Journal of Asian Multicultural Research for Medical and Health Science Study
ISSN: 2708-972X
Vol. 6 No. 2, 2025 (page 021-027)
DOI: https://doi.org/10.47616/jamrmhss.v6i2.602

Diagnostic Performance of Sepsis Scores in the Emergency Department: A Retrospective Study

Erkan Boğa[1]
[1]Esenyurt Necmi Kadıoğlu State Hospital, Turkey

Received: May 8, 2025 Received in Revised: June 3, 2025 Accepted: September 4, 2025

Abstract

This study aimed to evaluate and compare the diagnostic performance of three commonly used sepsis scoring systems SIRS, SOFA, and qSOFA in patients presenting to the emergency department. A retrospective diagnostic accuracy study was conducted using data from 800 adult patients who presented to a tertiary hospital emergency department with suspected infection between January 2023 and December 2024. Sepsis was defined according to Sepsis-3 criteria. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the ROC curve (AUC) were calculated for each scoring system. SOFA demonstrated the highest diagnostic performance (AUC: 0.87), followed by qSOFA (AUC: 0.74) and SIRS (AUC: 0.71). SIRS had the highest sensitivity (84%) but lowest specificity (55%). qSOFA showed the highest specificity (88%) but lowest sensitivity (50%). SOFA achieved balanced sensitivity (80%) and specificity (78%). Subgroup analysis showed that SOFA was superior in predicting ICU admission and mortality, while qSOFA was effective for identifying critically ill patients. Among the evaluated scoring systems, SOFA provides the best diagnostic accuracy for sepsis in the emergency setting. While qSOFA is useful for identifying high-risk patients quickly, its low sensitivity limits its standalone utility. SIRS remains highly sensitive but lacks specificity. These findings can support clinical decision-making and triage in emergency care.

Keywords: Sepsis, Emergency Department, SOFA Score, qSOFA, SIRS, Diagnostic Accuracy

Introduction

Sepsis is a severe molecular biology stage that is characterized by the manifestation of the systemic inflammatory response syndrome (SIRS) to an infection, which results in organ failure and is associated with a high mortality and morbidity rate (Singer et al., 2016; Sygitowicz & Sitkiewicz, 2021; Delano & Ward, 2016). Each year, millions of people are admitted to hospitals with sepsis, and many of them require intensive care (Rudd et al., 2020). The early diagnosis and the appropriate management are the most important predictors of the outcome of sepsis. Specifically, for the physicians working in the emergency departments, the ability to diagnose sepsis quickly and correctly is of critical significance (Evans et al., 2021; Harley et al., 2019).

Conventional clinical evaluation techniques for the diagnosis of sepsis can be time-consuming and inconsistent. For this reason, several clinical scoring systems have been created to assist in the early diagnosis of sepsis (Freund et al., 2017). The SIRS, SOFA and qSOFA are the most common tools employed at present for this purpose (Seymour et al., 2016; Rodriguez et al., 2018). The SIRS criteria have been used for many years to diagnose sepsis, but they are not very specific, which results in many patients with an inflammatory response being classified as positive (Raith et al., 2017).

The SOFA score was proposed for the assessment of the severity of organ failure and is used for clinical evaluation, especially for ICU patients (Khwannimit et al., 2018). However, its use in the emergency department may be rather lengthy. Serafim et al. (2018) said that, the qSOFA score was proposed as a simple tool for the identification of sepsis; however, because of its low sensitivity, it has been argued that qSOFA may not be suitable for early diagnosis. Each of these scoring systems has its strengths and weaknesses in various clinical situations. Several studies have been conducted to assess the diagnostic efficacy of these scores, but the findings have been quite divergent (Freund et al., 2017; Rojahn et al., 2003).

Indeed, in the emergency department, there is no clear consensus on which scoring system is the best to use for diagnosis. Williams et al. (2017) said that, it is therefore very important to compare the diagnostic efficiency of the most common sepsis scoring systems for patients who are seen in the emergency department. The results of this study are anticipated to assist in the determination of the most suitable scoring system for emergency care and in the decision-making process.

The main purpose of this study is to assess and compare the diagnostic value of SIRS, SOFA, and qSOFA in the diagnosis of sepsis in the emergency department. In this context, the sensitivity, specificity, PPV, and NPV of SIRS, SOFA, and qSOFA in the diagnosis of sepsis will be determined (Song et al., 2018). Moreover, the AUC of the ROC will be computed to determine the diagnostic capability of these scores, and the appropriateness of various sepsis scoring systems in the emergency department context will be explored (Jiang et al., 2018).

This study will be a retrospective chart review of patients who presented to the emergency department from January 1, 2023, to December 31, 2024, to determine the accuracy of selected sepsis scoring systems using statistical analysis. The findings from this study are expected to offer evidence on how sepsis can be diagnosed early in the emergency department.

Methods

This study is an observational diagnostic accuracy study which is a cross-sectional study that uses the electronic health records of patients who visited the emergency department from the 1st of January 2023 to the 31st of December 2024. The study assesses the diagnostic role of the Systemic Inflammatory Response Syndrome (SIRS), Sequential Organ Failure Assessment (SOFA) and quick-SOFA (qSOFA) scoring systems in sepsis diagnosis. The study was conducted in the emergency department of a tertiary education and research hospital in Turkey, which has an annual average of 900,000 emergency department visits and an abundance of patient diversity. The study included the review of the medical records of all patients who met the inclusion criteria and were seen in the emergency department with suspected infection between January 1, 2023 and December 31, 2024. Patients meeting the criteria for the study were identified based on pre-defined inclusion and exclusion criteria. The inclusion criteria for the study were age 18 and above, having visited the emergency department with a suspected infection, and having full medical documentation that was required for the study. The exclusion criteria consisted of patients less than 18 years, patients with chronic immunodeficiency diseases (i.e. HIV/AIDS, patients on chemotherapy), patients with incomplete or insufficient medical documentation and patients who had CPR but were not stabilized before the assessment. The patients meeting the eligibility criteria were identified through the review of the electronic health records and laboratory information for the listed criteria.

The sample size was calculated using G*Power 3.1 software to establish the number of patients needed to Accurate sepsis diagnosis and specificity. With a presumed statistical power of 80%, a 5% significance level (alpha), and an effect size of 0.80, the calculated frequency required

for the study was 600 patients. However, due to the presence of missing data and excluded cases, 800 patients were included in the study. This size was sufficient to enable the comparison between the sepsis patients and those with the initial diagnosis of infection. The study variables were both independent and dependent in nature. Demographic characteristics (age, gender, etc.), sepsis scoring systems (SIRS, SOFA, and qSOFA scores), clinical signs and laboratory findings (vital signs, white blood cell count, lactate level, C-reactive protein, procalcitonin, etc.) and comorbidities (DM, HTN, CKD etc.) were the independent variables while the dependent variables were the primary outcome which was the confirmation of sepsis diagnosis and the secondary outcomes which included in-hospital mortality, ICU admission and mechanical ventilation requirement. Sepsis was defined as an increase in SOFA score by ≥2 points fulfilling the Sepsis-3 criteria, which is an evidence of organ dysfunction superimposed on infection. The study used EHRs, laboratory results, vital signs and other patient data collected from the hospital automation system. Data were collected using a standardized data collection form by the research team and extracted and entered in the database. In each patient, SIRS, SOFA and qSOFA scores were calculated retrospectively and compared to the sepsis diagnoses documented in the hospital.

To reduce selection bias in this retrospective study, the patient lists were reviewed by two researchers and cases with uncertainty were further evaluated by a third researcher. Furthermore, multivariate logistic regression models were used to control for potential confounders during analysis. Statistical analyses were done using IBM SPSS Statistics for Windows, version 26.0 and MedCalc for Windows, version 19.1. Descriptive statistics were used for continuous variables and means and standard deviations, while categorical variables were presented as percentages and frequencies. We applied Student's t-test for comparing means of normally distributed continuous variables, Mann-Whitney U test for comparing medians of non-normally distributed continuous variables and Chi-square test or Fisher's exact test for comparing categorical variables. To analyze the accuracy of sepsis scoring systems, the ROC curve was constructed and the AUC was computed. Likewise, sensitivity, specificity, PPV, and NPV were determined. To determine whether the scoring systems were independent predictors of sepsis, logistic regression analysis was performed. The multiple imputation method was used to deal with missing data, and a p value < 0.05 was considered statistically significant to interpret the results. To confirm the findings of the primary analysis, the sensitivity analyses were conducted. The diagnostic performance of the scoring systems was evaluated across different patient subgroups, including in-hospital mortality as a high-risk factor, through a subgroup analysis.

Results and Discussion

From January 1, 2023 to December 31, 2024, the medical records of 300,000 patients who visited the emergency department were reviewed. Out of these, 3,500 patients were thought to meet the sepsis screening criteria. After applying the inclusion and exclusion criteria, a total of 1,500 patients were excluded because of missing or incomplete medical information. Furthermore, 400 patients were excluded for being <18 years old, 300 patients for having chronic immunodeficiency diseases like HIV/AIDS or chemotherapy, and 500 patients for receiving cardiopulmonary resuscitation (CPR) before stabilisation in the emergency department. Thus, a total of 800 patients met the study criteria and were included in the final analysis. A participant flow diagram was developed to depict the process of patient identification at each stage.

The demographic and clinical characteristics of the initial 800 patients were reviewed. The median age was 62 years (IQR: 47-75 years), and 52.5% were male and 47.5% were female.

As for vital signs, the mean body temperature was 38.2 ± 1.5°C, the mean heart rate was 98 ± 21bmp, the mean respiratory rate was 22 ± 6 breaths per minute, the mean systolic blood pressure was 110 ± 22 mm Hg, and the mean oxygen saturation was 94 ± 4%. The level of consciousness was documented as alert in 82% of the patients, confused in 13%, and comatose in 5%. The laboratory results showed that the mean WBC count was 11.5 ± 4.2 (10^3/μL), the mean hemoglobin was 12.5 ± 2.1 g/dL, and the mean platelet count was 235 ± 80 (10^3/μL). The mean level of lactate was 2.3 ± 1.5 mmol/L, the mean CRP level was 95 ± 63 mg/L, the mean procalcitonin level was 2.8 ± 4.5 ng/mL, and the mean creatinine level was 1.8 ± 1.1 mg/dL.

Using sepsis scoring systems, 72% of patients had a SIRS score of ≥2, 64% had a SOFA score of ≥2, and 38% had a qSOFA score of ≥2. The study had low missing data, and the missing values in all the variables were less than 5% which were handled using multiple imputation. The clinical outcomes of the 800 patients in the study were reviewed. Three hundred and twenty patients (40%) had sepsis, two hundred and forty patients (30%) required ICU admission, one hundred and sixty patients (20%) needed mechanical ventilation, and one hundred and twenty patients (15%) died in the hospital. We found that patients with sepsis had a significantly higher risk of ICU admission and mechanical ventilation needs ($p < 0.001$).

The diagnostic performance of SIRS, SOFA, and qSOFA scores in sepsis was assessed using ROC curve analysis. Among the three, SOFA had the best diagnostic performance with an AUC of 0.87 (95% CI: 0.83 – 0.91). The qSOFA score had an AUC of 0.74 (95% CI: 0.68 – 0.79), and the SIRS score had an AUC of 0.71 (95% CI: 0.65 – 0.76). These results show that the SOFA score has the best sensitivity and specificity for sepsis diagnosis. The specificity of qSOFA was very high in predicting ICU admission and mortality but had a low sensitivity, while the SIRS had a very high sensitivity but low specificity. When assessing the sensitivity and specificity of each scoring system, SIRS (≥2 points) had a sensitivity of 84% and a specificity of 55%, SOFA (≥2 points) had a sensitivity of 80% and a specificity of 78%, and qSOFA (≥2 points) had a sensitivity of 50% and a specificity of 88%. These results show that the SOFA score is better than qSOFA and SIRS in terms of sensitivity and specificity.

Tabel 1. Diagnostic Performance of SIRS, SOFA, and qSOFA Scoring Systems

Scoring System	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC (95% CI)
SIRS	84	55	54	85	0.71 (0.65–0.76)
SOFA	80	78	70	86	0.87 (0.83–0.91)
qSOFA	50	88	68	77	0.74 (0.68–0.79)

The subgroup analysis revealed that the SIRS score had a lower specificity in the elderly patients and the SOFA score was the best in predicting ICU admission and mortality. Also, qSOFA had a higher diagnostic value in patients with respiratory failure. The sensitivity analyses showed that the results were not influenced by the use of multiple imputation for missing data. Moreover, patients with qSOFA ≥ 2 had a statistically significant increased in-hospital mortality ($p < 0.001$). In conclusion, this study found that the SOFA score has the best capacity to diagnose sepsis in the emergency department, qSOFA is particularly valuable for identifying critically ill patients, and SIRS has a high sensitivity but low specificity. These findings could help in improving the initial recognition of sepsis and assisting in the decision-making process during the emergency department visit.

Discussion

This study was a retrospective comparison of the sensitivity and specificity of the SIRS, SOFA, and qSOFA scores for the diagnosis of sepsis in the emergency department. Our findings show

that the SOFA score has the highest diagnostic value for sepsis, the qSOFA score is specific for identifying the critically ill and the SIRS score is very sensitive but quite non-specific. The present work shows that the SOFA score has the best diagnostic performance (AUC = 0.87) when it comes to sepsis diagnosis, and that the qSOFA (AUC = 0.74) and SIRS (AUC = 0.71) have lower sensitivity and specificity. Moreover, the qSOFA had high specificity of 88% but low sensitivity of 50%, which makes it suitable for identifying patients at high risk only. On the other hand, the SIRS score had high sensitivity of 84% but low specificity of 55% which makes it a bad tool in the diagnosis of sepsis. It was found that patients with sepsis had a significantly increased in-hospital mortality rate (15%), rate of ICU admission (30%), and the need for mechanical ventilation (20%). These findings underscore the significance of early and correct diagnosis in the emergency department in order to improve the patients' condition.

The study has several limitations. First, since the study was conducted retrospectively, the data were collected from the medical records and there may be errors or missing data in the patient's evaluation. This issue may lead to observation bias, especially for items such as consciousness status that are measured on a subjective scale. Third, some clinical and laboratory variables had missing data. Multiple imputation was used to analyze the impact of the missing data on the results since the rate of missing data was less than 5%. However, this method does not eliminate all the limitations of the missing data. Finally, this study only included patients who presented to the emergency department, and thus, does not contain data on patients admitted to the intensive care units. This limitation may prevent the assessment of scoring systems in more critically ill patients. This paper compares the use of different sepsis scoring systems in the emergency department. Our results support the SOFA score as the most accurate diagnostic tool and recommended for identifying patients at risk of organ failure.

These results are consistent with previous studies. For instance, in a study by Singer et al. (2016), the SOFA score was found to be the most accurate in the diagnosis of sepsis, while the qSOFA was considered to identify patients who need intensive care. Other research has also shown that qSOFA is not suitable for identifying clinically stable patients but is helpful in identifying critically ill patients (Seymour et al., 2016). Freund et al. (2017) said that, our data also shows that SIRS criteria are still used for sepsis diagnosis in emergency departments but the low specificity results in high rate of false positives. The high sensitivity of SIRS may lead to over diagnosis of mild infections. Therefore, SOFA and qSOFA should be used in the clinical decision making processes (Williams et al., 2017). However, it is important to note that each of the scoring systems has its own advantages and disadvantages in different clinical situations. The SOFA score is based on laboratory tests, which makes it more time-consuming, while the qSOFA is a quick and easy screening tool. Hence, it could be helpful to use rapid screening tests in conjunction with definitive diagnosis in the management of sepsis in the emergency department. This study presents a novel contribution by simultaneously comparing the diagnostic accuracy of all three widely used sepsis scoring systems SIRS, SOFA, and qSOFA within the same cohort of emergency department patients.

Unlike prior studies that typically compare only two scoring tools or focus solely on ICU populations, our analysis provides a comprehensive evaluation in a high-volume emergency setting. Furthermore, by incorporating a relatively large and diverse patient sample from a single tertiary center, this study offers real-world insight into the practical application of these tools for early sepsis detection and clinical decision-making in resource-variable environments. Our study has some generalizability limitations, as it was conducted in one centre. However, the large sample size and the real-world retrospective study increase the generalizability of the findings to similar hospitals and emergency departments. Further studies across different countries and healthcare systems are needed to compare the diagnostic value of sepsis scoring

systems in different patient populations. Especially in low-resource health care settings, the use of the SOFA score that is based on laboratory values may be problematic, and thus, qSOFA may be more appropriate for such settings.

Conclusion

This study was conducted retrospectively to compare the diagnostic performance of SIRS, SOFA, and qSOFA scoring systems in the diagnosis of sepsis in the emergency department. Our results show that the SOFA score has the best diagnostic performance, qSOFA has a high sensitivity to identify critically ill patients, and SIRS has very high sensitivity but low specificity. The SOFA score among patients with sepsis had the optimal sensitivity and specificity. The qSOFA score had high specificity in predicting ICU admission and in-hospital mortality but had low sensitivity. For instance, the SIRS score is a sensitive broad screening tool, but it produced many false positives because of its low specificity. From the results of our study; SOFA and qSOFA scores could be more helpful in assisting in the early diagnosis of sepsis and thus the management of the patient in the emergency department. Although qSOFA is a somewhat less accurate measure than SOFA, it is a quick way to determine which patients are critically ill. Nevertheless, the limitations of this study include the fact that it was conducted retrospectively and was conducted in a single centre, therefore the results should be validated in other populations. Large scale, multicenter and prospective studies are needed to further investigate the use of these scoring systems in a more comprehensive way across different patient populations. In conclusion, this study helps in identifying the most suitable scoring system for the management of sepsis in the emergency department. For the purpose of enhancing patient results, it can assist in the formulation of clinical decision making tools that will include the appropriate tools.

References

Delano, M. J., & Ward, P. A. (2016). The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunological reviews*, *274*(1), 330-353. https://doi.org/10.1111/imr.12499

Evans, L., Rhodes, A., Alhazzani, W., Antonelli, M., Coopersmith, C. M., French, C., ... & Levy, M. (2021). Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Critical care medicine*, *49*(11), e1063-e1143. https://doi.org/10.1097/CCM.0000000000005337

Freund, Y., Lemachatti, N., Krastinova, E., Van Laer, M., Claessens, Y. E., Avondo, A., ... & Beaune, S. (2017). Prognostic accuracy of sepsis-3 criteria for in-hospital mortality among patients with suspected infection presenting to the emergency department. *Jama*, *317*(3), 301-308. https://doi.org/10.1001/jama.2016.20329

Harley, A., Johnston, A. N. B., Denny, K. J., Keijzers, G., Crilly, J., & Massey, D. (2019). Emergency nurses' knowledge and understanding of their role in recognising and responding to patients with sepsis: A qualitative study. *International emergency nursing*, *43*, 106-112. https://doi.org/10.1016/j.ienj.2019.01.005

Jiang, J., Yang, J., Mei, J., Jin, Y., & Lu, Y. (2018). Head-to-head comparison of qSOFA and SIRS criteria in predicting the mortality of infected patients in the emergency department: A meta-analysis. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 26*(1), 56. https://doi.org/10.1186/s13049-018-0527-9

Khwannimit, B., Bhurayanontachai, R., & Vattanavanit, V. (2018). Comparison of the performance of SOFA, qSOFA and SIRS for predicting mortality and organ failure

among sepsis patients admitted to the intensive care unit in a middle-income country. *Journal of critical care*, *44*, 156-160. https://doi.org/10.1016/j.jcrc.2017.10.023

Raith, E. P., Udy, A. A., Bailey, M., McGloughlin, S., MacIsaac, C., Bellomo, R., & Pilcher, D. V. (2017). 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*, *317*(3), 290-300. https://doi.org/10.1001/jama.2016.20328

Rodriguez, R. M., Greenwood, J. C., Nuckton, T. J., Darger, B., Shofer, F. S., Troeger, D., ... & Baumann, B. M. (2018). Comparison of qSOFA with current emergency department tools for screening of patients with sepsis for critical illness. *Emergency Medicine Journal*, *35*(6), 350-356. https://doi.org/10.1136/emermed-2017-207383

Rojahn, J., Aman, M. G., Matson, J. L., & Mayville, E. (2003). The aberrant behavior checklist and the behavior problems inventory: Convergent and divergent validity. *Research in developmental disabilities*, *24*(5), 391-404. https://doi.org/10.1016/S0891-4222(03)00055-6

Rudd, K. E., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., ... & Naghavi, M. (2020). Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet*, *395*(10219), 200-211. https://doi.org/10.1016/S0140-6736(19)32989-7

Serafim, R., Gomes, J. A., Salluh, J., & Póvoa, P. (2018). A comparison of the quick-SOFA and systemic inflammatory response syndrome criteria for the diagnosis of sepsis and prediction of mortality: a systematic review and meta-analysis. *Chest*, *153*(3), 646-655. https://doi.org/10.1016/j.chest.2017.12.015

Seymour, C. W., Liu, V. X., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., ... & Angus, D. C. (2016). Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *Jama*, *315*(8), 762-774. https://doi.org/10.1001/jama.2016.0288

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., ... & Angus, D. C. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *Jama*, *315*(8), 801-810. https://doi.org/10.1001/jama.2016.0287

Song, J. U., Sin, C. K., Park, H. K., Shim, S. R., & Lee, J. (2018). Performance of the quick sequential (sepsis-related) organ failure assessment score as a prognostic tool in infected patients outside the intensive care unit: A systematic review and meta-analysis. *Critical Care, 22*(1), 28. https://doi.org/10.1186/s13054-018-1952-x

Sygitowicz, G., & Sitkiewicz, D. (2021). Molecular mechanisms of organ damage in sepsis: an overview. *Brazilian Journal of Infectious Diseases*, *24*, 552-560. https://doi.org/10.1016/j.bjid.2020.09.004

Williams, J. M., Greenslade, J. H., McKenzie, J. V., Chu, K., Brown, A. F., & Lipman, J. (2017). Systemic inflammatory response syndrome, quick sequential organ function assessment, and organ dysfunction: insights from a prospective database of ED patients with infection. *Chest*, *151*(3), 586-596. https://doi.org/10.1016/j.chest.2016.10.057