIRS, MEWS, and SOFA score criteria were determined by calculating the number of criteria meeting the relevant threshold for each score component, which was limited to data clinically available at each time point. If unavailable, components were assumed to be normal.

We constructed a machine learning model using Least Absolute Shrinkage and Selection Operator (LASSO) for variable selection with data split into 80% training and 20% testing sets. The split was stratified to ensure proportional groupings of septic and nonseptic cases in both the training and test sets. We imputed missing data at any given time as normal to be consistent with clinical practice (eg, if total bilirubin is not ordered, clinicians using a SOFA tool would assume it is normal, not estimate an abnormal level based on the average of other patients), but for statistical robustness, we constructed another model imputing median values for missing data points. Further details on variable labeling and imputation methodology are provided in the Supplement Methods (Appendix E1, available at http://www.annemergmed.com). Variable selection for multivariable logistic regression modeling was performed on the training data set using the LASSO variable selection process.[23] Continuous variables were scaled and centered, and all candidate variables were considered for modeling, including demographic variables (age, race, and sex), vital signs (pulse rate, oxygen

saturation, systolic blood pressure, diastolic blood pressure, temperature, respiratory rate, and Glasgow Coma Score), laboratory values (white blood cells, platelets, bands, hemoglobin, sodium, potassium, glucose, creatinine, lactate, troponin, bilirubin, and leukocytes and nitrites in the urinalysis), and skin findings (tender abdomen, skin ulcer presence, any skin findings). The target variable was an explicit sepsis diagnosis code. A priori, we decided to create time censored multivariable logistic regression models using LASSO variable selection based on data available at 15 and 60 minutes. With regard to model fit, we considered area under the curve (AUC), accuracy, sensitivity, and specificity on test sets but did not use them to make decisions about the model. Further details are provided in Appendix E1.

Receiver operating curves (ROC) were produced for the diagnostic accuracy of physician gestalt versus standard sepsis screening tools (qSOFA, SIRS, SOFA, and MEWS) and machine learning algorithms. AUC with 95% confidence intervals (CIs) are reported. We

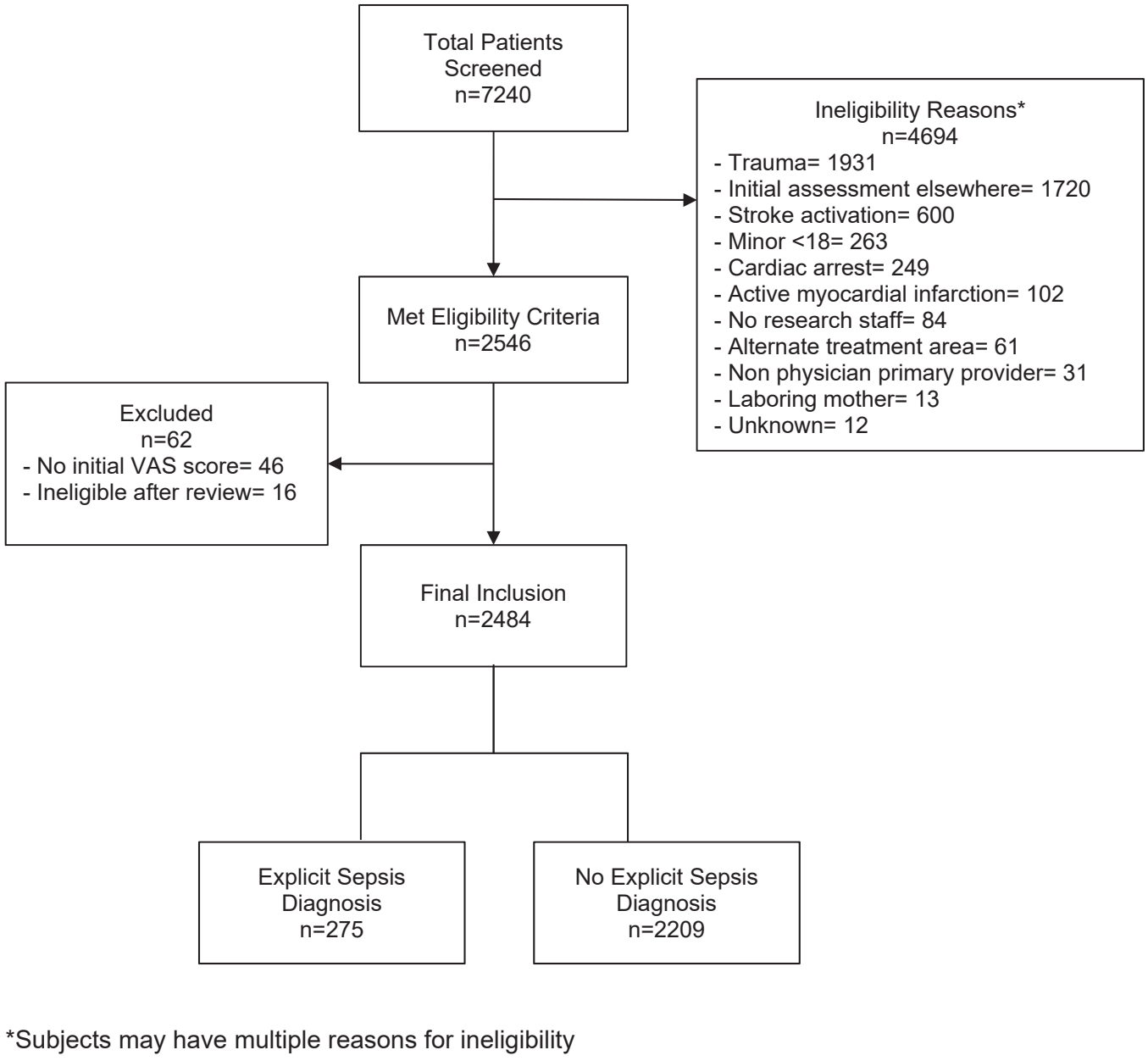

Figure 1. Study flow diagram. This diagram illustrates the flow of participants into this observational research study. Patients were initially screened and excluded were minors, patients presenting with trauma or obvious medical cause.

selected a clinically relevant VAS cutoff (more than 50%, or more likely than not) and a model optimized cutoff (45 at 15 minutes and 49 at 60 minutes) and compared the test characteristics at that threshold to qSOFA, SIRS, and SOFA scores at their respective thresholds (which are all 2 and greater). The MEWS score has a threshold of 5 for identifying patients at high risk of mortality and ICU needs; however, because there is no consensus threshold for this as a sepsis screening tool, we used a model optimized cut off (more than 2 at 15 minutes and more than 4 at 60 minutes). We evaluated the characteristics of the LASSO machine learning model using 2 thresholds: a cutoff of 50% probability to match our VAS cutoff and a model optimized cutoff (11% at both 15 minutes and 60 minutes). All optimized cutoffs were determined using the Youden index.[24] A predicted probability plot for administration of antibiotics was generated using a logistic regression model with splines to describe the relationship between VAS score and intravenous administration of antibiotics. Stata (StataCorp), version 15.1, was used for all data analyses with the exception of LASSO modeling, which was computed using R Studio and the *glmnet* package.[25,26]

Table 1. Clinical characteristics of subjects with and without explicit sepsis diagnosis.

Variable	Sepsis Diagnosis n= 275	No Sepsis Diagnosis n= 2209
Age, median (IQR) (y)	65 (52-75)	51 (33-65)
Male sex, no. (%)	158 (57)	1,320 (60)
Race, no. (%)		
Asian	6 (2)	45 (2)
Black, non-Hispanic	72 (26)	810 (37)
Hispanic	10 (4)	131 (6)
Native American or Native Alaskan	22 (8)	111 (5)
Pacific Islander and Native Hawaiian	0	2 (0.1)
White, non-Hispanic	148 (54)	864 (39)
Not reported	17 (6)	246 (11)
Vital signs, initial		
Systolic Blood Pressure (mmHg), median (IQR)	118 (96-140)	136 (119-153)
Pulse rate (beats/min), median (IQR)	114 (97-131)	100 (84-116)
Respiratory rate (beats/min), median (IQR)	22 (17-28)	19 (16-24)
Peripheral oxygen saturation (%), median (IQR)	95 (91-97)	97 (94-99)
Initial temperature (degrees C), median (IQR)	37.6 (36.4-38.6)	36.7 (36.4-37.0)
Glasgow coma scale, median (IQR)	14 (9-15)	15 (12-15)
Laboratory test results		
WBC (cellsx10^3/uL), median (IQR)	13.2 (9.4-18.2)	8.7 (6.6-11.7)
Creatinine (mg/dL), median (IQR)	1.6 (1.0-2.5)	1.1 (0.8-1.4)
Lactate (mmol/L), median (IQR)	3.3 (2.1-5.3)	2.4 (1.6-4.1)
Platelet count (cellsx10^3/uL), median (IQR)	226 (165-314)	252 (202-309)
Total Bilirubin (mg/dL), median (IQR)	0.6 (0.3-1.4)	0.4 (0.3-0.7)
Blood cultures obtained, no. (%)	224 (81)	482 (22)
Interventions		
Antibiotic administration, no. (%)	261 (95)	533 (24)
IV fluid administration in mL, median (IQR)	2,000 (1,500-3,000)	1,000 (0-1,923)
IV fluid administration of at least 1 L, no. (%)	235 (85)	1,366 (62)
Vasopressor administration, no. (%)	92 (33)	123 (6)
Intubation, no. (%)	119 (43)	378 (17)
Mortality, no. (%)	47 (17)	64 (3)

There was low missingness in our data with less than 2% in all variables except for bilirubin (61%), blood cultures obtained (18%).

Power and sample size. Based on our preliminary data, we estimated that approximately 10% of critically ill medical encounters would receive a discharge diagnosis of sepsis. Given an unknown distribution of VAS scores and desire to include enough clinically occult cases of sepsis, we planned to enroll approximately 250 patients with sepsis, or 2,500 patients overall, over the span of approximately 1 year. We calculated a priori that this would provide 96% power to detect a difference of 10 on the VAS scale between groups. Based on the observed cohort of 2,484 patients, at an alpha of 0.05, our study had >99%, 94%, and 80% power to detect differences in the AUC of 0.12, 0.1, and 0.08 between tests, respectively.

RESULTS

A total of 7,240 patient-physician encounters that took place in the ED resuscitation area were screened, and 2,484 met study eligibility. Most patients excluded were not in the intended study population, predominantly trauma patients (Figure 1). The median age was 53 (interquartile range [IQR] 35 to 67), and 60% were men. Two hundred seventy-five (11%) patients met the primary outcome of a sepsis ICD-10 discharge diagnosis marked as present on admission. Vital signs, laboratory studies, and key interventions are shown in Table 1. These same characteristics stratified by VAS status are shown in Table E1 (available at http://www.annemergmed.com). There was low missingness in our data, with less than 2% in all variables except for total bilirubin values (61%) and blood cultures obtained (18%).

The initial clinical attending physician VAS score within 15 minutes was obtained in 2,362 (94%) of cases, and the remaining 122 (6%) were initial resident physician VAS scores. Forty-eight attending physicians provided at least 5 VAS scores during the study, and a total of 59 attending physicians participated. Years of experience were well distributed among the attending physicians, with 18 (32%) having 0 to 3 years, 18 (32%) 4 to 10 years, and 22 (38%) at least 11 years. The median VAS score in the cohort of patients with sepsis was 81 (IQR 57 to 100) and 8 (IQR 0 to 26) in those without sepsis. The distribution of scores, stratified by attending and outcome, is illustrated in Figure E4, (available at http://www.annemergmed.com). We observed a strong correlation between initial (less than 15 minutes) VAS score and the likelihood of administering antibiotics in the ED (Figure 2).

Data for complete calculation of each of the screening tools were available for 2,484 (100%) patients for qSOFA, 1,461 (59%) for MEWS, 171 (7%) for SIRS, and 50 (2%) for SOFA at 15 minutes, compared with 2,484 (100%), 2,168 (87%), 1,595 (64%), and 157 (6%), respectively, at 1 hour. These data reflect real practice of what data are available when critical assessments must be made. The availability of various pieces of clinical data, as well as common interventions over time, are illustrated in Figure 3.

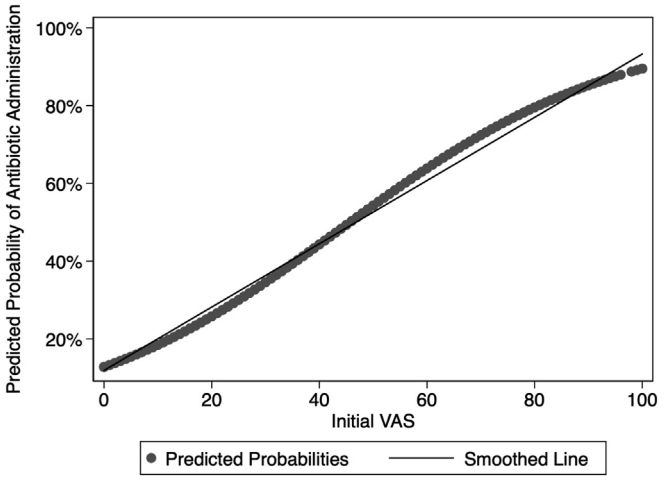

Figure 2. Predicted probability of antibiotic administration by initial VAS. Predictions calculated using logistic regression modeling with splines.

When limited to clinical data available at 15 minutes, initial VAS (AUC 0.90; 95% CI 0.88 to 0.92) significantly outperformed all tools, including LASSO (0.84; 95% CI 0.82 to 0.87), qSOFA (0.67; 95% CI 0.64 to 0.71), SIRS (0.67; 95% CI 0.64 to 0.70), SOFA (0.67; 95% CI 0.63 to 0.70), and MEWS (0.66; 95% CI 0.64 to 0.69). All 95% CI for initial VAS by attending were overlapping, showing no significant differences. The ROC curves are presented in Figure 4A. Broadening to data available at 60 minutes improved all tools' performance, though VAS maintained superior characteristics with ROC curves displayed in Figure 4B.

The test characteristics to predict the primary outcome at the predefined cutoffs at each time point are presented in Table 2. Both the clinically relevant cutoff of 50% and an optimized cutoff are presented for both the gestalt VAS score and the LASSO model. A sensitivity analysis was performed using the Angus definition of either implicit or explicit sepsis, and performance of models was similar to ROC curves presented in Figure E5, (available at http://www.annemergmed.com).

To better understand select cases where physician gestalt failed to recognize a patient subsequently diagnosed with sepsis, we conducted a manual chart review and provided a descriptive case series of patients with an initial VAS in the bottom 25th percentile of all assessments (coincident with a VAS = 0), who were ultimately assigned an explicit sepsis diagnosis present on arrival (10 patients, Table 3).

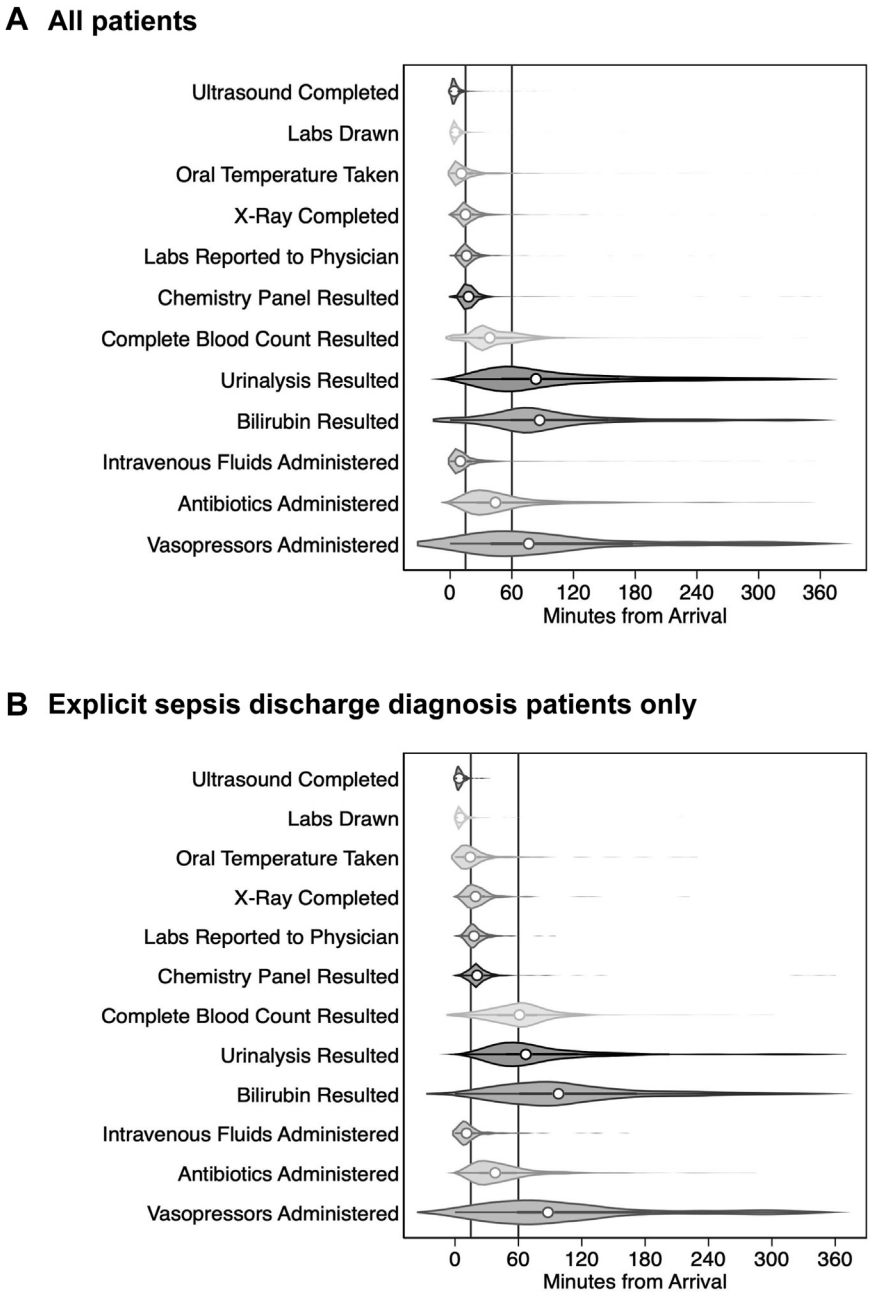

Figure 3. Violin plot illustrating the time to various ED diagnostic test results and interventions. Laboratory test results reported to physician are defined as the first time any paper results were handed to the clinician from a laboratory technician (venous blood gas and chemistry results). Complete blood cell count, chemistry panel, bilirubin, and urinalysis resulted times were defined based on electronic medical record timestamps. Treatment administered times were defined as time of initiation. Drop lines are added at 15 and 60 minutes to benchmark and compare panels. *A,* All patients. *B,* Explicit sepsis discharge diagnosis patients only.

LIMITATIONS

Our study has several limitations. First and foremost, these findings are only relevant to patients who are critically ill, and relative performance of gestalt in less severely ill or occult sepsis cases remains an area for future investigation. Relatedly, these results from a single academic medical center's specialized resuscitation area may not be generalizable to other practice settings, particularly in an era of ED crowding. With time to laboratory results of approximately 1 hour from arrival, these data represent an idealized setting, and we expect the comparison screening tools that rely on laboratory or radiology results to perform even worse in usual care settings. Importantly, although individual physician performance could have affected our

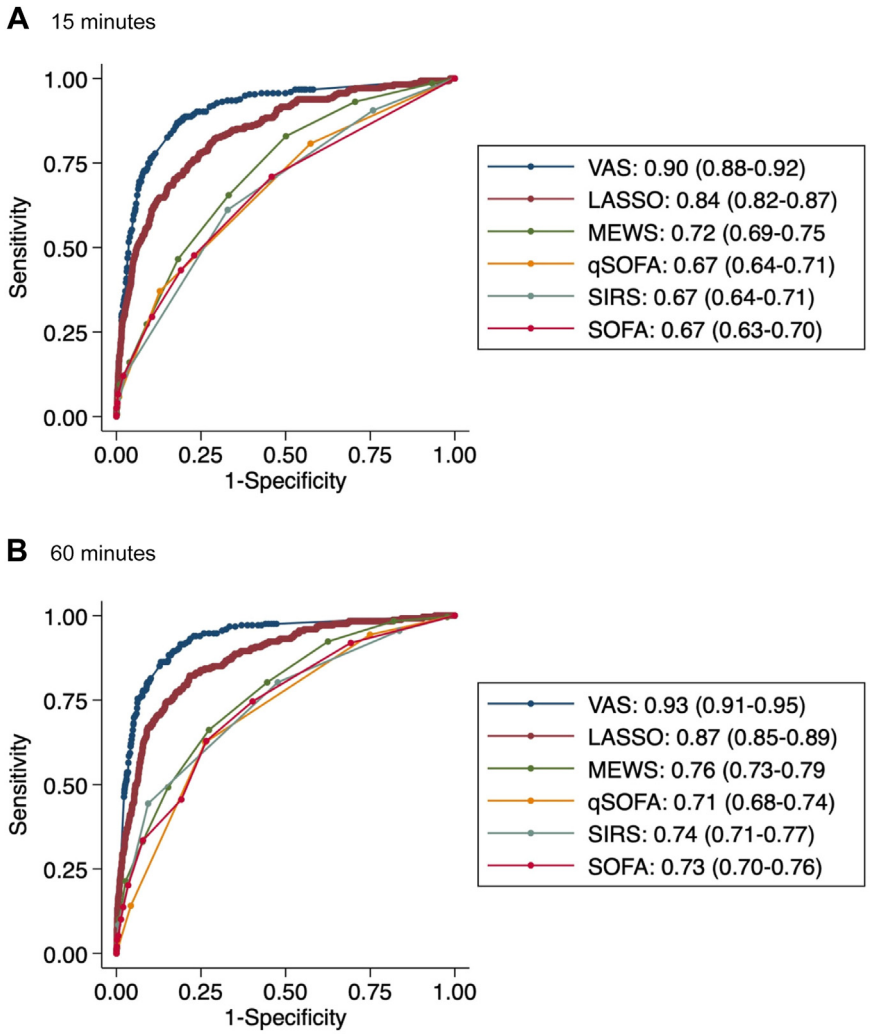

Figure 4. ROC curves comparing each of the screening tools at 15 and 60 minutes. Legends with AUC (95% CI). *A,* 15 minutes. *B,* 60 minutes.

results, the fact that approximately 50 different physicians provided at least 5 unique assessments mitigates this concern, in addition to the observation that there were no significant differences in the distribution of scores or AUC by physician. Given the syndromic definition of sepsis, alternative definitions of sepsis may also affect our findings. However, our results are highly relevant to quality improvement efforts toward improving SEP-1 compliance and were consistent when applied to implicit sepsis definitions. Using ICD-10 codes for diagnosis could be viewed as a limitation of this work; however, this is commonly used in performance metrics to assess compliance. Additionally, direct comparisons to SIRS and SOFA are in some ways nonsensical and self-referential because physician suspicion of infection is still required to diagnose sepsis. Other tools were not specifically designed for sepsis; however, their use for such purposes in some clinical practices makes this comparison relevant. Despite these limitations and lacking another true comparator, we maintained these comparisons as these tools, for better or worse, are often implemented as screening tools in various clinical settings. Although our machine learning model incorporated all data abstracted for this study, performance of nested tools that incorporate the patient's medical history and other electronic health record data points may have yielded different results and may be particularly relevant in inpatient settings when more longitudinal data and trends are available. Finally, it is possible our results suffered from a Hawthorne effect because the fact that physicians were queried as to their suspicion of sepsis could have increased that suspicion.

DISCUSSION

In this large single-center study of consecutively sampled, undifferentiated medical patients who were

Table 2. Test characteristics of physician gestalt on a visual analog scale and screening tools at 15 and 60 minutes.

	Cutoff Value	Sepsis n = 275	No Sepsis n = 2,209	Sensitivity	Specificity	PPV	NPV
15 Min		n (%)	n (%)	(95% CI)	(95% CI)	(95% CI)	(95% CI)
Gestalt VAS positive, no. (%)*	>50	227 (83)	333 (15)	82.5 (77.5-86.8)	84.9 (83.4-86.4)	40.5 (36.4-44.7)	97.5 (96.7-98.2)
Optimized Gestalt VAS positive, no. (%)*	>45	239 (87)	398 (18)	86.9 (82.3-90.7)	82.0 (80.3-83.6)	37.5 (33.7-41.4)	98.1 (97.3-98.6)
qSOFA positive, no. (%)	≥2	141 (51)	456 (21)	51.3 (45.2-57.3)	79.4 (77.6-81.0)	23.6 (20.3-27.2)	92.9 (91.6-94.0)
SIRS positive no. (%)	≥2	168 (61)	753 (34)	61.1 (55.1-66.9)	65.9 (63.9-67.9)	18.2 (15.8-20.9)	93.2 (91.8-94.4)
SOFA positive, no. (%)	≥2	152 (55)	624 (28)	55.3 (49.2-61.2)	71.8 (69.8-73.6)	19.6 (16.9-22.6)	92.8 (91.5-94.0)
MEWS positive, no. (%)	>2	228 (83)	1,107 (50)	82.9 (77.9-87.2)	49.9 (47.8-52.0)	17.1 (15.1-19.2)	95.9 (94.6-97.0)
LASSO positive, no. (%)*	>50	63 (23)	31 (1)	22.9 (18.1-28.3)	98.6 (98.0-99.0)	67.0 (56.6-76.4)	91.1 (89.9-92.2)
Optimized LASSO positive, no. (%)*	>11	209 (76)	502 (23)	76.0 (75.5-79.0)	77.3 (75.5-79.0)	29.4 (26.1-32.9)	96.3 (95.3-97.1)
60 Min							
Gestalt VAS positive, no. (%)*	>50	238 (96) N = 248	612 (33) N = 1831	85.1 (80.0-89.3)	87.2 (85.6-88.7)	47.4 (42.7-52.2)	97.7 (96.9-98.4)
Optimized Gestalt VAS positive, no. (%)*	>49	241 (97) N = 248	619(34) N = 1831	86.3 (81.4-90.3)	86.8 (85.2-88.4)	47.0 (42.4-51.7)	97.9 (97.1-98.5)
qSOFA positive, no. (%)	≥2	177 (64)	624 (28)	64.4 (58.4-70.0)	71.8 (69.8-73.6)	22.1 (19.3-25.1)	94.2 (92.9-95.2)
SIRS positive, no. (%)	≥2	225 (82)	1194 (54)	81.8 (76.7-86.2)	45.9 (43.9-48.1)	15.9 (14.0-17.9)	95.3 (93.9-96.5)
SOFA positive, no. (%)	≥2	194 (71)	795 (36)	70.5 (64.8-75.9)	64.0 (62.0-66.0)	19.6 (17.2-22.2)	94.6 (93.3-95.7)
MEWS positive, no. (%)	>4	178 (65)	598 (27)	64.7 (58.8-70.4)	72.9 (71.0-74.8)	22.9 (20.0-26.1)	94.3 (93.1-95.4)
LASSO positive, no. (%)*	>50	81 (30)	42 (2)	29.5 (24.1-35.2)	98.1 (97.4-98.6)	65.9 (56.8-74.3)	91.8 (90.6-92.9)
Optimized LASSO positive, no. (%)*	>11	222 (81)	461 (21)	80.7 (75.6-85.2)	79.1 (77.4-80.8)	32.5 (29.0-36.2)	97.1 (96.2-97.8)

*The cutoff point for a positive screen was 50% for VAS, given the face validity of 50% as being more likely than not. The LASSO machine learning algorithm is presented with characteristics at an empiric cutoff of 50%. Diagnostically optimized cutoffs using Youden Index are presented for the VAS and LASSO model as well. *PPV*, Positive predictive value; *NPV*, Negative predictive value.

Table 3. Clinical summary of patients assessed with a VAS of 0 at the 15-minute time point and subsequently assigned an explicit sepsis discharge diagnosis.

Clinical Summary	Antibiotics in the ED	Intubated ED/ICU	qSOFA	SIRS	SOFA	LASSO	VAS at 60 mins
71-year-old female presenting with hypoxia and respiratory distress. She was discharged from hospital the day prior on antibiotics after a 3-day hospitalization for Klebsiella bacteremia from a urinary source. The patient was admitted, and intravenous antibiotics were continued. She was discharged back to her facility the next day.	Yes	Yes	1	2	6	9.4	0
49-year-old female with a history of ulcerative colitis presenting with hypotension and gastrointestinal bleed. She was found to have an extensive amount of ischemic bowel that was resected. She remained unstable and in septic shock. Patient discharged after a month-long hospitalization.	Yes	Yes	0	0	3	2.7	N/A
30-year-old female presenting with acute respiratory failure requiring intubation. Throughout her hospital course, she was thought to have a combination of pneumonia, pulmonary edema, and obesity hypoventilation syndrome. She was discharged to subacute rehabilitation 2 weeks later.	Yes	Yes	2	2	7	10.8	37
64-year-old male with cirrhosis presenting with apnea and acute respiratory failure. There was concern for aspiration pneumonia leading to sepsis. He was discharged home after completing antibiotics and had a 9-day hospital stay.	Yes	Yes	0	2	6	2.8	0
69-year-old male with presumed aspiration in the setting of severe alcohol intoxication. His vital signs were concerning for sepsis from this respiratory source. He was discharged home 3 days later on oral antibiotics.	Yes	Yes	1	1	7	10.9	N/A
65-year-old male with cerebral palsy presenting with altered mental status who underwent septic workup and was found to have an atypical lung infection. He received COVID vaccine one day prior. Patient discharged 4 days later and completed an outpatient course of oral antibiotics.	Yes	No	2	3	1	37.8	73
57-year-old male with history of vasculitis, chronic kidney disease, abdominal aortic thrombosis with superior mesenteric artery occlusion and small bowel resection, right hemicolectomy, short bowel syndrome, high ostomy output who presented with acute hypoxic respiratory failure. He was found to have newly diagnosed heart failure and acute pulmonary edema. He was also diagnosed with central line associated bloodstream infection and had positive blood cultures. He completed a course of antibiotics and was discharged back to his facility after 9 days.	Yes	Yes	2	3	5	15.4	65

66-year-old male with history of end stage renal disease, diabetes, and a recent COVID infection who presented with acute hypoxic respiratory failure due to flash pulmonary edema thought to be related to volume overload and aortic stenosis. He was also found to have pneumonia on further imaging. Patient discharged on hospital day 1 and completed an outpatient course of oral antibiotics.	Yes	No	1	2	1	9	11.5	100
66-year-old male with history of hepatitis C and bipolar disorder who presented with agitation, altered mental status, and melena thought to be related to decompensated cirrhosis and a gastrointestinal bleed. He was found to have spontaneous bacterial peritonitis and pneumonia. He was treated with a course of broad-spectrum antibiotics. Patient was extubated on hospital day 8 and subsequently suffered a cardiac arrest on hospital day 13 and died.	Yes	Yes	2	1	9	4.6	0	

critically ill, we collected real-time data regarding clinical processes and physician gestalt of sepsis not captured in the electronic medical record or prior studies of sepsis screening tools. We found that physician gestalt within 15 minutes of ED arrival outperformed other sepsis screening methods, including a machine learning algorithm, with superior performance persisting at 1 hour. These data advance the current literature by directly comparing physician gestalt at the point-of-care to accepted sepsis screening approaches in a real-time, prospective fashion. Although most of these tools were not designed specifically for sepsis screening, they are commonly used for that purpose in clinical practice and served as the basis to use them as a comparison in this study. Furthermore, all consensus definitions of sepsis call for "clinical suspicion of infection," which has historically been poorly defined and relied on surrogate process metrics, such as the ordering of blood cultures or administration of intravenous antibiotics.

These data have several implications. First, delays in diagnosis or bundle compliance are unlikely to be improved on by implementing electronic medical record-based screening tools, at least in this idealized setting when patients are recognized to be critically ill by out-of-hospital or ED triage personnel. Future studies intending to study screening tools ought to be benchmarked against or in addition to physician judgment. Without this step, novel tools may lead to additional testing and practice changes without improvement in diagnostic accuracy. However, an opportunity for improvement of screening methodologies was demonstrated by identification of a subset of cases where certain screening tools identified patients missed by physicians. Further investigating how to best combine screening tools to minimize missed cases (improve sensitivity) would serve as a valuable area for future investigations.

Second, our data demonstrate it is critical that screening and diagnostics be evaluated at a time point during which time-sensitive clinical decisions related to the administration of intravenous fluid and antibiotics are made, with consideration of real-world test turnaround times. This study also illustrates the delays associated even with common, relatively rapid turnaround blood tests in an idealized setting where patients were immediately identified as critically ill and brought to a resuscitation area with a dedicated treatment team consisting of multiple nurses and physicians. In patients who are less severely ill, we expect the time for laboratory results to be even more delayed than in this study. Based on our data, to be truly efficacious, screening tools that outperform gestalt will either need to: a) avoid the need for blood draws, b) function through point-of-care testing, ideally using microfluidics that can be run on a fingerstick sample, or c) consist of innovative

subcutaneous sensors using technology now commonly used in diabetes, but made capable with an array of diagnostic tests.

Third, we unexpectedly observed a nearly linear relationship between physician suspicion of sepsis as measured by a VAS and propensity to administer intravenous antibiotics. These data support the concept that physicians conceptualize sepsis treatment on a continuum rather than a simple present or absent diagnosis. This observation has implications related to the design of interventions to improve early administration of timely therapies, which should augment rather than replace physician gestalt. As current tools seem to underperform physicians as a whole, screening tools should be applied after an assessment of gestalt to help aid in the identification of "tail" or "missed" cases. Based on these data, there remains a risk that blind implementation of these screening tools may underperform physician judgment, leading to indiscriminate antibiotic use, overzealous fluid resuscitation, and early diagnostic closure unless properly developed and tested.

In conclusion, this single-center study, physician gestalt within 15 minutes of patient arrival in the ED outperformed usual screening tools in identifying sepsis among critically ill, undifferentiated medical patients. Novel sepsis screening tools should incorporate physician gestalt.

Supervising editor: Alan E. Jones, MD. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From the Hennepin Healthcare (Knack, Scott, Driver, Prekker, Maruggi, Kaus, Tordsen, Puskarich), Minneapolis, MN; University of Florida, College of Medicine (Black), Jacksonville, FL; Northwestern University (Black), Feinberg School of Medicine, Chicago, IL; the University of Florida (Hopson), Gainesville, FL; and University of Minnesota (Puskarich), Minneapolis, MN.

Author contributions: MAP, BD, NS, MEP conceived and designed the study. MAP, NS, and BD supervised the conduct of the trial and data collection. EM, OK, and WT undertook data collection form production, data entry, and management including quality control. LB and CH provided statistical assistance with the machine learning model portion of the project. SK took the lead in data cleaning and statistical analysis. SK, MAP, and NS drafted the manuscript, and all authors contributed substantially to its revision. SK and MAP take responsibility for the paper as a whole.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data sharing statement: Partial or complete data sets and data dictionaries are available from one year after publication upon request to Dr. Puskarich at mike.puskarich@hcmed.org to investigators who provide an IRB letter of approval.

Funding and support: By *Annals'* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

Publication dates: Received for publication November 27, 2023. Revision received February 2, 2024. Accepted for publication February 14, 2024.

Presentation information: Data presented at SAEM annual conference, Austin, TX, May 18, 2023.

REFERENCES

1. Wardi G, Tainter CR, Ramnath VR, et al. Age-related incidence and outcomes of sepsis in California, 2008-2015. *J Crit Care*. 2021;62:212-217.
2. 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.
3. Fleischmann C, Scherag A, Adhikari NKJ, et al. Assessment of global incidence and mortality of hospital-treated sepsis. current estimates and limitations. *Am J Respir Crit Care Med*. 2016;193:259-272.
4. CDC. Technical resources & guidelines. Centers for disease control and prevention. Published August 23, 2023. Accessed November 22, 2023. https://www.cdc.gov/sepsis/clinicaltools/index.html
5. Hajj J, Blaine N, Salavaci J, et al. The "Centrality of Sepsis": a review on incidence, mortality, and cost of care. *Healthcare (Basel)*. 2018;6:90.
6. Seymour CW, Gesten F, Prescott HC, et al. Time to treatment and mortality during mandated emergency care for sepsis. *N Engl J Med*. 2017;376:2235-2244.
7. Baghdadi JD, Brook RH, Uslan DZ, et al. Association of a care bundle for early sepsis management with mortality among patients with hospital-onset or community-onset sepsis. *JAMA Intern Med*. 2020;180:707-716.
8. 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.
9. Lee SJ, Ramar K, Park JG, et al. Increased fluid administration in the first three hours of sepsis resuscitation is associated with reduced mortality: a retrospective cohort study. *Chest*. 2014;146:908-915.
10. Peach BC, Ng BP. Organizational factors associated with sepsis bundle compliance: a nationwide study. *Health Serv Res*. 2020;55:82-83.
11. Barbash IJ, Davis BS, Yabes JG, et al. Treatment patterns and clinical outcomes after the introduction of the medicare sepsis performance measure (SEP-1). *Ann Intern Med*. 2021;174:927-935.
12. Kaukonen KM, Bailey M, Pilcher D, et al. Systemic inflammatory response syndrome criteria in defining severe sepsis. *N Engl J Med*. 2015;372:1629-1638.
13. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. 2016;315:762-774.

14. Serafim R, Gomes JA, Salluh J, et al. A comparison of the quick-SOFA and systemic inflammatory response syndrome criteria for the diagnosis of sepsis and prediction of mortality: a systematic review and meta-analysis. *Chest*. 2018;153:646-655.
15. Subbe CP, Kruger M, Rutherford P, et al. Validation of a modified early warning score in medical admissions. *QJM*. 2001;94:521-526.
16. Shafaf N, Malek H. Applications of machine learning approaches in emergency medicine; a review article. *Arch Acad Emerg Med*. 2019;7:34.
17. Patton MJ, Liu VX. Predictive modeling using artificial intelligence and machine learning algorithms on electronic health record data: advantages and challenges. *Crit Care Clin*. 2023;39:647-673.
18. Schriger DL, Elder JW, Cooper RJ. Structured clinical decision aids are seldom compared with subjective physician judgment, and are seldom superior. *Ann Emerg Med*. 2017;70:338-344.e3.
19. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
20. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377-381.
21. ICD-10-CM/PCS MS-DRGv33 Definitions Manual. Accessed September 11, 2023. https://www.cms.gov/icd10manual/version33-fullcode-cms/fullcode_cms/P0328.html
22. Angus DC, Linde-Zwirble WT, Lidicker J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med*. 2001;29:1303-1310.
23. Taddy M. One-step estimator paths for concave regularization. *J Comput Graph Stat*. 2017;26:525-536.
24. Steyerberg EW, Harrell FE Jr, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774-781.
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1-22.
26. StataCorp L. Stata statistical software. College Station, TX: StataCorp: LP; 2017. Version 15. Accessed February 15, 2024.