Title

Evaluation of evolving sepsis screening criteria in discriminating suspected sepsis and mortality among adult patients admitted to the intensive care unit

Abstract

Background: Institutions struggle with successful use of sepsis alerts within electronic health records.

Objective: Test the association of sepsis screening measurement criteria in discrimination of mortality and detection of sepsis in a large dataset.

Design: Retrospective, cohort study using a large United States (U.S.) intensive care database. The Institutional Review Board exempt status was obtained from Kansas University Medical Center Human Research Protection Program (10-1-2015).

Setting: 334 U.S. hospitals participating in the eICU Research Institute

Participants: Nine hundred twelve thousand five hundred and nine adult intensive care admissions from 183 hospitals.

Methods: Exposures included: systemic inflammatory response syndrome criteria ≥ 2 (Sepsis-1); systemic inflammatory response syndrome criteria with organ failure criteria ≥ 3.5 points (Sepsis-2); and sepsis-related organ failure assessment score ≥ 2 and quick score ≥ 2 (Sepsis-3). Discrimination of outcomes were determined with/without (adjusted/unadjusted) baseline risk exposure to a model. The receiver operating characteristic curve (AUROC) and odds ratios (ORs) for each decile of baseline risk of sepsis or death were assessed.

Results: Within the eligible cohort of 912,509, a total of 86,219 (9.4%) patients did not survive their hospital stay and 186,870 (20.5%) met the definition of suspected sepsis. For suspected sepsis discrimination, Sepsis-2 (unadjusted AUROC 0.67, 99% CI: 0.66-0.67 and adjusted AUROC 0.77, 99% CI: 0.77-0.77) outperformed Sepsis-3 (SOFA unadjusted AUROC 0.61, 99%

CI: 0.61-0.61 and adjusted AUROC 0.74, 99% CI: 0.74-0.74), (qSOFA unadjusted AUROC 0.59, 99% CI: 0.59-0.60 and adjusted AUROC 0.73, 99% CI: 0.73-0.73). Sepsis-2 also outperformed Sepsis-1 (unadjusted AUROC 0.58, 99% CI: 0.58-0.58 and adjusted AUROC 0.73, 99% CI: 0.73-0.73). In between differences of AUROCs were statistically significantly different. Sepsis-2 ORs were higher for the outcome of suspected sepsis when considering deciles of risk than the other measurement systems.

Conclusions and Relevance: Sepsis-2 outperformed other systems in suspected sepsis detection and was comparable to SOFA in prognostic accuracy of mortality in adult intensive care patients.

What is already known

- Sepsis remains a leading cause of morbidity and mortality globally and early identification of sepsis has been cited as one of the biggest obstacles to timely therapeutic interventions.
- Nurses are essential to early identification of sepsis efforts.
- SOFA and qSOFA scoring systems have not been adequately compared to algorithms using a combination of systemic inflammatory response syndrome and organ failure criteria

What this paper adds

- An algorithm using modified Sepsis-2 screening criteria that includes expanded systemic inflammatory response syndrome and organ failure criteria detects suspected sepsis among adult intensive care patients as well as or better than traditional systemic inflammatory response syndrome criteria of 2 or more (Sepsis-1) or SOFA and qSOFA scores of 2 or greater (Sepsis-3).

- By combining additional physiologic measurements with organ failure criteria, more sophisticated algorithms can detect potential for sepsis with greater accuracy.

Introduction

Sepsis is a difficult to detect and costly syndrome that is one of the leading causes of death globally. An estimated 48.9 million cases of sepsis occurred worldwide in 2017 with 11.0 million sepsis-related deaths/19.7% of all deaths globally. [1] According to recent Medicare claims data (2012-2018), in U.S. hospitals sepsis rates have increased by 40%, sepsis cost burdens range from $27.7 to $41.5 billion annually, and higher risks of mortality, complications, readmissions, and resource utilization are realized when comparing sepsis to other diagnoses. [2-4] Early identification of sepsis is crucial to timely action, treatment, and control of systemic inflammation that leads to later stages of decompensation. [2, 5]

In 1992, four traditional systemic inflammatory response syndrome criteria (Sepsis-1) were used to define sepsis and in 2001 an international team of experts introduced additional diagnostic criteria that could be used with systemic inflammatory response syndrome criteria (Sepsis-2).[6-9] By 2012, sepsis was defined as “a systemic, deleterious host response to infection leading to severe sepsis (acute organ dysfunction secondary to documented or suspected infection) and septic shock (severe sepsis plus hypotension not reversed with fluid resuscitation)”.[10] In 2016 sepsis-related organ failure assessment (SOFA) score of 2 or more in intensive care unit (ICU) patients or a quick SOFA (qSOFA) in non-ICU patients with an infection was introduced with a new definition: a life-threatening organ dysfunction caused by a dysregulated host response to an infection leading to tissue injury and organ failure (Sepsis-3).[11, 12]

Nurses make up the largest number of the healthcare workforce and are the primary care providers in hospitals.[13] This, along with their practice of continual assessment makes

nurses essential in identifying patients with infections, organ failure, and sepsis.[14, 15] Nurses are often the first-line recipients of electronic algorithms aimed at detecting suspected sepsis.[16, 17] Recently, nurse experts recommended use of SOFA and qSOFA scoring for sepsis screening over systemic inflammatory response syndrome, National Early Warning Score (NEWS), or the Modified Early Warning Score (MEWS).[14] Prior to 2016, automated electronic sepsis alerts used a combination of systemic inflammatory response syndrome and organ failure criteria to identify sepsis versus systemic inflammatory response syndrome or organ failure alone.[17]

A body of literature related to development of automated electronic sepsis alerts is growing but the accuracy and reliability of these algorithms fall short.[18-23] Standardization of these alerts is lacking, and concerns related to alert fatigue are mounting.[20, 24-26] Many recent studies of ICU and non-ICU patients have demonstrated that SOFA and qSOFA scores do not demonstrate adequate precision in early detection of sepsis.[27-33] Johnson et al. (2018) found that despite clear advantages of using Sepsis-3 criteria in identification of sepsis, SOFA identified a larger, higher risk of mortality and less “pure” cohort of septic patients than use of administrative coded or Centers for Disease Control and Prevention criteria.[34] More recently the Surviving Sepsis Guidelines cite strong evidence against use of qSOFA for use in early detection of sepsis and wide variation in predictive accuracy of other tools.[35] While automated sepsis alerting has been shown to be beneficial,[36] additional examination of criteria from common sepsis measurement systems is needed to support development of more effective sepsis screening systems.

The overall objective of this study was to conduct a retrospective analysis using a large U.S. data repository to determine if an algorithm using modified Sepsis-2 screening criteria that includes expanded systemic inflammatory response syndrome and organ failure (OF) criteria can detect suspected sepsis among adult ICU patients as well as or better than traditional systemic inflammatory response syndrome criteria of 2 or more (Sepsis-1) or a SOFA and

qSOFA score of 2 or more (Sepsis-3). Prognostic accuracy of each measurement system for mortality for patients with and without suspected sepsis was also analyzed.

Methods

Study Design, Data Sources, and Population

A retrospective cohort study was performed in a large critical care clinical database known as the Philip's eICU Research Institute known as the eRI database.[37] More than 400 ICUs in the U.S. generate data through use of the Philips eCareManager® enterprise telehealth software application into this database. The eICU Research Institute database was used for this study was extensive (> 2 million cases in years 2010-2015) and as with most large data sets used for secondary analysis, was "uncleaned" when it was received. Massachusetts Institute of Technology Laboratory of Computational Physiology research partners, and co-authors of this publication, assisted in data extraction and preparation. A description of data extraction and management can be found in the *supplementary appendix.*

The Philip's eICU Research Institute database, used in this study, has been independently certified as meeting Health Insurance Portability and Accountability Act (HIPAA) safe harbor standards. The study was conducted according to the 1964 Helsinki declaration and its later amendments. The Institutional Review Board exempt status was obtained from Kansas University Medical Center Human Research Protection Program (10-1-2015). Many Institutional Review Boards recognize that de-identified data sets that are available for secondary analysis do not constitute human subjects research as defined at 45 CFR 46.102.

The eICU Research Institute complete dataset and the publicly available subset known as the eICU Collaborative Research Database (https://eicu-crd.mit.edu/)[38] were made available through the work of Philips Healthcare and collaborators at Massachusetts Institute of Technology's Laboratory for Computational Physiology. Philips eICU Research Institute

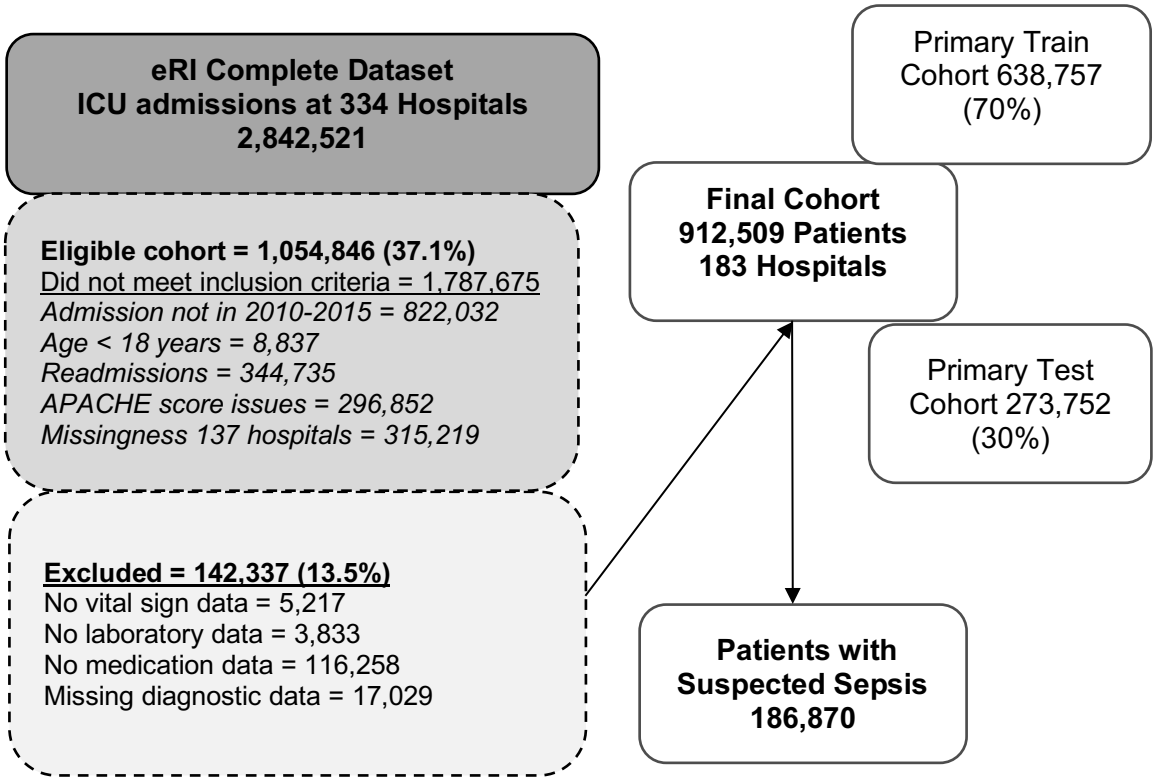
publications committee approved use of the complete data and Massachusetts Institute of Technology approved use of the publicly available subset. Analysis code for this study is available on GitHub13, and the dataset is made available on PhysioNet.[39, 40] Data scientists from Massachusetts Institute of Technology and Philips Healthcare worked together to extract data from the eICU Research Institute database using Structured Query Language (commonly referred to as "SQL"). Due to the size of the complete dataset, code in R statistical programming language was verified against the eICU Collaborative Research Database prior to applying it the eICU Research Institute dataset.[41]

All patient records admitted to the ICU between January 1, 2010 and December 31, 2015 were screened for study inclusion. Hospitals can participate with Tele-critical care centers using the Philips eCareManager® system at varying levels of service and as such may or may not participate in certain aspects of data collection and extraction; therefore, hospitals with data in the following relational database tables were used to determine inclusion: Acute Physiology Age Physiology Health Evaluation score (APACHE), medications, vital sign, and laboratory. APACHE scores are generated from many demographic and physiologic measures, as well as diagnosis, collected during the first 24 h of a patient's ICU stay[2]. Additional inclusion criteria were 18 years of age or older and only the first ICU admission during the same hospitalization.

Eligibility and the sequential order in which exclusion criteria were applied for the cohorts are described in Figure 1. When empty data tables for pertinent data needed for this study (vital sign, laboratory, medication, diagnosis) were realized at the hospital or patient level, we could not assume missing at random; therefore imputation methods were not used.[42] The reasons that entire data tables are found to be empty in the eICU Research Institute database and eICU Collaborative Research Database are described in the *supplementary appendix*. When data are missing not at random, introduction of bias with multiple imputation may be as great or greater than the bias in analyses of complete cases.[42] Given this we made the decision to include only

cases and hospitals with adequate evidence of vital sign, laboratory, medication or diagnostic data.

Figure 1. Eligible Population and Explanation of Cohorts



eRI, eICU Research Institute; ICU, intensive care unit; APACHE, acute physiology, age, chronic health evaluation; IVa version of APACHE algorithm

The APACHE scoring system is widely recognized as a good, predictive tool for in-hospital mortality and length of stay for patients in critical care and APACHE IV was consider a major update to the APACHE scoring system.[43, 44] The APACHE IVa version was a moderately minor but value-added update to the previous IV version.[45] APACHE IVa scoring issues were found in 297,000 patients. According to the Cerner APACHE® Foundations User Guide used during the years applicable to this study, reasons for APACHE score issues include no acute physiology score (APS), less than six hour ICU stays, most burn and transplant patients are not scored, no hospital mortality data, ICU to ICU transfers, and ICU discharge destination is another ICU.[46]

Lastly many ICUs board non-ICU level patients and as such do not collect APACHE data on boarding patients.

Only patients 18 years of age or greater were included and only cases that were not classified as ICU readmissions. At the patient level 142,337 patients were excluded for empty data tables for vital sign, laboratory, medication, or diagnostic data. After applying all inclusion/exclusion criteria we compared the characteristics of all available cases in the eICU Research Institute dataset for the designated period to the final cohort (912,509). In terms of demographics, relevant features, and outcomes we found the two populations to be similar. Further discussion related to decisions related to inclusion/exclusion criteria can be found in the *supplementary appendix*.

Frequency of Missing Data among Clinical and Laboratory Values

Complete patient information existed for most variables, however for chronic comorbidities we treated patients with missing or unknown status as “no”. For all other variables listed above (BMI, gender, ethnicity, admission source, hospital level information) with missing information we coded as “other/unknown”. For vital sign data, temperature had the highest number of missing values with 17,387 (< 2%) followed by SBP missing 184 values (< 1%) missing values; missing values were coded as normal. There were no missing heart rate, MAP, or respiratory values and only 9,882 (1%) had missing GCS. This can be explained as most participating sites collect APACHE IVa data and these values, are required for APACHE scoring. Missing laboratory values in the cohort: creatinine 88,585 (10%); bilirubin 534,627 (59%); lactate 697,840 (76%); partial pressure of oxygen in arterial blood ($PaO_{2)}$ 566,288 (62%); partial pressure of carbon dioxide in arterial blood ($PaCO_2$) 566,412 (62%); platelets 121,367 (13%); international normalized ratio (INR) 547,794 (60%); activated partial thromboplastin clotting time (aPTT) 645,624 (71%); white blood cell (WBC) 117,929 (13%); Bands 836,281 (92%); pH

576,985 (63%); base deficit 853,461 (94%); aspartate aminotransferase (AST) 528,607 (58%); alanine transaminase (ALT) 533,653 (58%); and albumin 548,672 (60%).

Patient and hospital level demographic information was missing at varying levels. There were 40,393 patients classified as other/unknown for ethnicity, 53 patients were missing information related to gender while another 170 were classified as other or unknown and 53 patients were missing ICU discharge disposition. There were 32,286 patients without BMI calculations due to missing heights and/or weights. There were 37,915 patients at hospitals classified as unknown for teaching status and 55,141 were missing hospital region information.

Determining Clinical Criteria for Sepsis Measurement Systems

Data cutoffs for variables for the SOFA and qSOFA scoring (Sepsis-3) and systemic inflammatory response syndrome criteria (Sepsis-1) were determined and baseline SOFA scores were assigned for three chronic health conditions using methodology consistent with other researchers .[11, 47] A net SOFA score ≥ 2 and a qSOFA score ≥ 2 were labeled positive. Patients with ≥ 2 systemic inflammatory response syndrome criteria met the threshold for systemic inflammatory response syndrome positive. Additional information related to SOFA (*eTable 1* in the *supplementary appendix*), qSOFA and systemic inflammatory response syndrome can be found in the *supplementary appendix*. Sepsis-2 was defined by expanded systemic inflammatory response syndrome and organ failure criteria (*eTable 2* in the *supplementary appendix*) and used to develop a sepsis screening algorithm in 2008 using Fuzzy Logic.[48]

To replicate the reasoning of care providers, Fuzzy Logic can be used to develop computer-based algorithms based on degree of correctness versus “absolutely true” or “absolutely false” decision trees.[49] Machine learning algorithms for glycemic control (artificial pancreas), decision-making algorithms in mechanical ventilation for patients with respiratory distress syndromes, and diagnostic algorithms used in implantable cardioverter defibrillators have been developed

using Fuzzy Logic.[50-52] For the development of the Sepsis-2 score, Fuzzy Logic was applied to expanded systemic inflammatory response syndrome criteria with most variables contributing to the score within a range of partial point values to a value of one based on how far they deviate from normal.

A total of 2.5 points was needed to meet the threshold for the inflammation criteria of the Sepsis-2 score. Each inflammation criterion (*eTable 2* in the *supplementary appendix*) was assigned zero points if normal and escalated from partial to 1 full point in a linear fashion based on how far values deviated from normal. Each organ failure criteria below were worth one point. After the inflammation threshold was met, one organ failure point was needed for a positive sepsis alert score. Organ failure criteria were dichotomous (1 point or 0). Additional information related to the Sepsis-2 score can be found in the *supplementary appendix*.

Defining Suspected Sepsis and Mortality Outcomes

The eICU Research Institute data set contains a subset of data integrated from electronic health records and other health information systems as well as data entered directly into the eCareManager® system within the first 24 hours of an ICU stay by trained critical care team members as part of APACHE and quality improvement data collection. Each entry of the problem list and the APACHE diagnosis in the eCareManager® system is time and date stamped. This allowed researchers to examine discrete data elements from a working problem list and the APACHE diagnosis for each patient. The use of documented diagnostic data provided a lens into clinical decision-making during the time of interest. Final administrative data coded at discharge were not used in this study.

Using Angus et al. (2000) [53] and Martin et al. (2001) [54] classifications we defined suspected sepsis as either having a severe sepsis or a septic shock diagnosis, or an infection diagnosis with an acute organ failure diagnosis recorded in the eCareManager® system problem list or listed as the APACHE diagnosis (primary reason for ICU admission) within the

24-hour period following admission to the ICU (*eTable 3 in the supplementary appendix*). This logic is consistent with Seymour et al. (2016) management of non-EHR datasets in the Sepsis-3 study where researcher used Angus ICD-9-CM diagnosis codes or prospective screening data to identify infection present on admission.[11]

The APACHE diagnosis was used for accurate classification of a patient's primary diagnosis or reason for ICU admission. The 448 unique APACHE diagnoses were categorized into groups using the ANZICS Adult Patient Database Data Dictionary for Software Programmers.[55] After selection for inclusion, records underwent a binary classification process to label them as suspected sepsis or not suspected sepsis. In-hospital mortality was defined as deceased at hospital discharge.

Statistical Analysis

Baseline models were developed using logistic regression to support risk-adjusted analysis for the outcomes of suspected sepsis and mortality. These models were constructed using available information at the time of ICU admission and variables were consistent with baseline model variables. Age, gender, body mass index, ethnicity, ICU admission source, physician specialty (critical care versus non-critical care), hospital size, hospital discharge year, and comorbid conditions (dialysis, aids, hepatic failure, diabetes, immunosuppression, leukemia, lymphoma, metastatic cancer, and selected cardiovascular and respiratory conditions), and use of thrombolytic therapy prior to ICU admission use were included in baseline models (adjusted analysis). Models were trained on a randomly selected subset of 70% of the eligible cohort and performance was assessed on the remaining 30%. The discriminatory capacities of each measurement system for suspected sepsis and mortality were measured.

Given that in such a large sample size a statistically significant Hosmer-Lemeshow test does not necessarily indicate that a predictive model is not useful,[56] we analyzed calibration

graphs of observed vs. predicted within deciles for suspected sepsis and mortality for each measurement system. The calibration graphs for suspected sepsis and mortality prediction were analyzed for agreement between observed endpoints and predictions and were found to be consistent.

Discrimination tests were useful for measuring the performance of prognostic algorithms and classification systems.[57] Discrimination for suspected sepsis and in-hospital mortality was assessed using the area under the receiver operator curve (AUROC) for each measurement system. Differences between AUROCs for each measurement system were evaluated individually (unadjusted analysis) and in conjunction with baseline risk models (adjusted analysis) using the method of Delong.[58]

Adjusted Odds Ratios (AOR) for suspected sepsis and Confidence Intervals (CI) were estimated on the training cohort and prediction and performance analyses (AUROC, sensitivity, specificity, PPV, and NPV) were performed on the testing cohort. As with Seymour et al. (2016), AUROCs would be considered to be poor at 0.6 to 0.7, adequate at 0.7 to 0.8, good at 0.8 to 0.9, and excellent at 0.9 or higher.[11, 59] Lastly, the strength of association between the measurement system and the outcomes were assessed under deciles of baseline risk for each outcome using the baseline-only models. All reported p-values were two-sided, and statistical significance was assessed at the 0.01 level.

To assess robustness of the model the analysis included: 1) suspected sepsis versus non-suspected sepsis patients, 2) mortality for patients with and without suspected sepsis (primary cohort) and mortality for patients with suspected sepsis (secondary cohort), and 3) using adjusted model (baseline risk) versus unadjusted analyses.

Results

Cohorts and Encounter Characteristics

Data pertaining to ICU admissions in 459 hospitals across the U.S. were recorded in the

eICU Research Institute Database for the period of 2010-2015. Of the final cohort of 912,509 patients from 183 hospitals, 186,870 (20.5%) were classified with suspected sepsis. As expected, the randomly assigned training (70%) and testing (30%) cohorts were balanced in their clinical and demographic characteristics. Comparison of patient and hospital level demographic and comorbid conditions data, measurement systems, illness severity, outcome, and diagnostic data for patients with and without suspected sepsis and survivors and non-survivors, are presented in Table 1.

Table 1. Comparison of Demographic, Measurement Systems, Illness Severity, Outcome, and Diagnostic Data

Characteristic	Non-Sepsis	Sepsis	Survivors	Non-Survivors
No.	725639	186870	826290	86219
Age, mean (SD)	62.2 (17.3)	65.7 (16.2)	62.3 (17.2)	69.5 (15.0)
Male, No. (%)	395985 (54.6)	94548 (50.6)	444766 (53.8)	45767 (53.1)
Ethnicity, No. (%)				
Caucasian	553821(76.3)	141546 (75.7)	629231 (76.2)	66136 (76.7)
African American	84902 (11.7)	20390 (10.9)	96168 (11.6)	9124 (10.6)
Hispanic	30536 (4.2)	10857 (5.8)	37359 (4.5)	4034 (4.7)
Asian	9221 (1.3)	2474 (1.3)	10468 (1.3)	1227 (1.4)
Native American	5248 (0.7)	1517 (0.8)	6085 (0.7)	680 (0.8)
Other	41911 (5.8)	10086 (5.4)	46979 (5.7)	5018 (5.8)
BMI No. (%)				
0-18.5	31764 (4.4)	12245 (6.6)	37672 (4.6)	6337 (7.3)
18.5-25	202170 (27.9)	55468 (29.7)	230063 (27.8)	27575 (32.0)
25-35	338933 (46.7)	77589 (41.5)	381728 (46.2)	34794 (40.4)
> 35	125888 (17.3)	36166 (19.4)	148643 (18.0)	13411 (15.6)
Other	26884 (3.7)	5402 (2.9)	28184 (3.4)	4102 (4.8)
ICU Admit Source No. (%)				
Floor	107913 (14.9)	46723 (25.0)	131068 (15.9)	23568 (27.3)
OR/Procedural	166553 (23.0)	9890 (5.3)	169828 (20.6)	6615 (7.7)
Direct Admit	80562 (11.1)	17476 (9.4)	87805 (10.6)	10233 (11.9)
ED	351730 (48.5)	104745 (56.1)	415282 (50.3)	41193 (47.8)
Other	5693 (0.8)	2070 (1.1)	6616 (0.8)	1147 (1.3)
SDU	13188 (1.8)	5966 (3.2)	15691 (1.9)	3463 (4.0)
Specialty Critical Care No. (%)	194821 (26.8)	72415 (38.8)	235008 (28.4)	32228 (37.4)
Hosp Discharge Year No. (%)				

2010	88588 (12.2)	22942 (12.3)	100104 (12.1)	11426 (13.3)
2011	95007 (13.1)	27029 (14.5)	109790 (13.3)	12246 (14.2)
2012	119084 (16.4)	30085 (16.1)	134912 (16.3)	14257 (16.5)
2013	133576 (18.4)	33722 (18.0)	151717 (18.4)	15581 (18.1)
2014	142947 (19.7)	35240 (18.9)	162398 (19.7)	15789 (18.3)
2015-2016	146437 (20.2)	37852 (20.3)	167369 (20.3)	16920 (19.6)
Teaching Hospital No. (%)	217937 (30.0)	58615 (31.4)	248468 (30.1)	28084 (32.6)
Hospital Size No. (%)				
Unknown	58989 (8.1)	13689 (7.3)	66107 (8.0)	6571 (7.6)
<100	26993 (3.7)	9779 (5.2)	34747 (4.2)	2025 (2.3)
100-249	163833 (22.6)	43284 (23.2)	190062 (23.0)	17055 (19.8)
250-500	131717 (18.2)	35396 (18.9)	151081 (18.3)	16032 (18.6)
>500	344107 (47.4)	84722 (45.3)	384293 (46.5)	44536 (51.7)
US Region No. (%)				
Midwest	313401 (43.2)	69674 (37.3)	352325 (42.6)	30750 (35.7)
Northeast	46163 (6.4)	27360 (14.6)	63922 (7.7)	9601 (11.1)
South	229437 (31.6)	54550 (29.2)	256647 (31.1)	27340 (31.7)
West	90846 (12.5)	25937 (13.9)	102962 (12.5)	13821 (16.0)
Unknown	45792 (6.3)	9349 (5.0)	50434 (6.1)	4707 (5.5)
Comorbid Conditions No. (%)				
Dialysis	22883 (3.2)	7751 (4.1)	26798 (3.2)	3836 (4.4)
AIDS	418 (0.1)	472 (0.3)	766 (0.1)	124 (0.1)
Hepatic Failure	14231 (2.0)	4834 (2.6)	15833 (1.9)	3232 (3.7)
Diabetes	159066 (21.9)	40750 (21.8)	183788 (22.2)	16028 (18.6)
Immunosuppression	13906 (1.9)	7475 (4.0)	17783 (2.2)	3598 (4.2)
Leukemia	4157 (0.6)	2427 (1.3)	5313 (0.6)	1271 (1.5)
Lymphoma	2413 (0.3)	1201 (0.6)	3042 (0.4)	572 (0.7)
Metastatic CA	12984 (1.8)	4523 (2.4)	14269 (1.7)	3238 (3.8)
Respiratory	156269 (21.5)	63588 (34.0)	195414 (23.6)	24443 (28.3)
Cardiovascular	161435 (22.2)	45771 (24.5)	183305 (22.2)	23901 (27.7)
Admitted with Myocardial Infarction and received Thrombolytics No. (%)	16351 (2.3)	384 (0.2)	15352 (1.9)	1383 (1.6)
SIRS ≥ 2 No. (%)	527901 (72.7)	165953 (88.8)	613583 (74.3)	80271 (93.1)
qSOFA ≥ 2 No. (%)	460152 (63.4)	152864 (81.8)	534943 (64.7)	78073 (90.6)
SOFA ≥ 2 No. (%)	467451 (64.4)	160765 (86.0)	546746 (66.2)	81470 (94.5)
Sepsis 2 ≥ 3.5 No. (%)	350438 (48.3)	152070 (81.4)	427824 (51.8)	74684 (86.6)
APACHE IVa (mean (SD))	52.0 (24.0)	69.5 (28.0)	51.90 (22.17)	90.36 (31.77)
Hospital Mortality No. (%)	51602 (7.1)	34617 (18.5)	0 (0.0)	86219 (100.0)
Hospital LOS (mean (SD))	7.0 (8.7)	10.3 (12.2)	7.69 (9.37)	7.97 (11.80)

ICU LOS (mean (SD))	2.8 (3.8)	4.2 (5.1)	2.93 (3.85)	4.49 (5.84)
Suspected Sepsis No. (%)	0 (0.0)	186870 (100.0)	152253 (18.4)	34617 (40.2)
APACHE Group No. (%)				
Cardiovascular	275329 (37.9)	20019 (10.7)	270546 (32.7)	24802 (28.8)
GI	83127 (11.5)	11546 (6.2)	87106 (10.5)	7567 (8.8)
Gynecological	2293 (0.3)	117 (0.1)	2381 (0.3)	29 (0.0)
Hematologic	5761 (0.8)	1058 (0.6)	6220 (0.8)	599 (0.7)
Metabolic	68815 (9.5)	6050 (3.2)	73481 (8.9)	1384 (1.6)
Musculoskeletal & Skin	9185 (1.3)	2270 (1.2)	10969 (1.3)	486 (0.6)
Neurologic	113662 (15.7)	9248 (4.9)	111809 (13.5)	11101 (12.9)
Renal/GU	17170 (2.4)	4955 (2.7)	20549 (2.5)	1576 (1.8)
Respiratory	89707 (12.4)	46348 (24.8)	118873 (14.4)	17182 (19.9)
Sepsis	14033 (1.9)	83565 (44.7)	79999 (9.7)	17599 (20.4)
Trauma	39475 (5.4)	1020 (0.5)	37405 (4.5)	3090 (3.6)
Undefined	7082 (1.0)	674 (0.4)	6952 (0.8)	804 (0.9)

No., number; SD, standard deviation; SIRS, systemic inflammatory response syndrome; qSOFA, quick SOFA; SOFA, sepsis-related organ failure assessment; Sepsis-2, modified algorithm based on sepsis-2 expanded systemic inflammatory response syndrome and organ failure criteria; APACHE, acute physiology age, chronic health, evaluation; ICU, intensive care unit; LOS, length of stay; GI, gastrointestinal; GU, genitourinary; Suspected Sepsis, All p-values were significant at the < 0.001 with the exception of values denoted with *.

Within the primary cohort 86,219 (9.5%) did not survive hospitalization. Of the 186,870 patients with suspected sepsis 34,617 (18.5%) did not survive hospitalization, 88.8% had ≥ 2 systemic inflammatory response syndrome criteria, 86.0% had SOFA positive and 81.8% had qSOFA positive net scores of ≥ 2, and 81.4% met the threshold for the Sepsis-2 algorithm. A higher percentage of suspected sepsis patients who did not survive hospitalization had a positive SOFA score (96.0%) versus meeting systemic inflammatory response syndrome criteria (95.5%), a positive qSOFA score (93.1%), or meeting the threshold for Sepsis-2 (93.1%). The Sepsis-2 threshold was less likely to be met in patients without suspected sepsis (48.3%) and in survivors (78.7%) versus negative scores for qSOFA (63.4% and 79.2% respectively) and SOFA (64.4% and 83.8% respectively), and negative systemic inflammatory response syndrome criteria (72.7% and 87.3% respectively).

Sensitivity, Specificity, negative predictive value, and positive predictive value for Each Measurement System

The sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) can be visualized in *eTable 4 in the supplementary appendix*. Systemic inflammatory response syndrome criteria of two or more had the highest sensitivity (89.0%) and the lowest specificity (27.1%) for predicting suspected sepsis with positive predictive value (PPV) 20.4% and negative predictive value 90.5%. The Sepsis 2 algorithm had a lower sensitivity score (81.6%) than SOFA (86.1%) but was more specific than SOFA (51.8% versus 35.6% respectively) and PPV for Sepsis-2 (30.2%) was higher than SOFA (25.7%) and Sepsis-2 NPV (91.6%) was also slightly higher than SOFA (90.9%). qSOFA demonstrated a lower sensitivity (82.0%) and PPV (25.1%) with a higher specificity (36.6%) than SOFA and the lowest NPV 88.7% when compared to the other measurement systems.

Systemic inflammatory response syndrome criteria of two or more had the lowest specificity (25.6%) with a sensitivity of 93.1% for predicting in-hospital mortality and PPV 11.5% and NPV 97.3%. The Sepsis 2 algorithm had lower sensitivity (86.5%) than SOFA (94.3%) but higher specificity (48.2%) than SOFA versus (33.8%) respectively. Sepsis-2 PPV was higher (14.8%) than SOFA (12.9%) while NPV for SOFA was higher (98.3%) than Sepsis-2 (97.2%) for predicting in-hospital mortality. qSOFA had the lowest sensitivity score (90.6%) for predicting in-hospital mortality. Specificity for qSOFA (35.2%) with PPV (12.7%) and NPV (97.3%).

Adjusted Odds Ratios (AOR) and Confidence Intervals (CIs) for Measurement Systems

The association between each measurement system's threshold and suspected sepsis was estimated after adjusting for baseline risk factors. Each system was positively associated with a suspected sepsis after the baseline risk adjustment, including two or more systemic inflammatory response syndrome criteria (AOR 2.85, 99% CI: 2.80 - 2.91), qSOFA (AOR 2.36,

99% CI: 2.32 - 2.40), SOFA (AOR 3.21, 99% CI: 3.15 - 3.26), and Sepsis-2 algorithm (AOR 4.46, 99% CI: 4.39 – 4.53). Each system was also positively associated with in-hospital mortality: two or more systemic inflammatory response syndrome criteria (AOR 4.71, 99% CI: 4.56 – 4.87), qSOFA (AOR 4.60, 99% CI: 4.47 – 4.73), SOFA (AOR 7.54, 99% CI: 7.28 - 7.82), and Sepsis-2 algorithm (AOR 5.81, 99% CI: 5.67 – 5.95).

Area Under the Receiver Operator Curve (AUROC) for Measurement Systems

Discrimination of suspected sepsis was significantly higher using the Sepsis-2 algorithm (AUROC 0.667, 99% CI: 0.664-0.669) than two or more systemic inflammatory response syndrome criteria (AUROC 0.581, 99% CI: 0.579-0.583), qSOFA (AUROC 0.593, 99% CI: 0.590-0.595), and SOFA (AUROC 0.609, 99% CI: 0.607-0.611); these differences were visually and statistically significant (Figure 2 and Table 2). Similarly, when considered along with baseline risk prediction of suspected sepsis (adjusted analysis), the Sepsis-2 algorithm (AUROC 0.771, 99% CI: 0.768-0.773) outperformed two or more systemic inflammatory response syndrome criteria (AUROC 0.730, 99% CI: 0.728- 0.733), qSOFA (AUROC 0.729, 99% CI: 0.726-0.732), and SOFA (AUROC 0.740, 99% CI: 0.737-0.743) in discrimination of suspected sepsis (Figure 2 and Table 2).

Visually, the Sepsis-2 algorithm (noted as SIRS+OF in Figure 2) odds ratios are higher for the outcome of sepsis than SOFA, qSOFA and two or more systemic inflammatory response syndrome criteria (noted as SIRS in Figure 2). This indicates better discrimination of sepsis by the Sepsis-2 algorithm when taking the baseline risk of sepsis into account. The bell-shaped appearance of the Sepsis-2 algorithm (noted as SIRS+OF in Figure 2) indicates escalating odds of sepsis until about the 6th decile when the odds ratios begin to decline. In deciles 1-3 SOFA and systemic inflammatory response syndrome odds ratios decrease and then increase in deciles 4-6 after which odds ratios decrease. Increasing odds ratios are seen deciles 1-5 for

qSOFA with a plateau across deciles 5-7 followed by increases in deciles 8 and 9 and then a decrease in decile 10.

Figure 2. Odds Ratios for Sepsis Comparing Encounters with Positive Measurement System Criteria or Scores (SIRS, qSOFA, SOFA, and SIRS+OF) for Training Cohort (All 638,757) across Deciles of Risk.

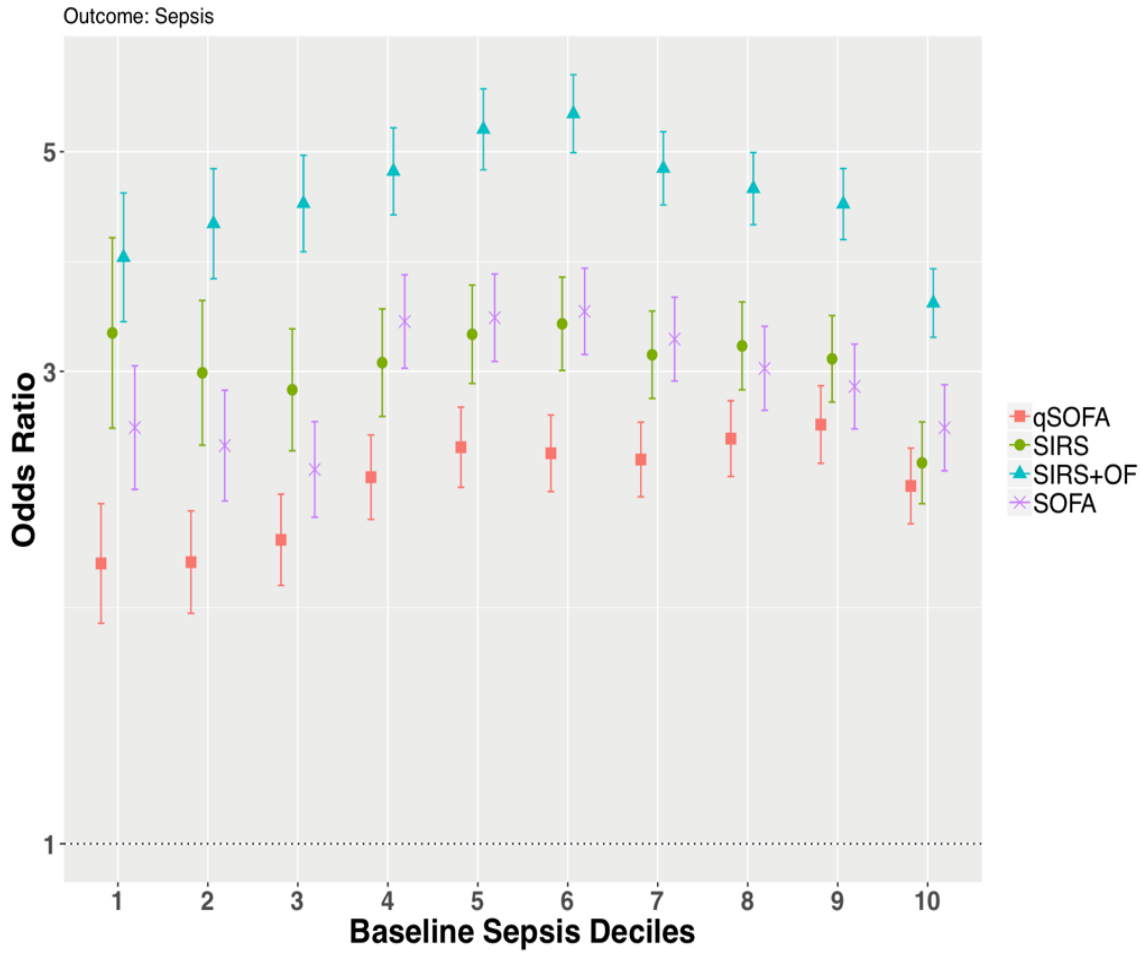


Table 2. Unadjusted and Adjusted AUROCs for Discrimination Characteristics of Measurement Systems Among Adult ICU Patients in the Testing Cohorts (273,752).

Characteristic	SIRS	qSOFA	SOFA	Sepsis-2
In-Hospital Sepsis				
Unadjusted AUROC	0.581	0.593	0.609	0.667
(99% CI)	(0.579-0.583)	(0.590-0.595)	(0.607-0.611)	(0.664-0.669)
Adjusted AUROC	0.730	0.729	0.740	0.771
(99% CI)	(0.728-0.733)	(0.726-0.732)	(0.737-0.743)	(0.768-0.773)
In-Hospital Mortality				
Unadjusted AUROC	0.593	0.629	0.641	0.673
(99% CI)	(0.591-0.596)	(0.626-0.632)	(0.639-0.643)	(0.670-0.676)
Adjusted AUROC	0.742	0.751	0.759	0.777
(99% CI)	(0.739-0.746)	(0.747-0.754)	(0.756-0.763)	(0.773-0.780)

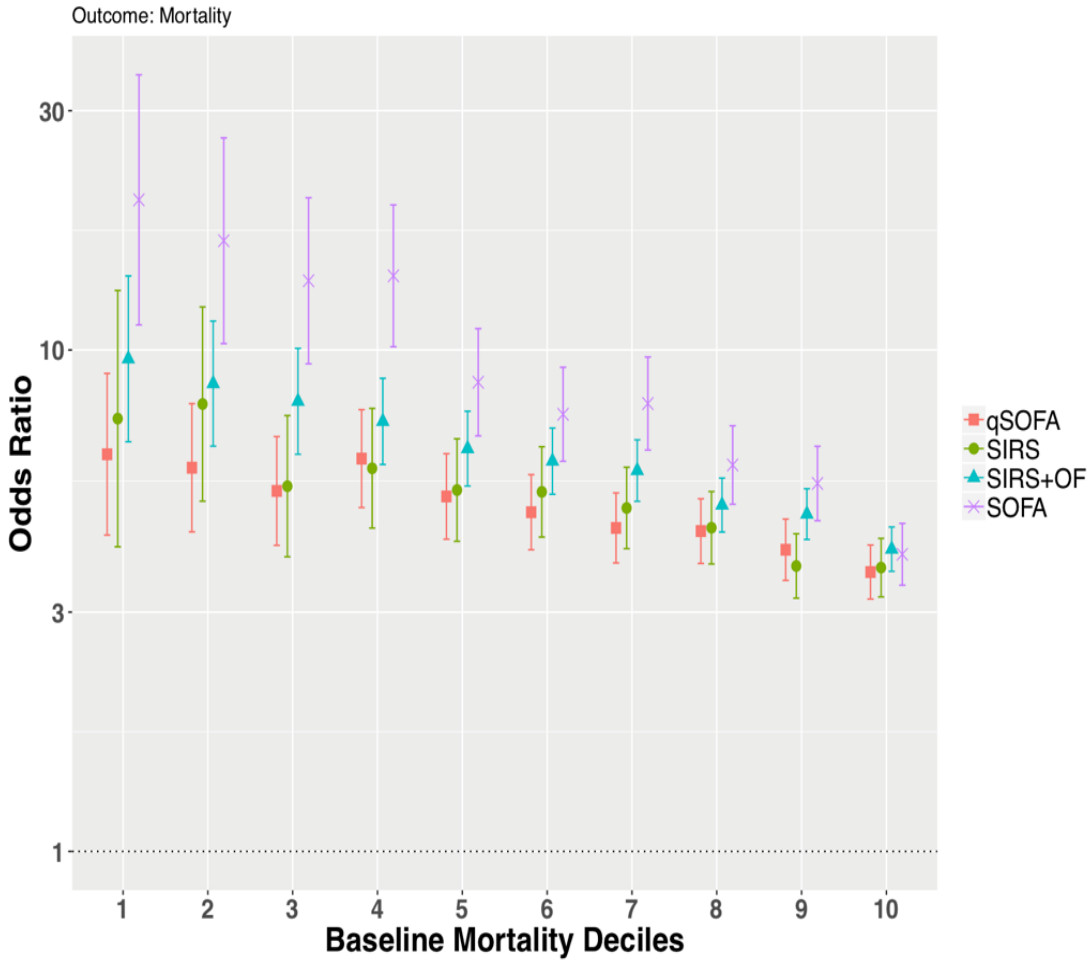
	qSOFA vs. SIRS	SOFA vs. qSOFA	SOFA vs. SIRS	SEPSIS-2 vs. SIRS	SEPSIS-2 vs. qSOFA	SEPSIS-2 vs. SOFA
In-Hospital Sepsis Between-Group Difference						
Unadjusted AUROC	0.012	0.016	0.028	0.086	0.074	0.058
Z-value	11.434	14.701	24.43	83.266	61.41	49.06
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Adjusted AUROC	-0.001	0.011	0.028	0.041	0.042	0.031
Z-value	-1.7694	15.33	11.937	52.772	50.241	36.222
p-value	0.07683	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
In-Hospital Mortality Between-Group Difference						
Unadjusted AUROC	0.036	0.012	0.047	0.080	0.044	0.033
Z-value	30.418	10.147	39.722	65.688	31.809	24.679
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Adjusted AUROC	0.008	0.00867	0.004	0.034	0.026	0.018
Z-value	7.8843	8.3814	14.999	30.923	21.591	14.737
p-value	3.16E-15	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

SIRS, systemic inflammatory response syndrome; OF, organ failure; SOFA, sepsis-related organ failure assessment; CI, confidence intervals; AUROC, area under the receiver operator curve; Z-Value calculated using Delong's test to compare differences between AUROC.

Likewise, predictive accuracy for mortality was higher using the Sepsis-2 algorithm (AUROC 0.673, 99% CI: 0.670-0.676) than two or more systemic inflammatory response syndrome criteria (AUROC 0.593, 99% CI: 0.591-0.596), qSOFA (AUROC 0.629, 99% CI: 0.626-0.632),

and SOFA (AUROC 0.641, 99% CI: 0.639-0.643); these differences were statistically significant. Similarly, when considered along with baseline risk prediction of mortality (adjusted analysis), the Sepsis-2 algorithm demonstrated greater discriminatory capacity for mortality (AUROC 0.777, 99% CI: 0.773 – 0.780) than two or more systemic inflammatory response syndrome criteria (AUROC 0.742, 99% CI: 0.739-0.746), qSOFA (AUROC 0.751, 99% CI: 0.747-0.754), SOFA (AUROC 0.759, 99% CI: 0.756-0.763); all differences were statistically significant (Table 2 and Figure 3). For the outcome of mortality in all patients, positive SOFA odds ratios were higher than two or more systemic inflammatory response syndrome criteria, qSOFA and and the Sepsis-2 algorithm (noted as SIRS+OF in Figure 3). Wider CIs and higher ORs in each measurement system in lower deciles with a descending pattern across the 10 deciles is depicted visually in Figure 3.

Figure 3. Odds Ratios for Mortality Comparing Encounters with Positive Measurement System Criteria or Scores (SIRS, qSOFA, SOFA, and SIRS+OF) for Training Cohort (All 638,757) across Deciles of Risk.



Discussion

In a retrospective cohort study of 1,012,410 hospitalized adults in 5 States in the U.S., researchers reported 30-day in-hospital mortality rates for sepsis patients ranging from 19-22% during the years of 2013-2015.[60] In an analysis of 2,566,689 sepsis cases using the Premier Healthcare Database (January 1, 2010, to September 30, 2016) mortality varied by severity in sepsis patients: 5.6% for sepsis without organ dysfunction, 14.9% for severe sepsis, and 34.2%

septic shock.[61] Data from the worldwide Intensive Care over Nations (ICON) audit reported rates of 18.5% for ICU mortality and 25.2% for hospital mortality for ICU patients with sepsis in a North American sample.[62] These rates are comparable to the 18.5% hospital mortality rate for sepsis patients admitted to adult ICUs reported in the results of this paper.

According to the Society of Critical Care Medicine's Critical Care Statistics webpage, mortality rates in patients admitted to adult ICUs average between 10% to 29%.[63] Hospital mortality of 10% was reported for 118,990 adult patients from 56 ICUs across 15 States in the U.S. in a study conducted by Lilly et al. (2014) using a dataset (2003 to 2008) that later became the Philip's eICU Research Institute database.[64] The hospital mortality rate of 9.5% for patients admitted to adult ICUs reported in the results of this paper is consistent with previous reported mortality rates.

The discriminatory capacity of binary measures related to the Sepsis-3 definition (SOFA and qSOFA scores of 2 or more), the Sepsis-1 definition (systemic inflammatory response syndrome threshold of 2 met) and the Sepsis- 2 definition (Sepsis-2 algorithm threshold 3.5) were assessed in a large population of adult ICU patients. The Sepsis-2 algorithm demonstrated better discrimination for suspected sepsis detection when compared to systemic inflammatory response syndrome, SOFA and qSOFA among adult ICU patients in 183 U.S. hospitals. Although SOFA scores may be a useful predictor of mortality, Sepsis-2 also demonstrated prognostic accuracy for mortality that was consistent with SOFA. The results of this study suggest that the use of expanded physiologic criteria combined with acute organ failure might perform better in early identification of sepsis in critically ill patients.

Some experts in the fields of critical care and sepsis have raised concerns that reliance on organ failure measurement systems (Sepsis-3) alone might identify sepsis too late or miss identifying some patients altogether.[35, 65] Physiologic data that include systemic inflammatory response syndrome and organ failure criteria are being extracted from electronic health records

and used as inputs into machine learning algorithms for sepsis detection with some of the best AUROCs reported as 0.68–0.99 in the intensive care unit.[66] Although Sepsis-2 in this study achieved only an adequate level (AUROC 0.77) in detection of sepsis, it was consistent with the best machine learning sepsis detection models and with SOFA (AUROC 0.79) in the Seymour et al. study.

For this study, researchers tested an algorithm that used selected expanded systemic inflammatory response syndrome criteria with OF criteria (systemic inflammatory response syndrome +OF) and compared its performance in discrimination of suspected sepsis and predictive accuracy of mortality with systemic inflammatory response syndrome criteria (Sepsis-1), qSOFA and SOFA scoring (Sepsis-3). By using diagnostic data to determine the presence of suspected sepsis (*eTable 3 in the supplementary appendix*) we were able to analyze the discriminatory capacity of each measurement system for detection of suspected sepsis which may lead to better tuning of predictive algorithms aimed at early sepsis detection. The results of this study demonstrated that the use of expanded physiologic criteria along with organ failure criteria could effectively detect suspected sepsis and prognosticate mortality in adult patients within the first 24 hours of ICU admission. As nurses tend to be the front-line care providers managing sepsis screening (manually or through use of automated electronic health record alerts)[14], this should be of interest to nurses and studies like this can be used to inform nurses on the usability of various measurement systems and algorithmic alerts aimed at early detection of sepsis.

Historically, simplified scoring systems were necessary because of ease of interpretability and lack of available real-time physiological and patient characteristic data that could be used to develop advanced algorithmic alerts. However, use of electronic health records and the growing science of machine learning, performance of scoring systems should be considered alongside simplicity and interpretability. Computerized alerting systems can

synthesize thousands of data points using complex algorithms and notify clinicians in real- or near real-time that their patients may have an evolving sepsis process. In ICU's, parsimonious scoring systems are easy to interpret and can be used without a computer. However, the use of algorithmic discrimination and better timing of sepsis detection could be augmented with machine-enhancement. Use of SOFA criteria alone diminishes reliance on important physiologic criteria that should be considered in development of predictive models and machine learning algorithms that can support more sophisticated alerting.

This study's cohort represented adult ICU patients located in coronary care, surgical, trauma, neuroscience, and medical intensive care units in over 180 hospitals across the United States (U.S.) from the years 2010-2015. The patients represented in this study were broadly distributed regarding hospital and community size, U.S. regions, presence or absence of teaching programs, and models of ICU staffing. Patients within varying age groups, gender differences, ethnic backgrounds, and comorbid conditions were represented in this study. This broad representation supports generalizability of these results to adult ICU patients in the U.S. Code sharing from this research can be used by other researchers to replicate this study using newer datasets.

There were several limitations of this study. First, the data were entered prospectively into the database and this investigation was retrospective. Although researchers used research transparency through code sharing along with careful selection of inclusion/exclusion criteria and a deep understanding of the dataset, bias related to confounding variables could not eliminated. Second, we only included cases that showed evidence of having data present in specific data tables versus including all potential cases and conducting imputation methods which reduced the primary cohort to 45.2% of the complete cohort. This decision could have introduced bias.[42] Third, the dataset did not include information related to mortality post hospital discharge, consequently in-hospital mortality was chosen for this endpoint.

Fourth, sepsis is a difficult condition to detect without clear delineation of onset. Thus, the diagnostic data from the eCareManager® problem list and/or APACHE diagnosis tables were used to determine presence of suspected sepsis in this cohort. Using diagnostic data documented within the first 24 hours of ICU admission limited the susceptibility to changing diagnosis and coding practices over time and allowed researchers to determine suspicion of sepsis within a distinct period. However, this was partially addressed by including analysis of mortality as an outcome. Fifth, demographic, comorbid conditions and diagnostic data within the model was dependent on accurate documentation. Sixth, we did not exclude patients with limitations of care or comfort care status, and this could have skewed the outcome of mortality. Seventh, the sepsis screening criteria used in the modified Sepsis-2 algorithm represents an attempt to codify criteria that were intentionally defined with ambiguity.

Conclusions

For patients with sepsis, early detection has been shown to be crucial for improving outcome and survival. However, sepsis can be difficult to diagnose quickly which can increase the risk of complications, shock, and death. The major finding in this study was that expanded systemic inflammatory response syndrome with organ failure criteria (Sepsis-2) performed better than SOFA and qSOFA scoring systems (Sepsis-3), and traditional systemic inflammatory response syndrome criteria (Sepsis-1) with thresholds of two in discrimination of sepsis. Unfortunately, performance for all measurement systems falls short and it is evident that our progress in standardizing sepsis screening criteria has been modest, at best. Early identification of sepsis has been cited as one of the biggest obstacles to timely therapeutic interventions aimed at saving lives and reducing complications and death. Nurses are effective agents in assisting with early identification of sepsis. The findings of this study support using a combination of expanded physiologic criteria with organ failure criteria in computer-enhanced

algorithms that can be used by nurses for early and effective sepsis detection.

Conflicts of Interest and Source of Funding:

The principal investigator has served in advisory and speaking roles for Philips Healthcare and serves as a clinical consultant for Baxter Healthcare Corporation. The principal investigator was a doctoral student with the University of Kansas, School of Nursing during research design, data extraction, and analysis. Three members of the research team received research funding from the National Institute of Health through NIBIB grant R01 EB017205 and Philips Healthcare. Several members of the research team were employed at MIT or Philips Healthcare during data extraction, analysis, and initial manuscript preparation but have moved on other roles. One member serves as an advisor to Ceiba Healthcare.

Acknowledgments:

The principal investigator would like to thank multiple colleagues from the University of Kansas: Devin C. Koestler, Ph.D for his instrumental critique and advice during the writing of statistical code in R and LaVerne Manos, RN, DNP; Qiuhua Shen, RN, Ph.D.; and Kelly A. Bosak, RN, Ph.D for their invaluable input related to study design. The research team would like to thank members of the eICU Research Institute Publications Committee for reviewing and approving this study and/or providing input on manuscript preparation.

References

1. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet**.** 2020;395(10219):200-211.
2. Buchman TG, Simpson SQ, Sciarretta KL, et al. Sepsis Among Medicare Beneficiaries: 1. The Burdens of Sepsis, 2012-2018. Crit Care Med**.** 2020;48(3):276-288.
3. Buchman TG, Simpson SQ, Sciarretta KL, et al. Sepsis Among Medicare Beneficiaries: 2. The Trajectories of Sepsis, 2012-2018. Crit Care Med**.** 2020;48(3):289-301.
4. Buchman TG, Simpson SQ, Sciarretta KL, et al. Sepsis Among Medicare Beneficiaries: 3. The Methods, Models, and Forecasts of Sepsis, 2012-2018. Crit Care Med**.** 2020;48(3):302-318.

5. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med. 2017;43(3):304-377.
6. Bone R, Balk R, Cerra F, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM consensus conference committee. American College of Chest Physicians/Society of Critical Care Medicine. Chest. 1992;101(6):1644-1655.
7. Bone R, Grodzin C, Balk R. Sepsis: a new hypothesis for pathogenesis of the disease process. Chest. 1997;112(1):235-243.
8. Russell J. Management of sepsis. N Engl J Med 2006;355(16):1699-1713.
9. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Crit Care Med. 2003;31(4):1250-1256.
10. Dellinger RP, Levy MM, Rhodes A, et al. Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock: 2012. Critical Care Medicine. 2013;41(2):580-637.
11. Seymour C, Liu V, Iwashyna T, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):762-774.
12. Singer M, Deutschman C, Seymour C, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801-810.
13. Haddad L, Annamaraju P, Toney-Butler T. Nursing Shortage. StatPearls Publishing. 2023 Feb 22. Available at: https://www.ncbi.nlm.nih.gov/books/NBK493175/. Accessed September 17, 2022.
14. Schorr CA, Seckel MA, Papathanassoglou E, Kleinpell R. Nursing Implications of the Updated 2021 Surviving Sepsis Campaign Guidelines. Am J Crit Care. 2022;31(4):329-336.
15. American Association of Critical-Care Nurses. Nurses on the Front Line of Sepsis. American Association of Critical-Care Nurses (AACN). Available at: https://www.aacn.org/clinical-resources/sepsis. Accessed March 25, 2023.
16. Hwang MI, Bond WF, Powell ES. Sepsis Alerts in Emergency Departments: A Systematic Review of Accuracy and Quality Measure Impact. West J Emerg Med. 2020;21(5):1201-1210.
17. Alberto L, Marshall AP, Walker R, Aitken LM. Screening for sepsis in general hospitalized patients: a systematic review. Journal of Hospital Infection. 2017;96(4):305-315.
18. Guirgis FW, Jones L, Esma R, et al. Managing sepsis: Electronic recognition, rapid response teams, and standardized care save lives. Journal of Critical Care. 2017;40:296-302.
19. Manaktala S, Claypool SR. Evaluating the impact of a computerized surveillance algorithm and decision support system on sepsis mortality. Journal of the American Medical Informatics Association. 2017;24(1):88-95.
20. Austrian JS, Jamin CT, Doty GR, Blecker S. Impact of an emergency department electronic sepsis surveillance system on patient mortality and length of stay. Journal of the American Medical Informatics Association. 2018;25(5):523-529.
21. Downing NL, Rolnick J, Poole SF, et al. Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. BMJ Quality & Safety. 2019;28(9):762.
22. Despins LA. Automated Detection of Sepsis Using Electronic Medical Record Data: A Systematic Review. The Journal for Healthcare Quality (JHQ). 2017;39(6).
23. Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Internal Medicine. 2021;181(8):1065.

24. Nelson B. Preventing sepsis alert fatigue. The Hospitalist; 2018.
25. The Lancet Respiratory M. Crying wolf: the growing fatigue around sepsis alerts. Lancet Respir Med**.** 2018;6(3):161.
26. Kluwer W. Why your EHR is not enough to improve sepsis outcomes. Wolters Kluwer N.V. February 13, 2019. Available at: https://www.wolterskluwer.com/en/expert-insights/why-your-ehr-is-not-enough-to-improve-sepsis-outcomes. Accessed February 19, 2021.
27. Koch C, Edinger F, Fischer T, et al. Comparison of qSOFA score, SOFA score, and SIRS criteria for the prediction of infection and mortality among surgical intermediate and intensive care patients. World journal of emergency surgery : WJES**.** 2020;15(1):63-63.
28. Williams JM, Greenslade JH, McKenzie JV, Chu K, Brown AFT, Lipman J. Systemic Inflammatory Response Syndrome, Quick Sequential Organ Function Assessment, and Organ Dysfunction: Insights From a Prospective Database of ED Patients With Infection. Chest**.** 2017;151(3):586-596.
29. Finkelsztein EJ, Jones DS, Ma KC, et al. Comparison of qSOFA and SIRS for predicting adverse outcomes of patients with suspicion of sepsis outside the intensive care unit. Crit Care**.** 2017;21(1):73.
30. Moskowitz A, Patel PV, Grossestreuer AV, et al. Quick Sequential Organ Failure Assessment and Systemic Inflammatory Response Syndrome Criteria as Predictors of Critical Care Intervention Among Patients With Suspected Infection. Crit Care Med**.** 2017;45(11):1813-1819.
31. Parks Taylor S, McWilliams A, Taylor BT, et al. Predictive Accuracy of Quick Sequential Organ Failure Assessment for Hospital Mortality Decreases With Increasing Comorbidity Burden Among Patients Admitted for Suspected Infection. Crit Care Med**.** 2019;47(8):1081-1088.
32. Centner FS, Schoettler JJ, Fairley AM, et al. Impact of different consensus definition criteria on sepsis diagnosis in a cohort of critically ill patients-Insights from a new mathematical probabilistic approach to mortality-based validation of sepsis criteria. PLoS One**.** 2020;15(9):e0238548.
33. Gando S, Shiraishi A, Abe T, et al. The SIRS criteria have better performance for predicting infection than qSOFA scores in the emergency department. Scientific Reports**.** 2020;10(1):8095.
34. Johnson AEW, Aboab J, Raffa JD, et al. A Comparative Analysis of Sepsis Identification Methods in an Electronic Database. Critical care medicine**.** 2018;46(4):494-499.
35. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Medicine**.** 2021;47(11):1181-1247.
36. Zhang Z, Chen L, Xu P, et al. Effectiveness of automated alerting system compared to usual care for the management of sepsis. npj Digital Medicine**.** 2022;5(1):101.
37. Philips Healthcare. eICU Research Institute: Big data that's changing the face of critical care. Koninklijke Philips N.V.,. Available at: https://www.usa.philips.com/healthcare/solutions/enterprise-telehealth/eri. Accessed May 4, 2022.
38. The MIT Lab for Computational Physiology eICU Collaborative Research Database. Available at: http://eicu-crd.mit.edu/. Accessed March 7, 2018.
39. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Badawi O, Mark RG. eICU Collaborative Research Database (Version 2.0). PhysioNet. 2018/09/11. Available at: https://physionet.org/content/eicu-crd/2.0/. Accessed February 18, 2021.
40. Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet. Components of a New Research Resource for Complex Physiologic Signals**.** 2000;101(23):e215-e220.

41. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/.
42. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ**.** 2009;338:b2393.
43. Vasilevskis E, Kuzniewicz M, Cason B, et al. Mortality probability model III and simplified acute physiology score II: assessing their value in predicting length of stay and comparison to APACHE IV. Chest**.** 2009;136(1):89-101.
44. Zimmerman J, Kramer A, McNair D, Malila F. Acute physiology and chronic health evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med**.** 2006;34(5):1297-1310.
45. Feng S, Dubin JA. Identifying early-measured variables associated with APACHE IVa providing incorrect in-hospital mortality predictions for critical care patients. Scientific Reports**.** 2021;11(1):22203.
46. Manganaro L, Stark M. APACHE® Foundations User Guide Cerner Corporation; 2010:49.
47. Raith EP, Udy AA, Bailey M, et al. Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. JAMA**.** 2017;317(3):290-300.
48. Badawi O, Fromm R, Rincon T, et al. Using annotated electronic patient data to develop a predictive model for identifying severe sepsis 463. Crit Care Med**.** 2009;37(12):A224.
49. Cintula P, Fermüller CG, Noguera C. Fuzzy Logic. Metaphysics Research Lab, Stanford University. July 18, 2017. Available at: https://plato.stanford.edu/archives/fall2017/entries/logic-fuzzy/. Accessed March 31, 2018.
50. Nimri R, Phillip M. Artificial pancrease: fuzzy logic and control of glycemia. Curr Opin Endocrinol Diabetes Obes**.** 2014;21(4):251-256.
51. Nguyen B, Bernstein D, Bates J. Controlling mechanical ventilation in acute respiratory distress syndrome with fuzzy logic. J Crit Care**.** 2014;29(4):551-556.
52. Bardossy A, Blinowska A, Kuzmicz W, et al. Fuzzy logic-based diagnostic algorithm for implantable cardioverter defibrillators. Artif Intell Med**.** 2014;60(2):113-121.
53. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med**.** 2001;29(7):1303-1310.
54. Martin G, Mannino D, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med**.** 2003;348(16):1546-1554.
55. Pilcher D, Litton E, Millar J, Huckson S, Gian S. APD Data Dictionary for Software Programmers. In: Committee ACM, ed. ANZICS Core - Adult Patient Database Carlton South VIC ANZICS Centre for Outcome and Resource Evaluation; 2017:154.
56. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited*. Critical Care Medicine**.** 2007;35(9):2052-2056.
57. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain**.** 2008;8(6):221-223.
58. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics**.** 1988;44(3):837-845.
59. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology**.** 1982;143(1):29-36.
60. Kahn JM, Davis BS, Yabes JG, et al. Association Between State-Mandated Protocolized Sepsis Care and In-hospital Mortality Among Adults With Sepsis. JAMA**.** 2019;322(3):240-250.

61. Paoli CJ, Reynolds MA, Sinha M, Gitlin M, Crouser E. Epidemiology and Costs of Sepsis in the United States-An Analysis Based on Timing of Diagnosis and Severity Level. Crit Care Med**.** 2018;46(12):1889-1897.
62. Sakr Y, Jaschinski U, Wittebole X, et al. Sepsis in Intensive Care Unit Patients: Worldwide Data From the Intensive Care over Nations Audit. Open Forum Infect Dis**.** 2018;5(12):ofy313.
63. Halpern NA. Critical Care Statistics. Society of Critical Care Medicine. Available at: https://www.sccm.org/Communications/Critical-Care-Statistics. Accessed March 23, 2023.
64. Lilly CM, McLaughlin JM, Zhao H, Baker SP, Cody S, Irwin RS. A multicenter study of ICU telemedicine reengineering of adult critical care. Chest**.** 2014;145(3):500-507.
65. Genga KR, Russell JA. Update of Sepsis in the Intensive Care Unit. J Innate Immun**.** 2017;9(5):441-455.
66. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive care medicine**.** 2020;46(3):383-400.