

Using Statistical and Machine Learning Methods to Evaluate the Prognostic Accuracy of SIRS and qSOFA

Akash Gupta, MS¹, Tieming Liu, PhD¹, Scott Shepherd, DO, FACP², William Paiva, PhD²

¹Industrial Engineering and Management, ²Center for Health Systems Innovation, Oklahoma State University, Stillwater, OK, USA

Objectives: The objective of this study was to compare the performance of two popularly used early sepsis diagnostic criteria, systemic inflammatory response syndrome (SIRS) and quick Sepsis-related Organ Failure Assessment (qSOFA), using statistical and machine learning approaches. **Methods:** This retrospective study examined patient visits in Emergency Department (ED) with sepsis related diagnosis. The outcome was 28-day in-hospital mortality. Using odds ratio (OR) and modeling methods (decision tree [DT], multivariate logistic regression [LR], and naïve Bayes [NB]), the relationships between diagnostic criteria and mortality were examined. **Results:** Of 132,704 eligible patient visits, 14% died within 28 days of ED admission. The association of qSOFA ≥ 2 with mortality (OR = 3.06; 95% confidence interval [CI], 2.96–3.17) greater than the association of SIRS ≥ 2 with mortality (OR = 1.22; 95% CI, 1.18–1.26). The area under the ROC curve for qSOFA (AUROC = 0.70) was significantly greater than for SIRS (AUROC = 0.63). For qSOFA, the sensitivity and specificity were DT = 0.39, LR = 0.64, NB = 0.62 and DT = 0.89, LR = 0.63, NB = 0.66, respectively. For SIRS, the sensitivity and specificity were DT = 0.46, LR = 0.62, NB = 0.62 and DT = 0.70, LR = 0.59, NB = 0.58, respectively. **Conclusions:** The evidences suggest that qSOFA is a better diagnostic criteria than SIRS. The low sensitivity of qSOFA can be improved by carefully selecting the threshold to translate the predicted probabilities into labels. These findings can guide healthcare providers in selecting risk-stratification measures for patients presenting to an ED with sepsis.

Keywords: Sepsis, Systemic Inflammatory Response Syndrome, Severity of Illness Index, Medical Informatics, Artificial Intelligence

Submitted: March 14, 2018

Revised: April 19, 2018

Accepted: April 20, 2018

Corresponding Author

Akash Gupta, MS

Industrial Engineering and Management, Oklahoma State University, Stillwater, OK 74074, USA. Tel: +1-405-334-8339, E-mail: akashg@okstate.edu

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2018 The Korean Society of Medical Informatics

1. Introduction

Patients with sepsis are at considerable risk for severe complications and death. In-hospital mortality rates for sepsis patients range from 10% to 20% [1], and between 2007 and 2013, the number of hospital admissions due to sepsis increased nearly 49% to more than 352 per 100,000 persons per year [2]. At about \$20 billion, or 5.2% of national hospital costs, sepsis is considered the most expensive condition treated in US hospitals [3].

Early detection of sepsis is critical because each hour delay in treatment increases mortality by 7.6% [4]. To detect the disease in early stage, many early diagnostic criteria were

proposed [5-8]. The two early diagnostic criteria, systemic inflammatory response syndrome (SIRS) and quick Sepsis-related Organ Failure Assessment (qSOFA), are dominantly used in clinics to assess the criticality of disease. The foundation of SIRS relies on inflammatory response to infection while the basis of qSOFA relies on the organ failures. Table 1 shows the clinical criteria used for SIRS and qSOFA. The presence of a criterion at the designated threshold yields a score of 1, otherwise 0. SIRS scores range from 0 to 4, with scores of 2 indicative of SIRS and, subsequently, an increased likelihood of mortality. qSOFA scores range from 0 to 3, with scores of 2 indicative of high risk for mortality. Of the seven indicators between the two assessments, only respiratory rate is common to both, with the threshold slightly higher in qSOFA.

Singer et al. [5] proposed qSOFA and compared its performance against SIRS. The authors argued of using qSOFA in clinics because of its high specificity. However, the several practitioners and researchers argued against on accepting qSOFA because of its low sensitivity that could lead to many patients undiagnosed. Given the criticality of early sepsis detection and divided opinion of researchers, a detailed comparison of SIRS and qSOFA can establish clarity in practitioner's mind for the use of more accurate diagnostic criteria. In this paper, combination of statistical and machine learning [9,10] approaches was employed to evaluate the performance of SIRS and qSOFA. Adopting the approach from [11] for finding the threshold to translate predicted probability into labels, the modeling approaches provided a balance between sensitivity and specificity. This study has great clinical implications as it provides evidence in favor of effective screening criteria that enables healthcare providers to effectively manage limited resources of Emergency Department (ED).

Table 1. SIRS and qSOFA criteria

| Criterion | Threshold | |
|--|------------|-------|
| | SIRS | qSOFA |
| Body temperature (°C) | <36 or >38 | - |
| Heart rate (beats/min) | >90 | - |
| While blood cell count (10 ³ /μL) | <4 or >12 | - |
| Respiratory rate (breaths/min) | >20 | ≥22 |
| Systolic blood pressure (mmHg) | - | ≤100 |
| Glasgow Coma Scale | - | ≤13 |

SIRS: systemic inflammatory response syndrome, qSOFA: quick Sepsis-related Organ Failure Assessment.

II. Methods

1. Study Design

This study used data from the Cerner Corporation's HIPAA-compliant Health Facts database. At the time of this study, the database comprised electronic health records for 379 million encounters from 480 affiliated hospitals across the United States. Comprehensive clinical records for individual encounters include date- and time-stamped information on admission and discharge, discharge location, laboratory data, diagnoses (using International Classification of Diseases, ninth revision, clinical modification [ICD-9-CM] diagnostic codes), patient demographics, and additional clinical and billing information.

2. Study Setting and Population

All visits for patients admitted to the ED between January 1, 2009 to December 31, 2015 with primary or secondary diagnoses of sepsis (ICD-9-CM code 995.91), severe sepsis (ICD-9-CM code 995.92), septic shock (ICD-9-CM code 785.52), and unspecified sepsis (ICD-9-CM 038.xx) were included in the initial data extraction. According to the third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), sepsis is defined as SOFA ≥ 2 and infection. However, determining SOFA score from electronics health record is a major challenge. The SOFA score requires eight variables. Some of them are lab results that are very sparsely available (missing percentage >50%). In addition, the information about respiratory support and the quantity of dopamine are not recorded properly. Deriving population after imputation of such sparse data could potentially lead to population biasness. Therefore, we used ICD-9-CM code for selecting the sepsis population. Extracted data included age, gender, US Census region (Midwest, Northeast, South, and West), hospital location (urban/rural), SIRS/qSOFA clinical variables and discharge type. The outcome of interest was 28-day in-hospital mortality. We selected 28 days as an endpoint because it is widely used in literature of sepsis [12].

We used mortality to infer about diagnostic accuracy of SIRS and qSOFA. Although qSOFA was introduced by Sepsis-3 to detect sepsis early, however, Sepsis-3 also made an assumption that tied the mortality and the early diagnosis of sepsis [13]. The study assumes that mortality and sepsis detection has direct relation. High rate of mortality implies high risk of sepsis. The Sepsis-3 study compares the prognostic accuracy of SIRS, SOFA, LODS (logistic organ dysfunction syndrome) and qSOFA based on their discrimination power of mortality prediction. Hence, the diagnostic

accuracy of SIRS and qSOFA is evaluated based their mortality prediction.

The study included patients 18 years of age and older with length of stay less than or equal to 28 days. Of the 230,451 extracted encounters, 97,747 encounters were excluded (20,639 under 18 years old; 12,276 with length of stay greater than 28 days; and 64,832 with non-emergency admission). Our analytical sample consisted of 132,704 encounters occurring in the ED (Figure 1). The study used first measurement to compute SIRS and qSOFA. The aim of SIRS and qSOFA is to identify patient at high risk of developing sepsis in early stage. Hence, we compared the mortality prediction of both criteria based on the measurement recorded just after the admission.

3. Data Analysis

Data preparation and statistical analysis were performed using R version 3.2.5 (<https://cran.r-project.org>). The two approaches, odds ratio (OR) and modeling methods (multivariate logistic regression, decision tree, and naïve Bayes), were used to compare the associations of SIRS and qSOFA with 28-day in-hospital mortality. The OR is a statistical parameter used to determine the strength of association between two categorical variables. Note that the association is statistically significant if the range of 95% confidence interval (CI) of OR does not contain the value 1.

To investigate the robustness of two diagnostic criteria, the patient cohort was divided into 10 classes of varying baseline risk [14]. The baseline risk was estimated using patient demographic and hospital characteristic including age, sex, hospital location, and US Census region. The OR was computed to compare the association of mortality with SIRS (≥ 2 vs. < 2) and qSOFA (≥ 2 vs. < 2) across each decile.

For further investigation, the modeling approaches (multi-

variate logistic regression, decision tree [15], and naïve Bayes classifier [16]) were employed on SIRS and qSOFA variables. Following the literature, the performance matrices of SIRS and qSOFA were evaluated above the baseline risk [11,17]. The set of variables used for modeling purpose are summarized as follows:

- (i) Baseline variables: Age, sex, hospital location (urban/rural), and US Census region.
- (ii) Base + SIRS variables: Baseline variables plus SIRS clinical variables.
- (iii) Base + qSOFA variables: Baseline variables plus qSOFA clinical variables.

We assessed the discriminatory power of each model using the area under the receiver operating characteristic (AUROC) curve. Other performance parameters such as sensitivity, specificity, positive predictive value, and negative predictive value were also used. Table 2 represents the confusion matrix. Equations (1)–(6) are the expressions of performance metrics derived from Table 2. The threshold to determine the predicted class from estimated probability was decided using the ROC plot. The cutoff that resulted a point in the ROC plot closest to (0, 1) was selected as threshold [11]. In (0, 1), the first index represents false positive rate and second represents true positive rate.

$$OR = \frac{TP \times TN}{FP \times FN} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$Positive\ predictive\ value = \frac{TP}{TP + FP} \tag{4}$$

$$Negative\ predictive\ value = \frac{TN}{TN + FN} \tag{5}$$

$$Accuracy = \frac{TN + TP}{TN + FN + TP + FP} \tag{6}$$

To summarize the analysis, we first compared the association of mortality with SIRS ≥ 2 and qSOFA ≥ 2 using OR. Later, we divided the patient visits at the ED into 10 classes to

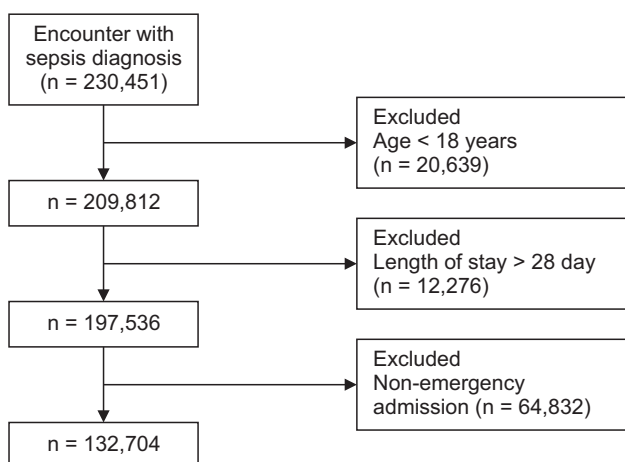


Figure 1. Encounter flow chart.

Table 2. Confusion matrix

| | | Actual value | |
|-----------------|-----|--------------|----|
| | | Yes | No |
| Predicted value | Yes | TP | FP |
| | No | TN | FN |

TP: true positive, FP: false positive, TN: true negative, FN: false negative.

investigate the trend of OR of SIRS ≥ 2 and qSOFA ≥ 2 across varying initial risk. The results obtained from OR analysis provided contradictory behavior for sensitivity and specificity of SIRS and qSOFA. Therefore, instead of considering SIRS or qSOFA score, we used SIRS and qSOFA variables to estimate the probability of risk using modeling approaches.

The percentages of missing data for SIRS and qSOFA variables were as follows: white blood cell count (37%), Glasgow Coma Scale (30%), heart rate (26%), blood pressure (14%), temperature (7%), and respiratory rate (3%). To manage the problem of missing values in clinical variables, multivariate imputation by chained equations (MICE) was used [18].

III. Results

1. Population Characteristics

Table 3 compares demographic, geographic, and clinical characteristics of non-expired and expired encounters. The mortality rate for ED encounters was 14%, with a median age of 75 years (interquartile range [IQR], 60–83 years) and

50.4% male. In comparison, the non-expired group had a median age of 65 years (IQR, 52–78 years) and was 48.7% male. The Mood's median test [19] showed a significant difference in median age between the groups, with age increasing mortality risk. Using 95% CI of OR, we found that gender, hospital location, and census region were significantly associated with mortality. The results showed that males were at higher risk than females (OR = 1.07; 95% CI, 1.01–1.10), and encounters at urban hospitals had lower levels of risk than rural hospitals (OR = 0.77; 95% CI, 0.75–0.80). Similarly, we compared the association of mortality with hospital census region and found encounters in the Midwest and West to have lower risk for mortality, while those in the Northeast and South had higher risk.

Table 3 also shows the association of individual variables of SIRS and qSOFA with mortality. For the qSOFA, each criterion showed a positive association with mortality: blood pressure (OR = 2.00; 95% CI, 1.94–2.08), Glasgow Coma Scale (OR = 2.96; 95% CI, 2.87–3.06), and respiratory rate (OR = 1.66; 95% CI, 1.61–1.72). Whereas, for SIRS, heart

Table 3. Population characteristic

| Variables | Non-expired (n = 114,030) | Expired (n = 18,674) | OR | 95% CI |
|--|------------------------------|-------------------------|------|-----------|
| Demographics | | | | |
| Age (yr) | 65 (52–78) | 75 (60–83) | | |
| Male | 55,554 (48.7) | 9,424 (50.4) | 1.07 | 1.01–1.10 |
| Hospital location | | | | |
| Urban | 118,193 (81.8) | 22,028 (77.7) | 0.77 | 0.75–0.80 |
| Census region | | | | |
| Midwest | 21,857 (19.2) | 2,691 (14.4) | 0.71 | 0.68–0.74 |
| Northeast | 35,862 (31.5) | 6,357 (34.0) | 1.13 | 1.09–1.16 |
| South | 40,820 (35.8) | 7,410 (39.7) | 1.18 | 1.14–1.22 |
| West | 15,491 (13.6) | 2,216 (11.9) | 0.86 | 0.82–0.90 |
| SIRS criteria met | | | | |
| Temperature (<36°C or >38°C) | 27,676 (24.3) | 3,950 (21.2) | 0.84 | 0.80–0.87 |
| Heart rate (>90 beats/min) | 69,712 (61.1) | 11,791 (63.1) | 1.09 | 1.05–1.12 |
| Respiratory rate (>20 breaths/min) | 37,253 (32.7) | 8,356 (44.8) | 1.67 | 1.62–1.72 |
| WBC count (< 4 or >12×10 ³ /μL) | 67,509 (59.2) | 11,218 (60.0) | 1.04 | 1.01–1.07 |
| qSOFA criteria met | | | | |
| Systolic BP (≤100 mmHg) | 23,451 (20.6) | 6,387 (34.2) | 2.00 | 1.94–2.08 |
| Glasgow Coma Scale (≤13) | 20,745 (18.2) | 7,417 (39.7) | 2.96 | 2.87–3.06 |
| Respiratory rate (≥22 breaths/min) | 35,667 (31.3) | 8,046 (43.1) | 1.66 | 1.61–1.72 |

Values are presented as median (interquartile range) or number (%).

OR: odds ratio, CI: confidence interval, SIRS: systemic inflammatory response syndrome, WBC: white blood cell count, qSOFA: quick Sepsis-related Organ Failure Assessment, BP: blood pressure.

rate (OR = 1.09; 95% CI, 1.05–1.12), respiratory rate (OR = 1.67; 95% CI, 1.62–1.72), and white blood cell (OR = 1.04; 95% CI, 1.01–1.07) were associated with a higher risk of mortality. While respiratory rate had the smallest association with mortality among qSOFA criteria, it had the largest association among SIRS criteria. The association of the SIRS criterion of body temperature (<36°C or >38°C) with mortality (OR = 0.84; 95% CI, 0.80–0.87) was counterintuitive as it suggests that high or low body temperature reduces the risk of mortality.

Figure 2 shows the distribution of encounters by SIRS and qSOFA scores. There were 23,260 (17.5%) encounters that met the qSOFA criteria for mortality risk (qSOFA ≥2), whereas 80,015 (39.7%) encounters met the definition of (SIRS ≥2). Figure 3 shows the distribution of mortality for the SIRS and qSOFA scores, with rates markedly increasing across qSOFA scores compared to the relatively uniform dis-

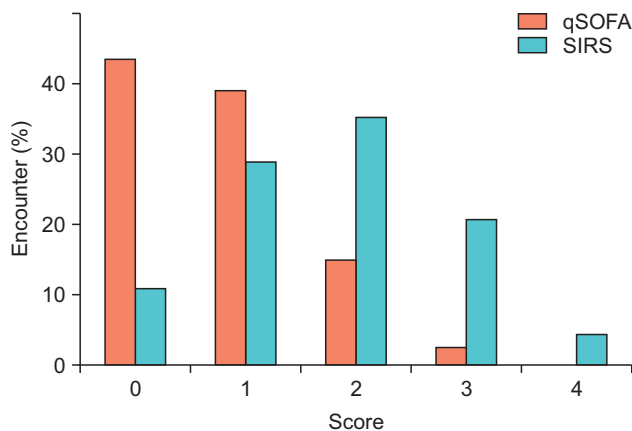


Figure 2. Distribution of encounters across SIRS and qSOFA scores. SIRS: systemic inflammatory response syndrome, qSOFA: quick Sepsis-related Organ Failure Assessment.

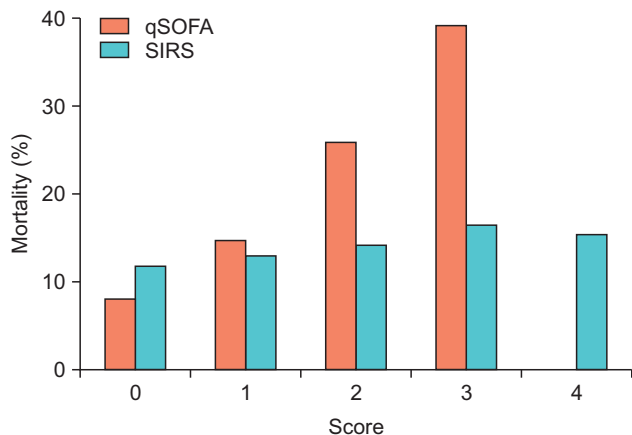


Figure 3. Distribution of mortality across SIRS and qSOFA scores. SIRS: systemic inflammatory response syndrome, qSOFA: quick Sepsis-related Organ Failure Assessment.

tribution of SIRS scores. For qSOFA scores, mortality rates were 40.1% for the highest possible score of 3 and 8% for a score of 0. In contrast, for SIRS scores, mortality rates were 18.9% for the highest possible score of 4 and 11.3% with a score of 0.

2. Comparison of Prognostic Accuracy of SIRS and qSOFA Scores Using OR

Confusion matrices for both SIRS and qSOFA are shown in Table 4. Both SIRS and qSOFA were significantly associated with mortality. The association was considerably stronger for qSOFA ≥2 (OR = 3.06; 95% CI, 2.96–3.17) than for SIRS ≥2 (OR = 1.22; 95% CI, 1.18–1.26). The classification accuracy, sensitivity, and specificity for qSOFA were 0.78, 0.35, and 0.85, respectively. The same performance parameters for SIRS were 0.44, 0.64, and 0.40, respectively. Using the cutoff of ≥2 for both measures, the qSOFA outperformed SIRS on accuracy and specificity, but demonstrated lower sensitivity.

Figure 4 compares the association of SIRS and qSOFA with mortality over deciles of baseline risk. The deciles of pa-

Table 4. Confusion matrix for SIRS and qSOFA

| | | Expired | |
|----------|-----|---------|--------|
| | | Yes | No |
| SIRS ≥2 | Yes | 12,021 | 67,994 |
| | No | 6,653 | 46,036 |
| qSOFA ≥2 | Yes | 6,464 | 16,796 |
| | No | 12,210 | 97,234 |

SIRS: systemic inflammatory response syndrome, qSOFA: quick Sepsis-related Organ Failure Assessment.

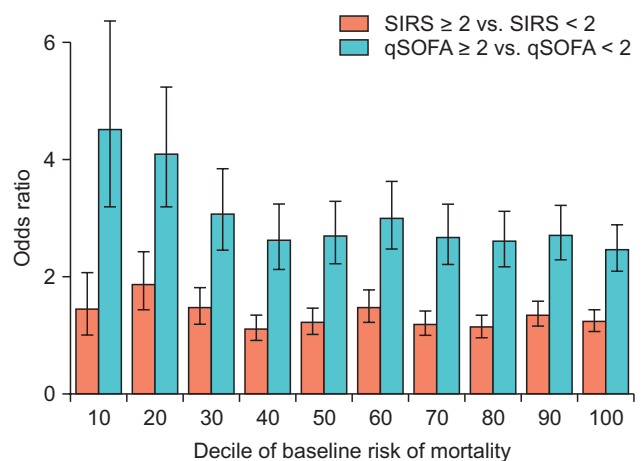


Figure 4. Odds ratio for in-hospital mortality for each decile of baseline risk. SIRS: systemic inflammatory response syndrome, qSOFA: quick Sepsis-related Organ Failure Assessment.

Table 5. Performance measures of SIRS and qSOFA (10-fold cross-validation)

| Performance metrics | Baseline + SIRS | | | Baseline + qSOFA | | |
|---------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | DT | LR | NB | DT | LR | NB |
| AUROC (95% CI) | 0.622 (0.615–0.630) | 0.643 (0.637–0.652) | 0.653 (0.635–0.650) | 0.612 (0.605–0.618) | 0.696 (0.689–0.703) | 0.696 (0.688–0.703) |
| Sensitivity | 0.459 | 0.617 | 0.623 | 0.394 | 0.642 | 0.616 |
| Specificity | 0.703 | 0.585 | 0.580 | 0.819 | 0.634 | 0.664 |
| Positive predictive value | 0.203 | 0.196 | 0.195 | 0.263 | 0.223 | 0.230 |
| Negative predictive value | 0.891 | 0.903 | 0.903 | 0.892 | 0.915 | 0.914 |
| Accuracy | 0.669 | 0.590 | 0.586 | 0.759 | 0.631 | 0.657 |

SIRS: systemic inflammatory response syndrome, qSOFA: quick Sepsis-related Organ Failure Assessment, DT: decision tree, LR: logistic regression, NB: naive Bayes, AUROC: area under the receiver operating characteristic curve, CI: confidence interval.

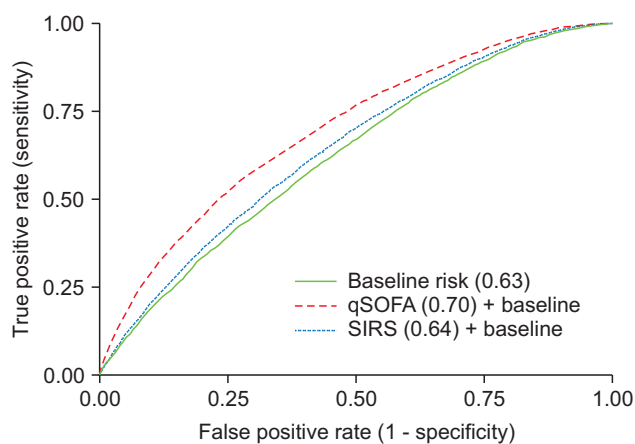


Figure 5. The area under the receiver operating characteristic (AUROC) curve using multivariate logistic regression. SIRS: systemic inflammatory response syndrome, qSOFA: quick Sepsis-related Organ Failure Assessment.

tients were derived from baseline variables and multivariate logistic regression. For each decile, the OR for qSOFA score (≥ 2 vs. < 2) was greater than for SIRS score (≥ 2 vs. < 2). The OR of qSOFA ranged from 4.3 among those in the lowest baseline risk decile to 2.4 among those in the highest decile, while ORs for SIRS across deciles were more or less constant.

3. Comparison of Prognostic Accuracy of SIRS and qSOFA Using Modeling Approach

Table 5 shows the performance measures of different modeling techniques on two set of variables: baseline + SIRS variables and baseline + qSOFA variables. Considering either sensitivity or specificity a model selecting criteria is a debatable subject in the medical domain. Based on sensitivity, multivariate logistic regression and naïve Bayes performed well. Based on specificity, decision tree showed the best result. Figure 5 shows the receiver operating characteristic

curve using multivariate logistic regression for baseline risk, baseline + SIRS, and baseline + qSOFA. The qSOFA criteria demonstrate better discrimination power than SIRS—baseline (AUROC = 0.63; 95% CI, 0.61–0.63), baseline + SIRS (AUROC = 0.64; 95% CI, 0.64–0.65), and baseline + qSOFA (AUROC = 0.70; 95% CI, 0.69–0.70).

IV. Discussion

This study compared the prognostic power of SIRS and qSOFA. The results obtained using OR and modeling approaches indicated that qSOFA was more accurate than SIRS for assessing the risk of mortality among patients at the ED. In addition, the results showed that qSOFA criteria provide better balance between sensitivity and specificity.

The individual variable analysis of SIRS and qSOFA revealed interesting results (Table 3). We found that body temperature was not directly associated with 28-day in-hospital mortality. This may explain the poor performance of SIRS criteria and slight dip in mortality between scores of 3 and 4 on the SIRS. This finding is parallel to the result presented by Young et al. [20] where authors showed the evidence that elevated body temperature among patients with infection is associated with reduced in-hospital mortality. A close inspection of individual SIRS criteria showed that all variables except respiratory rate were weakly associated with in-hospital mortality. However, among qSOFA criteria, all variables were strongly associated with in-hospital mortality, with the Glasgow Coma Scale having the strongest association. The trends of mortality showed a steep increase in mortality with qSOFA scores, while steady change in mortality with SIRS score. The increase of each unit of qSOFA score provides more information about the mortality risk than the unit change in SIRS score. We also investigated the possibility of

either SIRS or qSOFA performing well to a group with population of specific characteristics. The investigation revealed that the association of qSOFA with mortality is greater than association of mortality with SIRS across all groups of different initial risk of mortality.

Further investigation found that although qSOFA ≥ 2 has stronger association with mortality than SIRS ≥ 2 for predicting in-hospital mortality among ED patients, it had lower sensitivity. Due to the poor sensitivity of qSOFA, patients with sepsis might remain undiagnosed. This misdiagnosis in the early stage could lead to life-threatening outcomes as timely treatment is critical [21]. However, the high sensitivity of SIRS could lead to unnecessary burdening of ICUs due to improper referrals. Researchers differ in their preference between sensitivity and specificity while selecting diagnostic criteria. Freund et al. argued that the high specificity of qSOFA criteria make it suitable to replace SIRS for efficient stratification of sepsis patients in the ED [1]. On the other hand, Askim et al. [22] preferred sensitivity and presented results against the use of qSOFA. Apart from the contradictory nature of SIRS and qSOFA about sensitivity and specificity, other performance characteristics such as positive and negative likelihood ratios showed evidence in favor of qSOFA. Later, we explain how carefully selection of threshold to determine labels from predicted probabilities, obtained from modeling approaches, can facilitate balance between sensitivity and specificity and make qSOFA more suitable for clinical use.

All three modeling approaches presented results in favor of qSOFA criteria. The AUROC was greater, in most cases, for qSOFA than SIRS. The findings are aligned with work that established the foundation of qSOFA criteria [13]. Since most patients with sepsis are initially assessed in the ED [23], several other studies compared the performance of SIRS and qSOFA. Freund et al. [1] noted the AUROC curve for predicting in-hospital mortality among ED patients was 0.80 for qSOFA and 0.65 for SIRS. One of the reasons for the difference in results between this study and ours is the use of different baseline risk variables. Churpek et al. [24] found that for non-ICU patients, the AUROC was 0.69 for qSOFA and 0.65 for SIRS, which is closely aligned with our results obtained using multivariate logistic regression. The sensitivity and specificity obtained from modeling approach is more balanced for both SIRS and qSOFA than obtained directly considering SIRS ≥ 2 and qSOFA ≥ 2 (Table 4). The reason for this is careful selection of threshold, instead of considering default 0.5, to determine the predicted labels from predicted probabilities. Adapting [11], we computed threshold that re-

sulted in a point in roc plot nearest to (0, 1). From Table 5, it is clear that sensitivity and specificity are greater for qSOFA than SIRS for most of the modeling approaches.

Our findings support the conclusions of other studies that qSOFA criteria can better differentiate low- and high-risk patients in the ED across varying levels of baseline risk. We know that early diagnosis of sepsis is the most effective way to reduce mortality [25]. Therefore, since qSOFA does not require any laboratory results, the use of qSOFA criteria in the ED could lead to early identification of critically ill patients and, consequently, improved outcomes.

This study has some limitations that should be considered when interpreting the results. We selected encounters using ICD-9 codes. However, these codes have been criticized in the literature due to lack of clear definitions [26,27]. Future studies are needed that overcome this limitation by using more precise methods of identifying patients at risk for sepsis. For example, in their assessment of clinical criteria for sepsis, Seymour et al. [13] used a combination of antibiotics and body fluid cultures occurring within a specific time-frame to define suspected infection.

Due to data coming only from hospitals using Cerner's EHR system, there could be potential sources of bias in the data. Although we included data from hospitals located in all four US Census regions, we cannot generalize the results to all US hospitals as there could be distinct differences between hospitals using the Cerner system compared to other hospitals. For some variables, the missing percentage was about 35%, therefore, the results can present slight biasness due to imputation. To minimize the induced biasness, we used imputation technique that substitutes unknown observation based on the known observation of the encounter. We relaxed the effect of intervention by assuming that each patient was subjected to a similar intervention. However, this is not usually the case, as intervention varies depending on illness severity. Therefore, we created deciles of baseline risk to compare the association of scores with in-hospital mortality over varying levels of baseline risk.

Our findings suggest that the discrimination power for qSOFA is greater than SIRS. The baseline risk analysis showed the robustness of qSOFA. Given the continued use of SIRS to assess mortality risk, the negative association of body temperature with mortality is of particular concern as this suggests normal body temperatures could increase risk of mortality. These findings contribute to our understanding of the prognostic power of qSOFA as a rapid bedside assessment requiring no laboratory data. With statistical and machine learning methods, we showed the advantages and

disadvantages of SIRS and qSOFA in terms of sensitivity and specificity. The qSOFA criteria performed better across varying risk of patients that signifies the robustness of the qSOFA criteria. This study also identified the Glasgow Coma Scale as the most important variable within the qSOFA clinical variables. We also found that careful selection of threshold to translate the predicted probabilities in labels can facilitate better balance between sensitivity and specificity. These findings have important implications for the implementation and use of sepsis-related clinical scoring systems.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

The authors would like to thank Krista Schumacher, PhD, from Oklahoma State University for editing support. The authors also like to acknowledge the data support from Elvena Fong (Oklahoma State University) and timely suggestions of Zhuqi Miao, PhD, from Oklahoma State University.

References

- Freund Y, Lemachatti N, Krastinova E, Van Laer M, Claessens YE, Avondo A, et al. Prognostic accuracy of Sepsis-3 criteria for in-hospital mortality among patients with suspected infection presenting to the emergency department. *JAMA* 2017;317(3):301-8.
- Fleischmann C, Thomas-Rueddel DO, Hartmann M, Hartog CS, Welte T, Heublein S, et al. Hospital incidence and mortality rates of sepsis: an analysis of hospital episode (DRG) statistics in Germany from 2007 to 2013. *Dtsch Arztebl Int* 2016;113(10):159-66.
- Torio CM, Andrews RM. National inpatient hospital costs: the most expensive conditions by payer, 2011 (Statistical Brief #160). Rockville (MD): Agency for Healthcare Research and Quality; 2006.
- Castellanos-Ortega A, Suberviola B, Garcia-Astudillo LA, Holanda MS, Ortiz F, Llorca J, et al. Impact of the Surviving Sepsis Campaign protocols on hospital length of stay and mortality in septic shock patients: results of a three-year follow-up quasi-experimental study. *Crit Care Med* 2010;38(4):1036-43.
- Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;315(8):801-10.
- Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest* 1992;101(6):1644-55.
- Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Heidelberg, Germany: Springer; 1996.
- Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. *Ann R Coll Surg Engl* 2006;88(6):571-5.
- Kim YM, Kathuria P, Delen D. Machine learning to compare frequent medical problems of African American and Caucasian diabetic kidney patients. *Healthc Inform Res* 2017;23(4):241-8.
- Kuo CY, Yu LC, Chen HC, Chan CL. Comparison of models for the prediction of medical costs of spinal fusion in Taiwan Diagnosis-Related Groups by machine learning algorithms. *Healthc Inform Res* 2018;24(1):29-37.
- Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006;163(7):670-5.
- Vincent JL. Endpoints in sepsis trials: more than just 28-day mortality? *Crit Care Med* 2004;32(5 Suppl):S209-13.
- Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, 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):762-74.
- Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, 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.
- Piri S, Delen D, Liu T, Zolbanin HM. A data analytics approach to building a clinical decision support system for diabetic retinopathy: developing and deploying a model ensemble. *Decis Support Syst* 2017;101:12-27.
- Murphy KP. Naive Bayes classifiers. Vancouver, Canada: University of British Columbia; 2006.
- Amland RC, Sutariya BB. Quick Sequential [Sepsis-Related] Organ Failure Assessment (qSOFA) and St.

- John Sepsis Surveillance Agent to detect patients at risk of sepsis: an observational cohort study. *Am J Med Qual* 2018;33(1):50-7.
18. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011;20(1):40-9.
 19. Milton JS, Arnold JC. *Introduction to probability and statistics: principles and applications for engineering and the computing sciences*. 4th ed. New York (NY): McGraw-Hill; 2003.
 20. Young PJ, Saxena M, Beasley R, Bellomo R, Bailey M, Pilcher D, et al. Early peak temperature and mortality in critically ill patients with or without infection. *Intensive Care Med* 2012;38(3):437-44.
 21. Nelson JL, Smith BL, Jared JD, Younger JG. Prospective trial of real-time electronic surveillance to expedite early care of severe sepsis. *Ann Emerg Med* 2011;57(5):500-4.
 22. Askim A, Moser F, Gustad LT, Stene H, Gundersen M, Asvold BO, et al. Poor performance of quick-SOFA (qSOFA) score in predicting severe sepsis and mortality: a prospective study of patients admitted with infection to the emergency department. *Scand J Trauma Resusc Emerg Med* 2017;25:56.
 23. Macdonald SP, Arendts G, Fatovich DM, Brown SG. Comparison of PIRO, SOFA, and MEDS scores for predicting mortality in emergency department patients with severe sepsis and septic shock. *Acad Emerg Med* 2014;21(11):1257-63.
 24. Churpek MM, Snyder A, Han X, Sokol S, Pettit N, Howell MD, et al. Quick Sepsis-related Organ Failure Assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. *Am J Respir Crit Care Med* 2017;195(7):906-11.
 25. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001;345(19):1368-77.
 26. Ollendorf DA, Fendrick AM, Massey K, Williams GR, Oster G. Is sepsis accurately coded on hospital bills? *Value Health* 2002;5(2):79-81.
 27. Sukanya C. Validity of principal diagnoses in discharge summaries and ICD-10 coding assessments based on national health data of Thailand. *Healthc Inform Res* 2017;23(4):293-303.