Scientific Reports

https://doi.org/10.1038/s41598-026-63089-8

Article in Press

Prognostic accuracy of qsofa for in-hospital mortality among adult patients presenting with suspected infection to the Emergency Department in the Middle East

Received: 7 September 2025

Accepted: 16 July 2026

Published online: 01 August 2026

Cite this article as: Algarni A.M., Alfaifi M.S., Bshabshe A. *et al.* Prognostic accuracy of qsofa for in-hospital mortality among adult patients presenting with suspected infection to the Emergency Department in the Middle East. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-63089-8

Abdullah M. Algarni, Musa S. Alfaifi, Ali Bshabshe, Othman M. Omair, Mohammed A. Alsultan, Hasan M. Alzahrani, Hadi E. Alali, Abdullah A. Alsabaani, Ali M. Alqarni, Tim R. Harris, Salah A. Alghanem, Bandar S. Mufareh, Ayesha M. Almemari, Abdulrahman A. Sindi, Ibrahim U. Ozturan, Abdullah A. Alhadhira, Asaad S. Shujaa, Abdullah H. Alotaibi, Moosa M. Awladthani, Ahmed A. Alsaad, Abdullah A. Almarshed, AlHanouv M. AlQahtani, Bader Alyahya, Saad A. Assiri, Feras H. Abuzeyad, Sara N. Kazim, Abdulrahman A. Al-Fares, Faisal Y. Almazroua, Naif T. Marzook, Abdullah A. Basri, Abdallah M. Elsafti, Nader Mohammed, Ayisha Ameen, Abdulaziz S. Alalshaikh, Cansu A. Özturan, Yousef I. Alawad, Awad AlOmari, Malek A. Alkhateeb, Moonis M. Farooq, Latifa Ali AlMutairi, Muneera M. Alasfour, Mohammad I. AlHaber, Umma Kulthum A. Umar, Nidal H. Bokhary, Abdulrhman Almutairi, Hisham F. Alyahya, Wejdan S. Alzahrani, Fawziah Alsalmi, Ahmed S. ALZomia, Abdulmajeed M. Omair, Faysal M. Alasmari, Sultan Y. Alfifi, Mohammed S. Al-Nujimi & Farid Foroutan

ARTICLE IN PRESS

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Prognostic Accuracy of qsofa for In-Hospital Mortality Among Adult Patients Presenting With Suspected Infection to the Emergency Department in the Middle East

Abdullah M. Algarni[1], Musa S. Alfaifi[2], Ali A. Al Bshabshe[3]✉, Othman M. Omair[4], Mohammed A. Alsultan[5], Hasan M. Alzahrani[3], Hadi E. Alali[2], Abdullah A. Alsabaani [3], Ali M. Alqarni [6], Tim R. Harris[7], Salah A. Alghanem[8], Bandar S. Al Mufareh[9], Ayesha M. Almemari10, Abdulrahman A. Sindi[11], Ibrahim U. Ozturan[12], Abdullah A. Alhadhira [13], Asaad S. Shujaa[13], Abdullah H. Alotaibi[14], Moosa M. Awladthani[15], Ahmed A. Alsaad [16], Abdullah A. Almarshed[17], AlHanouv M. AlQahtani[18], Bader A. Alyahya[19], Saad A. Assiri [20], Feras H. Abuzeyad [21], Sara N. Kazim [22], Abdulrahman A. Al-Fares [23], Faisal Y. Almazroua [24], Naif T. Marzook[25], Abdullah A. Basri[8], Abdallah M. Elsafti [7], Nader Mohammed [7], Ayisha Ameen[7], Abdulaziz S. Alalshaikh19, Cansu A. Özturan[12], Yousef I. Alawad [26], Awad AlOmari [27], Malek A. Alkhateeb [20], Moonis M. Farooq [21], Latifa Ali AlMutairi[28], Muneera M. Alasfour[28], Mohammad I. AlHaber10, Umma Kulthum A. Umar10, Nidal H. Bokhary11, Abdulrhman Almutairi2, Hisham F. Alyahya24, Wejdan S. Alzahrani25, Fawziah Alsalmi25, Ahmed S. ALZomia4, Abdulmajeed M. Omair[4], Faysal M. Alasmari[3], Sultan Y. Alfifi[3], Mohammed S. Al-Nujimi[3], and Farid Foroutan [29]

Corresponding Author: Ali A. Al Bshabshe[3]✉ **Email:** albshabshe@yahoo.com

[1]Family Medicine Department, Aseer Central Hospital, Abha, Saudi Arabia. [2]Emergency Medicine Department, Armed Forces Hospital Southern Region, Khamis Mushait, Saudi Arabia. [3]College of Medicine, King Khalid University, Abha, Saudi Arabia. [4]Emergency Medicine Department, Aseer Central Hospital, Abha, Saudi Arabia. [5]Saudi Commission for Health Specialties, Riyadh, Saudi Arabia. [6]Radiology Department, Prince Mashary Bin Saud Hospital, Belgraishi, Saudi Arabia. [7]Emergency Medicine Department, Hamad Medical Corporation, Doha, Qatar. [8]Emergency Medicine Department, Bahrain Defence Force Hospital, Al Riffa, Bahrain. [9]Emergency Medicine Department, Royal Commission Hospital in Jubail, Jubail, Saudi Arabia. [10]Emergency Medicine Department, Shaikh Shakhbout Medical City, Abu Dhabi, United Arab Emirates. [11]College of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia. [12]Kocaeli University, Faculty of Medicine, Emergency Medicine Department, Kocaeli, Turkey. [13]Emergency Medicine Department, Johns Hopkins Aramco Healthcare, Dhahran, Saudi Arabia. [14]Emergency Medicine Department, King Abdullah University Hospital, Riyadh, Saudi Arabia. [15]Critical Care Department, Armed Forces Hospital Oman, Muscat, Oman. [16]Emergency Medicine Department, King Fahad Specialist Hospital, Dammam, Saudi Arabia. [17]Emergency Medicine Department, King Fahad Medical City, Riyadh, Saudi Arabia. [18]Emergency Medicine Department, North Medical Tower Hospital, Arar, Saudi Arabia. [19]College of Medicine, King Saud University, Riyadh, Saudi Arabia. [20]Emergency Medicine Department, Sulaiman Al Habib Medical Group, Riyadh, Saudi Arabia. [21]Emergency Medicine Department, King Hamad University Hospital, Muharraq, Bahrain. [22]Emergency Medicine Department, Rashid Hospital, Dubai, United Arab Emirates. [23]Critical Care Department, Al Amiri Hospital, Kuwait City, Kuwait. [24]Emergency Medicine Department, King Saud Medical City, Riyadh, Saudi Arabia. [25]Emergency Medicine Department, King Fahad Armed Forces Hospital, Jeddah, Saudi Arabia. [26]Emergency Medicine Administration, King Fahad Medical City, Riyadh, Saudi Arabia. [27]Critical Care Department, Sulaiman Al Habib Medical Group, Riyadh, Saudi Arabia. [28]Emergency Medicine Department, Al Amiri Hospital, Kuwait City, Kuwait. [29]Health Research Methods, Evidence and Impact, McMaster University, Hamilton.

ARTICLE IN PRESS

ARTICLE IN PRESS

Abstract

The accuracy of systemic inflammatory response syndrome (SIRS) criteria, quick sequential organ failure assessment (qSOFA), National Early Warning Score (NEWS and NEWS2), and Modified Early Warning Score (MEWS) in predicting sepsis outcomes among Middle Eastern populations remains uncertain. This international, multicenter, prospective cohort study included 3,274 consecutive adult patients with suspected infections admitted through 20 emergency departments across Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, Turkey, and the United Arab Emirates. The primary outcome was 30-day in-hospital mortality. Of the participants (median age 62 years; 55.5% male), mortality was 8.8%. qSOFA showed the highest AUROC (0.75; 95% CI, 0.72–0.78), outperforming SIRS (0.63; 95% CI, 0.59–0.66) and MEWS (0.69; 95% CI, 0.66–0.72; P < .001), and was comparable to NEWS2 (0.75; 95% CI, 0.72–0.78; P = .99). Sensitivity was greater for NEWS ≥5 (73.9%), NEWS2 ≥5 (73.5%), and SIRS ≥2 (72.5%), compared with qSOFA ≥2 (42.2%) and MEWS ≥5 (40.1%). Conversely, specificity was highest for qSOFA ≥2 (89.0%) and MEWS ≥5 (83.0%), while SIRS ≥2 had the lowest specificity (48.6%). These findings suggest that although qSOFA and MEWS offer high specificity, their limited sensitivity risks missing many sepsis cases in emergency triage. NEWS2 provided the most balanced performance, although score selection should be guided by the intended clinical priority, including sensitivity for early identification, specificity for risk stratification, and local workflow.

Keywords

Systemic Inflammatory Response Syndrome (SIRS), Quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score (NEWS and NEWS2), and Modified Early Warning Score (MEWS)

ARTICLE IN PRESS

Introduction

Sepsis is a significant health problem worldwide, accounting for approximately half of hospital deaths and 10% of intensive care unit (ICU) admissions, with an in-hospital mortality rate of 10-50%.[1,2] Outcomes are improving in wealthier nations with widely used guidelines, improved awareness, and earlier detection and treatment.[3-5]Over two-thirds of sepsis cases may initially present to the emergency department (ED).[6] Using physiological and metabolic markers to screen for and risk assess patients with suspected infection is integral to care

ARTICLE IN PRESS

provision and early identification in the ED.[7,8] Commonly used screening tools include systemic inflammatory response syndrome (SIRS) criteria, national early warning scores one and two (NEWS[2]), and modified early warning score (MEWS).

The definition of sepsis has evolved since the early 1990s. In 2016, the Society of Critical Care Medicine/European Society of Intensive Care Medicine (SCCM/ESICM) task force redefined sepsis based on organ dysfunction and mortality prediction.[9] Sepsis is currently defined as life-threatening organ dysfunction caused by the dysregulated host response to infection. This definition has replaced the Sepsis-2 definitions proposed in 2001.[10] Although SIRS ≥2 were moderately sensitive, identifying 80-90% of high-risk ED patients with sepsis who died and/or required ICU admission, they had poor specificity.[11,12] It was also criticized for reflecting an appropriate host response frequently adaptive to various infectious and non-infectious processes. In contrast, the Sepsis-3 definitions emphasize the life-threatening organ dysfunction aspect of sepsis, which cannot be solely attributed to inflammation and infection[9] Due to these limitations, the task force recommended altering the risk assessment tool from the SIRS criteria to the acute change in sequential organ failure assessment (SOFA) score, with a point of 2 or more indicating sepsis-related organ dysfunction and increased in-hospital mortality. The shortened 'qSOFA' score was described for use outside the ICU to identify patients with suspected infection at increased risk of mortality where only physiological signs may be available.

The SCCM/ESICM task force recommended validating the Sepsis-3 definitions in different settings and populations. In line with this recommendation, we conducted this comparative validation study to assess the performance of SIRS criteria, qSOFA score, NEWS(2), and MEWS among ED patients with suspected infection in the Middle East populations.

Methods

ARTICLE IN PRESS

Study design and setting

We have described the methodology of the Prognostic Accuracy of qSOFA, SIRS, and Early Warning Scores (EWSs) for In-Hospital Mortality Among Adult Patients Presenting With Suspected Infection to the ED (PASSEM) study in detail elsewhere.[13] Briefly, PASSEM was an international, multicenter, prospective cohort study designed to evaluate the predictive performance of SIRS criteria, qSOFA score, NEWS(2), and MEWS for in-hospital mortality among adult patients presenting to the ED with suspected infection (ClinicalTrials.gov identifier: NCT05172479). The study recruited participants from 20 EDs across seven countries (Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, Turkey, and the United Arab Emirates), with recruitment and follow-up lasting 12 months at each center. The study protocol was approved by the relevant institutional and licensing committees at all participating locations. The Aseer Institutional Review Board served as the primary IRB, and all participating centers obtained ethical approval from their respective local IRBs prior to recruitment. The study was conducted in accordance with relevant guidelines and regulations. All methods were carried out in accordance with these guidelines and regulations. Written informed consent was obtained from all participants and/or their legal guardians prior to enrollment. This study was investigator-initiated and self-funded. The reporting of this study followed the TRIPOD-Cluster statement for studies validating a prediction model.[14]

ARTICLE IN PRESS

Diagnosis of infection

The PASSEM study followed a pragmatic approach in diagnosing suspected infection, best reflecting actual clinical practice and offering a high degree of generalizability. Participants were recruited if the attending ED physicians presumptively suspected or diagnosed infection based on the initial participant presentation without requiring exact criteria. The study team reviewed the notes post-recruitment to confirm the initial infection diagnosis. Where there was

ARTICLE IN PRESS

disagreement, this was resolved by two experts reviewing the medical record on day 30 through a comprehensive assessment of all clinical, physiological, microbiological, and diagnostic imaging data. In cases of disagreement between the two experts, consensus was sought. In all cases, the diagnosis of infection was blinded to the output of the prediction models and the outcomes of participants.

Study population and enrollment of participants

Study population

We enrolled all consecutive adult patients (age ≥18 years) who presented to the ED with suspected infection and who were subsequently admitted to the hospital between December 2021 and August 2023. Patients who presented to the ED due to non-infectious causes, pregnant women, those who were transferred from other hospitals or electively admitted to the hospital, prisoners, those who chose not to provide the informed study consent, or patients with a "Do-Not-Resuscitate" code status were excluded. We also excluded patients whose initial diagnosis of infection in the ED was not confirmed at the end of follow-up (30 days). We considered only the first hospitalization for each participant in cases of multiple readmissions during the study period.

ARTICLE IN PRESS

Enrollment of participants

A study team member screened all consecutive patients for eligibility and reviewed their physiological signs on arrival at the ED triage. Subsequently, the research team approached and enrolled potentially eligible participants. A web-based electronic data capture system (EDC) assigned each patient a participant number.15 The investigator then entered the participant's data in an electronic case report form. We collected the participant's initial physiological signs measured at ED triage or on arrival in the resuscitation area if directed to bypass triage. We also recorded the first biochemical and hematological measurements (partial

ARTICLE IN PRESS

pressure of carbon dioxide and white blood cell [WBC] count). If the participant was not eligible for recruitment, we closed the participant record in the EDC and captured the cause of exclusion. After enrollment, the study team monitored hospitalized participants' status using their unique medical registration number. The in-hospital follow-up care team was kept blinded to the initial score.

Outcomes

The primary outcome was 30-day in-hospital mortality. Participants who were transferred to another hospital or remained hospitalized at the 30-day follow-up cutoff were classified as not having the primary outcome, because the primary outcome was restricted to confirmed in-hospital death within 30 days at the recruiting hospital. Secondary outcomes included ICU admission within 30 days and prolonged ICU stay (≥72 hours). Due to administrative issues, we could not assess the all-cause mortality within 30 and 90 days.

Statistical analyses

Continuous variables were presented as median (IQR), and categorical variables as numbers (percentage). For better interpretability and comparability, continuous variables were adjusted by winsorizing at the 1% and 99% levels and then standardized with a mean of 0 and a standard deviation of 1.16 A complete-case analysis was conducted due to minimal missing data proportion (29 participants; <1%).[17]

The study's outcomes were expressed by the overall area under the receiver operating characteristic curve (AUROC) for each studied score with their corresponding asymptotic normal 95% confidence intervals.[18] The original regression coefficients used during the study scores' development could not be obtained despite multiple emails sent to get them from the corresponding authors. Hence, the calibration of these scores was not assessed. To evaluate studied scores cutoffs for outcomes, including SIRS ≥2, qSOFA ≥2, NEWS(2) ≥5, and MEWS

ARTICLE IN PRESS

≥5, we reported sensitivity, specificity, and positive and negative predictive values (PPV and NPV, respectively) along with their exact binomial 95% confidence interval.[19,20] Additionally, we reported the net reclassification indexes (NRIs) for these cutoffs and their bias-corrected and accelerated bootstrapped (1,000 replication) 95% confidence intervals, with recruiting hospitals being the resampling clusters.[21,22]

The respective odds ratios of the studied score for the study's outcomes were estimated with a mixed-effect logistic regression model. First, we built a nested baseline model (recruited hospitals nested in the countries). Fixed-effect variables included age and the Charlson Comorbidity Index (fractional polynomial), sex, race, hypertension, and infection site. Nested and random-effect models were evaluated and showed no significant differences; as a result, the country was omitted from the model. After fitting the baseline model, we added the studied score to the model (dichotomized to <2 and ≥2 for SIRS and qSOFA; and <5 and ≥5 for NEWS(2) and MEWS; as incremental score, and their components). The Kaplan-Meier method was employed to estimate time to 30-day in-hospital death, and the log-rank test was used to assess differences between groups. For these time-to-event analyses, participants who died in hospital within 30 days were treated as events at the recorded day of death. Participants discharged alive or transferred to another hospital before 30 days were censored at the date of discharge or transfer, and participants still hospitalized without in-hospital death at day 30 were censored at day 30. The crude primary outcome proportion was calculated separately as the number of confirmed in-hospital deaths within 30 days divided by the full analytic cohort.

Evidence to support a specific sample size choice for model validation studies is limited. One approach suggests a minimum of 100 events and 100 non-events are required to reach an adequate sample size,[23,24] with more than 200 events preferable.[25] In the current study, we targeted more than 200 events (in-hospital death within 30 days) for sample size calculation. Previous work by Freund et al. showed that recruiting 879 participants yielded 74 events (28-

ARTICLE IN PRESS

day in-hospital mortality) when power was set at 90% to ensure precise performance estimation.[26] Therefore, if we require at least 200 events and consider 20% of lost to follow-up and missing data, a sample size of 2,851 should suffice. All analyses were performed using Stata (version 18.0; StataCorp, College Station, TX, USA). A 2-tailed P <0.05 (with the Bonferroni correction when there are multiple comparisons) was considered statistically significant.

Results

Study cohort characteristics

Of the 3,782 participants screened, 307 (8.1%) were excluded due to not fitting the eligibility criteria. An additional 201 participants were ruled out; 172 (4.6%) were found to have no infection upon follow-up completion, and 29 (0.8%) had incomplete data. Consequently, 3,274 (86.6%) participants were included in the final analysis, as shown in Figure 1 and eFigure 1 in the Supplement. Table 1 presents the characteristics of the study cohort. The primary outcome (30-day in-hospital mortality) occurred in 287 (8.8%); 554 (16.9%) were admitted to the ICU, of whom 406 (73.3%) were admitted for 72 hours or more. By the 30-day in-hospital follow-up cutoff, 29 participants (0.9%) had been transferred to another hospital and 233 (7.1%) remained hospitalized; per protocol, these 262 participants were classified as not having the primary outcome. Participants who died in the hospital were more likely to be older men and had more severe comorbidities. They also had higher physiological scores and required ICU admission during their hospital stay. The most frequently observed evidence of infection was radiographic, with respiratory infections being the predominant source. Of all the participants, 53% had SIRS ≥2, 13.7% had qSOFA ≥2, 41.8% had NEWS(2) ≥5, and 19.1% had MEWS ≥5.

Study outcomes

ARTICLE IN PRESS

The median in-hospital follow-up was six days (IQR, 3-12 days). The crude 30-day in-hospital mortality proportion was 8.8% (287 of 3,274 participants). The Kaplan-Meier 30-day survival estimate was 69.83% under the censoring rules described above; this curve-based estimate is not the complement of the crude mortality proportion because patients discharged alive or transferred before 30 days were censored at their last in-hospital day. Participants who met predefined scores (≥2 for SIRS and qSOFA; ≥5 for NEWS[2] and MEWS) had significantly lower survival than those who did not (Figure 2 and eFigure 2 in Supplement) and higher odds of in-hospital mortality and ICU admission (Table 2 and eTable 1a-c in Supplement). Moreover, in-hospital mortality and ICU admission progressively increased with increasing physiological derangement but at variable rates of increase by each score (eFigure 3 in the Supplement). The qSOFA had a higher AUROC for in-hospital mortality than SIRS and MEWS but similar to NEWS(2) (Figure 3 and eFigure 4 in the Supplement). For ICU admission, NEWS2 had a higher AUROC than qSOFA, SIRS, and MEWS but similar to NEWS. Moreover, NEWS(2) had a higher AUROC than SIRS and MEWS but similar to qSOFA for prolonged ICU stay. The AUROC of NEWS and NEWS2 did not differ significantly in predicting any of the study's outcomes. NEWS(2) ≥5 and SIRS ≥2 were the most sensitive, while qSOFA ≥2 and MEWS ≥5 were the most specific for in-hospital mortality and ICU admission (Table 3 and eTable 2 in the Supplement). However, NEWS(2) ≥5 have the best trade-off between sensitivity and specificity. All the studied scores had intermediate to poor performance for prolonged ICU stay. qSOFA ≥2 has the highest PPV for in-hospital mortality and ICU admission. However, PPV was similar for all studied scores for prolonged ICU stay. Moreover, all analyzed scores had similar NPV. Relative to SIRS ≥2, NEWS(2) ≥5 provided the largest increase in relative NRI for in-hospital mortality and ICU admission but not prolonged ICU stay (eFigure 5a-c in the Supplement). On the other hand, qSOFA ≥2 and MEWS ≥5 provided the largest increase in absolute NRI for in-hospital mortality and ICU admission, but not prolonged ICU stay.

ARTICLE IN PRESS

Discussion

Key findings

In predicting in-hospital mortality among ED patients with suspected infection, qSOFA score demonstrated similar discrimination as NEWS(2) and better than SIRS criteria and MEWS by AUROC. However, qSOFA ≥2 had lower sensitivity than NEWS(2) ≥5 and SIRS ≥2, and similar to MEWS ≥5. The specificity of NEWS(2) ≥5 was better than SIRS ≥2 and reasonable compared to qSOFA ≥2 and MEWS ≥5. None of these scores performed sufficiently as a stand-alone risk assessment to identify the risk of death or ICU admission. Moreover, the study found no difference between NEWS and NEWS2. Therefore, the hypothesis that qSOFA score is more accurate than SIRS criteria and EWSs for predicting the studied outcomes in the ED is rejected.

Relationship with derivation studies

The present study's findings differ from the original cohorts, where the scores were derived regarding outcomes and predictive performance.[27-30] Such differences may be due to discrepancies in eligibility criteria, the settings in which participants were enrolled (ED, ICU, or general wards), how the scores were calculated (initial, best, or worst one), endpoint definitions, and follow-up duration. Our study utilized a pragmatic approach to defining suspected infection and aimed to validate the scores specifically in ED triage. Additionally, the initial physiological signs and laboratory indicators were captured for scores calculations, with participant recruitment spanning 20 EDs across seven countries. These differences lend different case mixes, which might affect the performance and generalizability of the scores in the current study compared to the derivation cohorts.[14]

Relationship with previous studies

ARTICLE IN PRESS

Multiple studies have externally validated the scores under study in the ED. In a recent meta-analysis of observational studies conducted in the ED, qSOFA ≥2 had poor sensitivity and moderate specificity. In contrast, SIRS ≥2 had moderate sensitivity and poor specificity for in-hospital mortality.[31] In a subsequent systematic review with high heterogeneity, NEWS ≥5 showed superior sensitivity compared to qSOFA ≥2. However, qSOFA ≥2 outperformed in specificity when predicting mortality and ICU admissions.[32] These findings generally correspond with the results of the current study.

NEWS2 is an updated version of NEWS that accounts for possible underlying type 2 respiratory failure (T2RF) in screened patients.[33] The current study did not find a difference between NEWS and NEWS2 in predicting the study's outcomes. There is limited data regarding the comparative performance of NEWS and NEWS2 in the ED; however, a retrospective study of hospitalized patients found no difference in predicting 24-hour in-hospital mortality, irrespective of T2RF presence.[34] These findings raise questions about the usefulness of implementing NEWS2.[35,36]

Interpretation and implementation

The SIRS criteria were part of the sepsis definitions for over two decades.[37,38] With the introduction of the Sepsis-3 definitions, there has been a shift toward using qSOFA score as a sepsis screening tool, similar to how SIRS criteria were previously employed.39 However, qSOFA is not included in the Sepsis-3 definitions and was not originally developed to screen for potential sepsis. Instead, it was recommended as a simple bedside tool to identify patients with suspected infection who are at risk for adverse sepsis outcomes outside the ICU.[39,40]

Employing qSOFA to screen patients with suspected infections during ED triage would miss many cases of sepsis, partially due to its involvement of criteria showing current decompensation, like hypotension, rather than indicating potential future deterioration.[31,41] Our

ARTICLE IN PRESS

findings support the recommendation of the 2021 Surviving Sepsis Campaign against using qSOFA as a single screening tool for sepsis or septic shock compared with SIRS, NEWS, or MEWS.[42] However, a pragmatic approach could involve leveraging the sensitivity of NEWS(2) to identify patients for rapid review in the triage and resuscitation area while implementing qSOFA or MEWS to identify those most likely to benefit from early referral to critical care teams. the selection of an early warning score should be guided by the intended clinical purpose, including whether the priority is to maximize sensitivity for early identification (rule-out strategy) or specificity for risk stratification (rule-in strategy), as well as local workflow considerations and available resources .It is also crucial to highlight that in our study, SIRS ≥2 may be falsely positive (lowering specificity) due to the prevalent benign ethnic neutropenia (BEN) in Middle Eastern populations.[43,44] BEN is an inherited cause of chronic mild to moderate neutropenia, which poses no risk of infections or other cytopenias. Individuals affected often have a WBC count below 4 x 10^3/µL. This notation is backed by a Norwegian study, which found that excluding the WBC count from the SIRS criteria and using a cutoff value of 2 or more moderately increased their specificity in predicting 7-day mortality among ED patients.[45] It can be argued that qSOFA and NEWS(2) have similar discrimination, as evidenced by their comparable AUROC. However, as suggested by some experts, the sensitivity of any scoring system should be prioritized when triaging patients with suspected infection in ED, as the ramifications of false negatives, such as overlooked diagnoses and treatment delays, are much more severe than those of false positives, which might lead to unneeded antibiotic and fluid administration.[46] One possible drawback of NEWS(2) is their incorporation of greater physiological signs than SIRS and qSOFA, leading to complexities in memorization and computation. Nevertheless, this could account for their superior performance, especially in the context of contemporary electronic health records systems. the study findings should be interpreted within the context of hospitalized ED patients with

ARTICLE IN PRESS

suspected infection and that caution is warranted when extrapolating the results to broader triage populations that include patients with less severe illness who do not require hospital admission

Strength and limitations

The current study has many strengths, including its prospective multicenter study design, participant assessment (ED triage) location, the low percentage of missing data, and the comprehensive data set. Nevertheless, there are some limitations. First, how we defined infection may have affected the scores' performance. However, we adopted this approach to be practical and to best reflect clinical practice. Secondly, SIRS performance may be improved if we exclude the WBC count, but this was not the study's goal. Thirdly, the study focused on the predictive performance of different scores without examining the effect of their implementation on adverse sepsis outcomes; a controlled trial would be needed for this. Fourthly, although obtaining physiological and laboratory markers at ED triage might be premature for predicting adverse sepsis outcomes, such scores are most helpful in identifying patients with suspected infection early. Fifthly, because participants transferred to another hospital or still hospitalized at 30 days were classified as not having the primary outcome, deaths occurring after transfer or beyond the 30-day in-hospital follow-up window may not have been captured. Finally, while experts were aware of the individual components of the scores, they were blind to the overall scores themselves.

ARTICLE IN PRESS

Conclusion

ARTICLE IN PRESS

While qSOFA and MEWS demonstrated high specificity, their sensitivity was low, suggesting they might miss many sepsis cases if employed as ED screening tools. The NEWS(2) were the most sensitive, and SIRS performed close to them. Because NEWS(2) showed high sensitivity with better specificity than SIRS and is widely used across diverse medical conditions, it may be incorporated into clinical judgment during ED triage to support sepsis identification and prompt treatment, while recognizing that no score should be used as a stand-alone screening tool.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ARTICLE IN PRESS

References

1.Angus DC, van der Poll T. Severe sepsis and septic shock. The New England journal of medicine. 2013;369(9):840-851.

2.Cohen J, Vincent JL, Adhikari NK, et al. Sepsis: a roadmap for future research. Lancet Infect Dis. 2015;15(5):581-614.

3.Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. The New England journal of medicine. 2003;348(16):1546-1554.

4.Kaukonen KM, Bailey M, Suzuki S, Pilcher D, Bellomo R. Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. JAMA. 2014;311(13):1308-1316.

5.Stevenson EK, Rubenstein AR, Radin GT, Wiener RS, Walkey AJ. Two decades of mortality trends among patients with severe sepsis: a comparative meta-analysis*. Crit Care Med. 2014;42(3):625-631.

6.Wang HE, Shapiro NI, Angus DC, Yealy DM. National estimates of severe sepsis in United States emergency departments.Crit Care Med. 2007;35(8):1928-1936.

7.Rhee C, Dantes R, Epstein L, et al. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009- 2014. JAMA. 2017;318(13):1241-1249.

8.Gauer R, Forbes D, Boyer N. Sepsis: Diagnosis and Management. American family physician. 2020;101(7):409-418.

ARTICLE IN PRESS

9.Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801-810.

10.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.

11.Churpek MM, Snyder A, Han X, 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):906-911.

12.Goulden R, Hoyle MC, Monis J, et al. qSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. Emerg Med J. 2018;35(6):345-349.

13.Algarni AM, Alfaifi MS, Al Bshabshe AA, et al. Prognostic accuracy of qSOFA score, SIRS criteria, and EWSs for in-hospital mortality among adult patients presenting with suspected infection to the emergency department (PASSEM) Multicenter prospective external validation cohort study protocol. PLoS One. 2024;19(1):e0281208. doi:10.1371/journal.pone.0281208

14.Debray TPA, Collins GS, Riley RD, et al. Transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster): explanation and elaboration. BMJ (Clinical research ed). 2023;380:e071058.

15.Castor EDC. Castor Electronic Data Capture. https://castoredc.com. Published 2023. Accessed 24 March 2023.

16.Steyerberg EW. Coding of Categorical and Continuous Predictors. In: Steyerberg EW, ed. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating. Second ed. Switzerland: Springer Nature; 2019:175-190.

17.Frank E. Harrell J. Missing Data. In: Frank E. Harrell J, ed. Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Second ed. Switzerland: Springer Nature; 2015:45-61.

18.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.

19.Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika.1934;26(4):404-413.

20.Deeks JJ, Altman DG. Sensitivity and specificity and their confidence intervals cannot exceed 100%. BMJ (Clinical research ed). 1999;318(7177):193-194.

21.Alba A, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. JAMA. 2017;318(14):1377-1384.

ARTICLE IN PRESS

ARTICLE IN PRESS

22.Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology (Cambridge, Mass). 2014;25(1):114-121.

23.Morise AP, Detrano R, Bobbio M, Diamond GA. Development and validation of a logistic regression-derived algorithm for estimating the incremental probability of coronary artery disease before and after exercise testing. J Am Coll Cardiol. 1992;20(5):1187-1196.

24.Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol. 2005;58(5):475-483.

25.Snell KIE, Archer L, Ensor J, et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. J Clin Epidemiol. 2021;135:79-89.

26.Freund Y, Lemachatti N, Krastinova E, 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):301-308.

27.Knaus WA, Sun X, Nystrom O, Wagner DP. Evaluation of definitions for sepsis. Chest. 1992;101(6):1656-1662.

28.Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):762-774.

29.Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. Qjm. 2001;94(10):521-526.

30.Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS--Towards a national early warning score for detecting adult inpatient deterioration. Resuscitation. 2010;81(8):932-937.

31.Fernando SM, Tran A, Taljaard M, et al. Prognostic Accuracy of the Quick Sequential Organ Failure Assessment for Mortality in Patients With Suspected Infection: A Systematic Review and Meta-analysis. Ann Intern Med. 2018;168(4):266-275.

32.Sabir L, Ramlakhan S, Goodacre S. Comparison of qSOFA and Hospital Early Warning Scores for prognosis in suspected sepsis in emergency department patients: a systematic review. Emerg Med J. 2021.

33.(RCP) RCoP. NHS England approves use of National Early Warning Score (NEWS) 2 to improve detection of acutely ill patients. https://www.rcplondon.ac.uk/news/nhs-england-approves-use-national-early-warning-score-news-2- improve-detection-acutely-ill. Published 2018. Accessed 17 November 2021, 2021.

34.Pimentel MAF, Redfern OC, Gerry S, et al. A comparison of the ability of the National Early Warning Score and the National Early Warning Score 2 to identify patients at risk of in-hospital mortality: A multi-centre database study. Resuscitation. 2019;134:147-156.

ARTICLE IN PRESS

35.Chen J. Does the National Early Warning Score 2 system serve its purpose? Resuscitation. 2019;134:145-146.

36.O'Driscoll R, Bakerly N, Murphy P, Turkington P. NEWS2 needs to be tested in prospective trials involving patients with confirmed hypercapnia. Resuscitation. 2019;139:369-370.

37.Bone RC, Balk RA, Cerra FB, 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.

38.Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference.Intensive Care Med. 2003;29(4):530-538.

39.Keeley A, Hine P, Nsutebu E. The recognition and management of sepsis and septic shock: a guide for non-intensivists.Postgrad Med J. 2017;93(1104):626-634.

40.Vincent JL, Martin GS, Levy MM. qSOFA does not replace SIRS in the definition of sepsis. Crit Care. 2016;20(1):210.

41.Singer M, Shankar-Hari M. qSOFA, Cue Confusion. Ann Intern Med. 2018;168(4):293-295.

42.Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med. 2021;47(11):1181-1247.

43.Atallah-Yunes SA, Ready A, Newburger PE. Benign ethnic neutropenia. Blood Rev. 2019;37:100586.

44.Donadieu J, Beaupain B, Fenneteau O, Bellanné-Chantelot C. Congenital neutropenia in the era of genomics: classification, diagnosis, and natural history. Br J Haematol. 2017;179(4):557-574.

45.Askim Å, Moser F, Gustad LT, 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(1):56.

46.Sartelli M, Kluger Y, Ansaloni L, et al. Raising concerns about the Sepsis-3 definitions. World J Emerg Surg. 2018;13:6.

Funding : No Funding

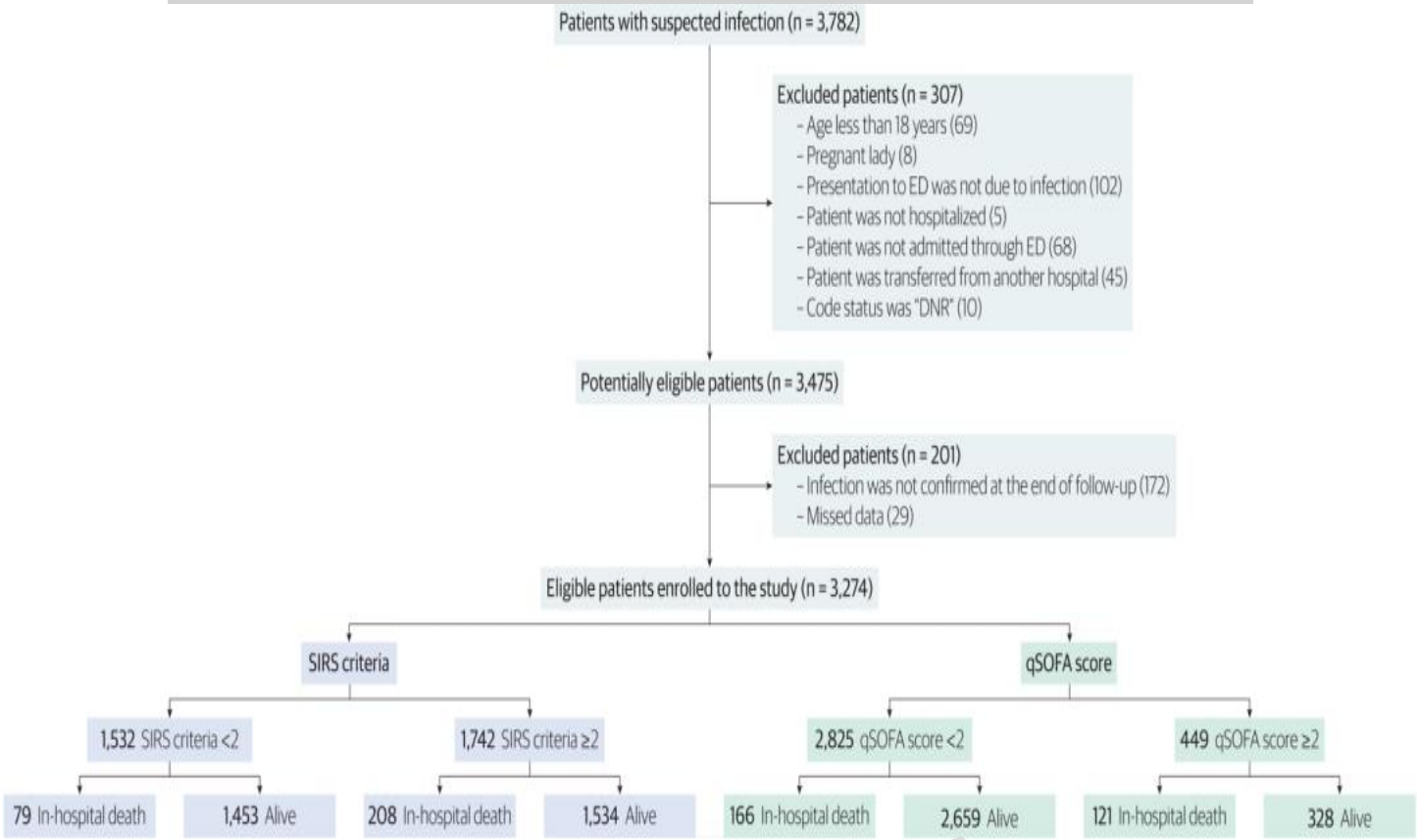
Figure

ARTICLE IN PRESS

Figure 1. Flow diagram of the study

ARTICLE IN PRESS

ARTICLE IN PRESS

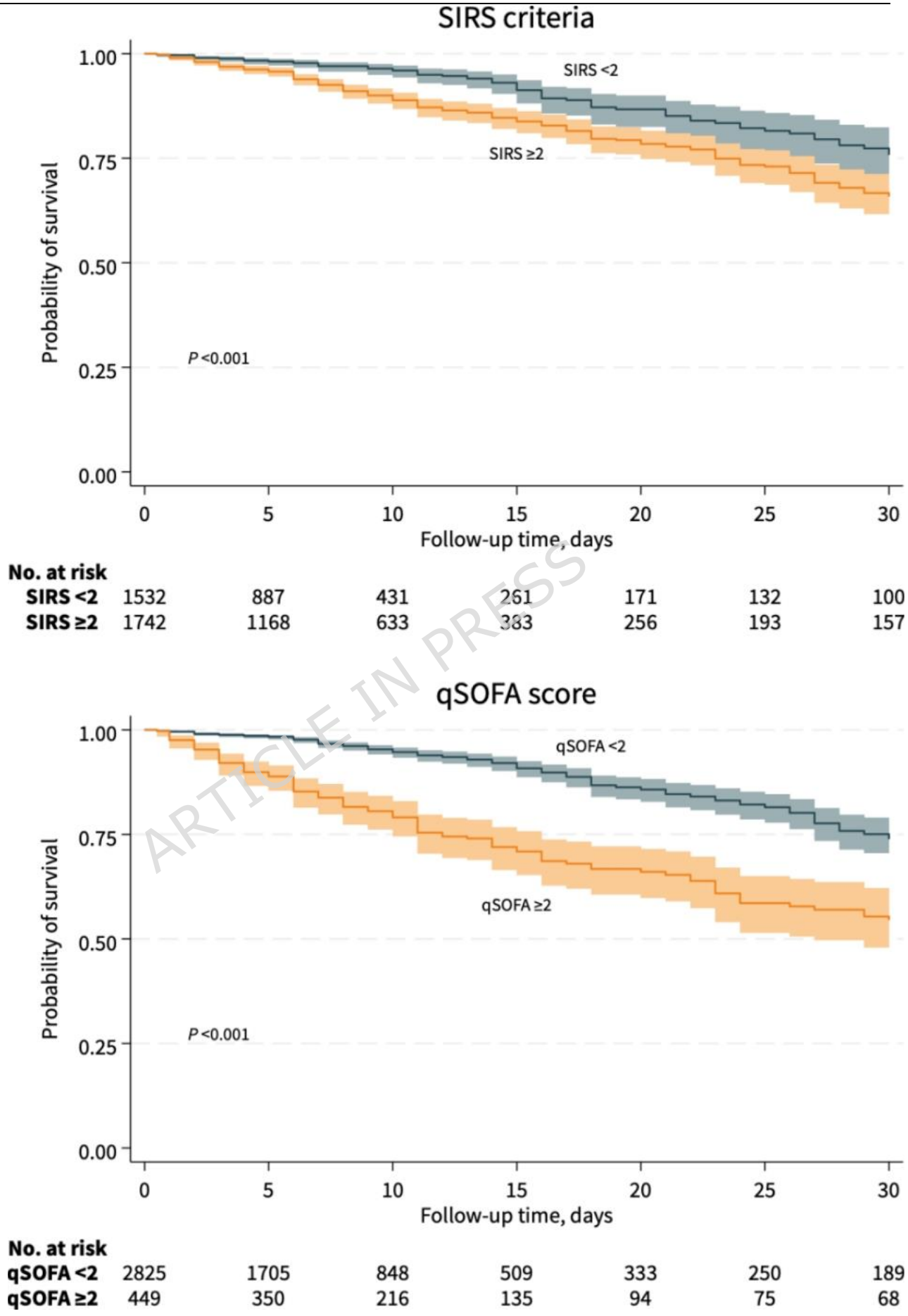


DNR: Do not resuscitate; ***ED:*** Emergency department; ***qSOFA:*** Quick sequential organ failure assessment; ***SIRS:*** Systemic inflammatory response syndrome.

ARTICLE IN PRESS

ARTICLE IN PRESS

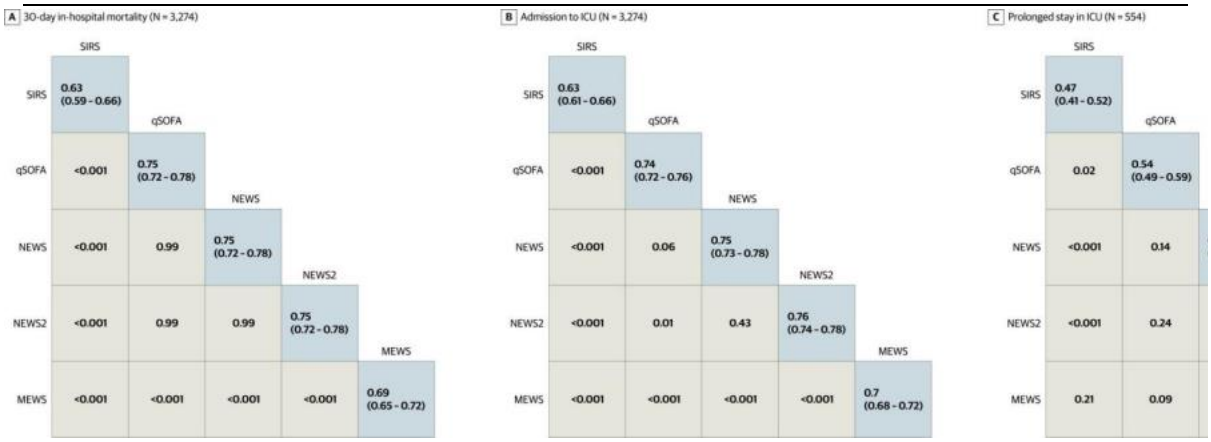
Figure 2: Kaplan-Meier survival estimates for participants who met SIRS ≥2 and qSOFA ≥2 versus those who did not



qSOFA: Quick sequential organ failure assessment; ***SIRS:*** Systemic inflammatory response syndrome.

ARTICLE IN PRESS

Figure 3: The area under the receiver operating characteristic curve of SIRS, qSOFA, NEWS(2), and MEWS for the study's outcomes



The area under the receiver operating characteristic curve (AUROC) data is in the blue-shaded diagonal cells (with 95% asymptotic normal confidence intervals). The corrected *P* values for multiple comparisons using the Bonferroni method are below the AUROC cells.

ICU: Intensive care unit; ***NEWS(2):*** National early warning score one and two; ***MEWS:*** Modified early warning score; ***qSOFA:*** Quick sequential organ function assessment; ***SIRS:*** Systemic inflammatory response syndrome.

ARTICLE IN PRESS

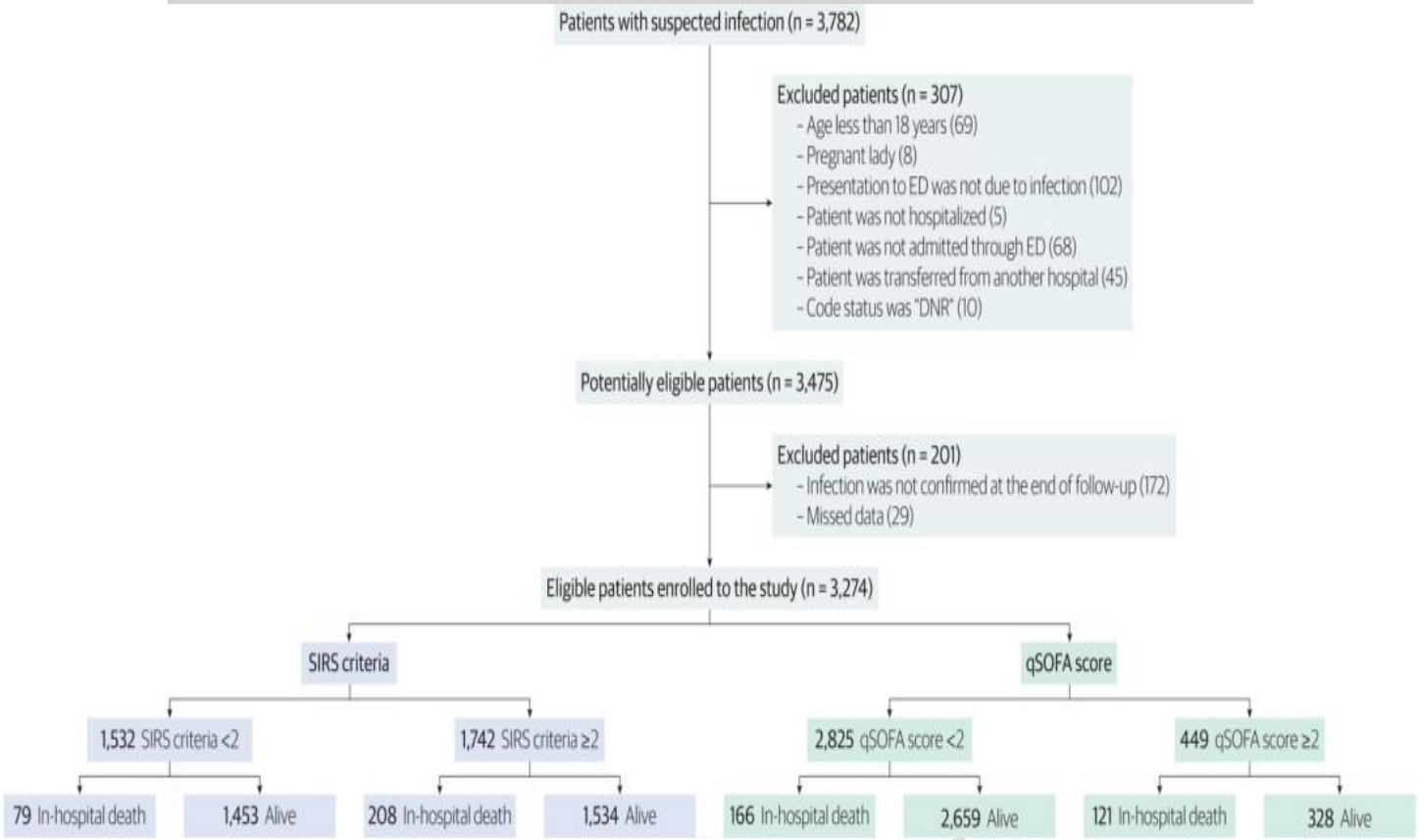
ARTICLE IN PRESS

Figure

Figure 1. Flow diagram of the study

ARTICLE IN PRESS

ARTICLE IN PRESS

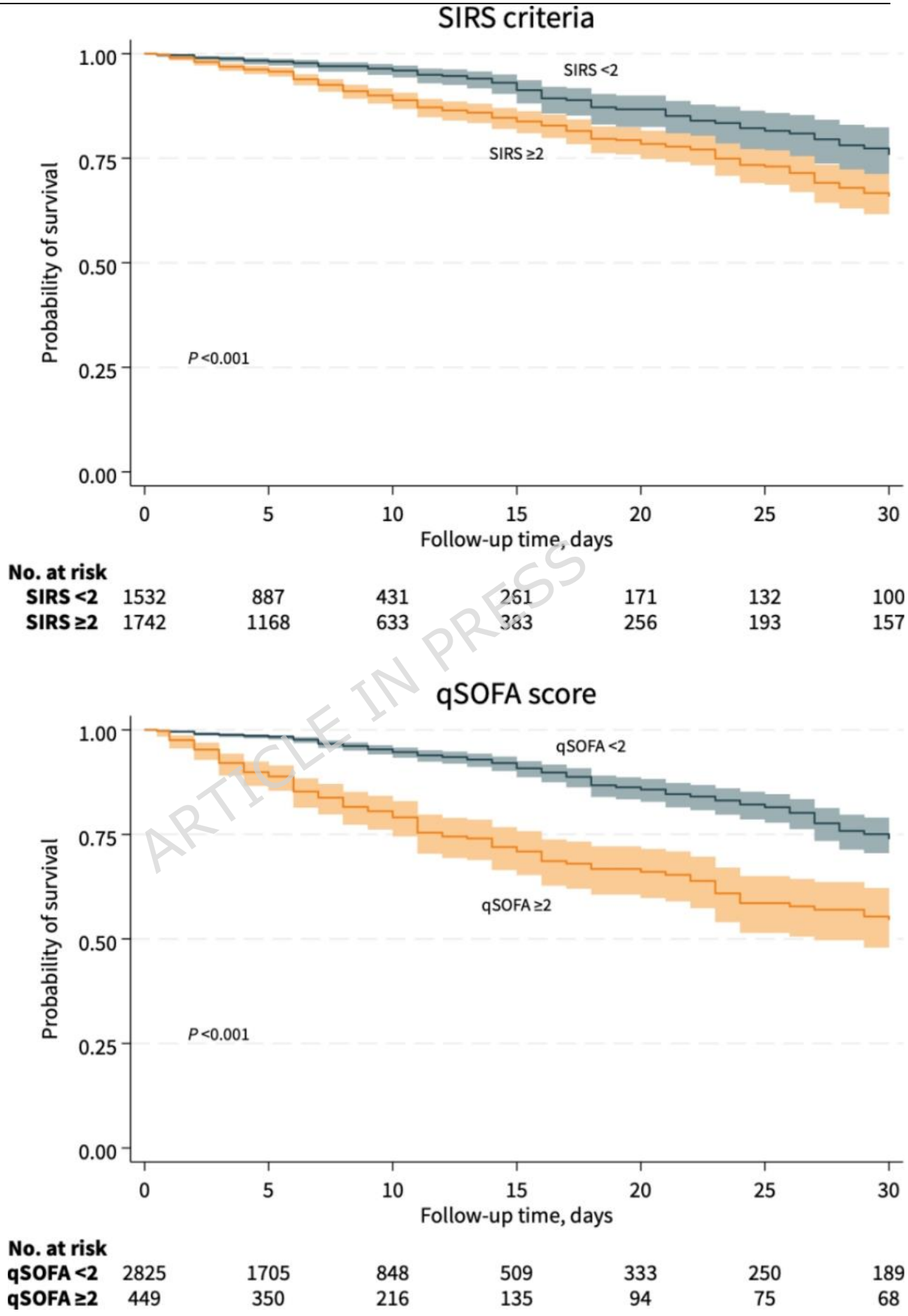


DNR: Do not resuscitate; ***ED:*** Emergency department; ***qSOFA:*** Quick sequential organ failure assessment; ***SIRS:*** Systemic inflammatory response syndrome.

ARTICLE IN PRESS

ARTICLE IN PRESS

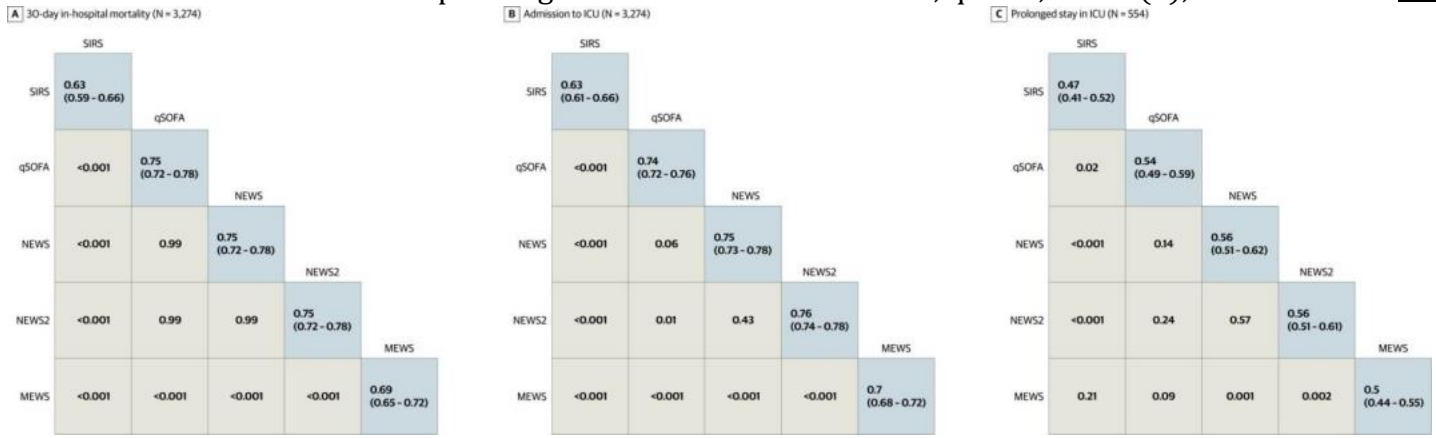
Figure 2: Kaplan-Meier survival estimates for participants who met SIRS ≥2 and qSOFA ≥2 versus those who did not



qSOFA: Quick sequential organ failure assessment; ***SIRS:*** Systemic inflammatory response syndrome.

ARTICLE IN PRESS

Figure 3: The area under the receiver operating characteristic curve of SIRS, qSOFA, NEWS(2), and MEWS for the study's outcomes



The area under the receiver operating characteristic curve (AUROC) data is in the blue-shaded diagonal cells (with 95% asymptotic normal confidence intervals). The corrected *P* values for multiple comparisons using the Bonferroni method are below the AUROC cells.

ICU: Intensive care unit; ***NEWS(2):*** National early warning score one and two; ***MEWS:*** Modified early warning score; ***qSOFA:*** Quick sequential organ function assessment; ***SIRS:*** Systemic inflammatory response syndrome.

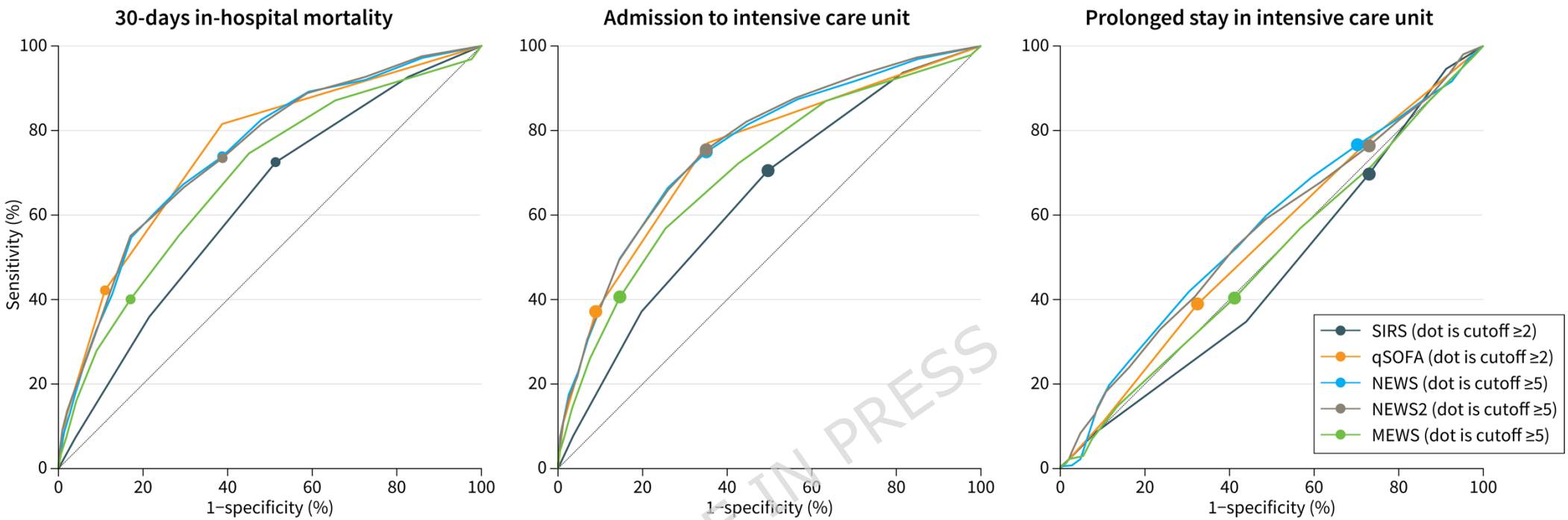
ARTICLE IN PRESS

ARTICLE IN PRESS

ARTICLE IN PRESS

ARTICLE IN PRESS

eFigure 4. ROC for 30-day in-hospital mortality, ICU admission, and prolonged ICU stay



ARTICLE IN PRESS

ARTICLE IN PRESS

eFigure 5a: Net reclassification indexes for 30-day in-hospital mortality

1 30-days in-hospital mortality

Patients with events (n = 287; 8.8% of total)

A SIRS *vs.* qSOFA

SIRS \ qSOFA	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	61	18	79 (27.5)
High risk (>9%)	105	103	208 (72.5)
Totals, qSOFA (%)	166 (57.8)	121 (42.2)	287 (100)

	Patients with events, No.	Patients without events, No.
Correct reclassification	18	1,278
Incorrect reclassification	105	67
Net reclassification	-87	1,206

Additive NRI (95% CI) 10.06 (-0.75 - 17.17) Net benifit (%, 95% CI) 0.88 (-0.04 - 1.59)
Absolute NRI (%, 95% CI) 34.18 (30.1 - 38.23) Relative utility (%, 95% CI) 10 (-0.39 - 18.8)

B SIRS *vs.* NEWS

SIRS \ NEWS	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	34	45	79 (27.5)
High risk (>9%)	41	167	208 (72.5)
Totals, NEWS (%)	75 (26.1)	212 (73.9)	287 (100)

	Patients with events, No.	Patients without events, No.
Correct reclassification	45	658
Incorrect reclassification	41	281
Net reclassification	4	377

Additive NRI (95% CI) 14.02 (6.44 - 20.54) Net benefit (%, 95% CI) 1.2 (0.45 - 1.8)
Absolute NRI (%, 95% CI) 11.64 (6.25 - 15.87) Relative utility (%, 95% CI) 14 (5.1 - 20.8)

C SIRS *vs.* NEWS2

SIRS \ NEWS2	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	35	44	79 (27.5)
High risk (>9%)	41	167	208 (72.5)
Totals, qSOFA (%)	76 (26.5)	211 (73.5)	287 (100)

	Patients with events, No.	Patients without events, No.
Correct reclassification	44	655
Incorrect reclassification	41	280
Net reclassification	3	375

Additive NRI (95% CI) 13.60 (6.61 - 20.73) Net benefit (%, 95% CI) 1.1 (0.45 - 1.7)
Absolute NRI (%, 95% CI) 11.55 (6.60 - 15.26) Relative utility (%, 95% CI) 13.6 (4.9 - 20.5)

D SIRS *vs.* MEWS

SIRS \ MEWS	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	67	12	79 (27.5)
High risk (>9%)	98	110	208 (72.5)
Totals, qSOFA (%)	165 (57.5)	122 (42.5)	287 (100)

	Patients with events, No.	Patients without events, No.
Correct reclassification	12	941
Incorrect reclassification	98	25
Net reclassification	-86	916

Additive NRI (95% CI) 1.91 (-7.02 - 7.72) Net benefit (%, 95% CI) 0.06 (-0.28 - 0.82)
Absolute NRI (%, 95% CI) 28.47 (25.48 - 31.53) Relative utility (%, 95% CI) 0.7 (-3.2 - 10.1)

Patients without events (n = 2,987; 91.2% of total)

SIRS \ qSOFA	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	1,386	67	1,453 (48.6)
High risk (>9%)	1,278	261	1,534 (51.4)
Totals, qSOFA (%)	2,659 (89)	328 (11)	2,987 (100)

SIRS \ NEWS	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	1,172	281	1,453 (48.6)
High risk (>9%)	658	1,157	1,534 (51.4)
Totals, NEWS (%)	1,830 (61.3)	1,157 (38.7)	2,987 (100)

SIRS \ NEWS2	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	1,173	280	1,453 (48.6)
High risk (>9%)	655	879	1,534 (51.4)
Totals, NEWS2 (%)	1,828 (61.2)	1,159 (38.8)	2,987 (100)

SIRS \ MEWS	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	1,428	25	1,453 (48.6)
High risk (>9%)	941	593	1,534 (51.4)
Totals, MEWS (%)	2,369 (79.3)	618 (20.7)	2,987 (100)

Overall (n = 3,287; 100% of total)

SIRS \ qSOFA	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	1,447	85	1,532 (46.8)
High risk (>9%)	1,378	364	1,742 (53.2)
Totals, qSOFA (%)	2,825 (86.3)	449 (13.7)	3,274 (100)

SIRS \ NEWS	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	1,206	326	1,532 (46.8)
High risk (>9%)	699	1,043	1,742 (53.2)
Totals, NEWS (%)	1,905 (58.2)	1,369 (41.8)	3,274 (100)

SIRS \ NEWS2	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	1,208	324	1,532 (46.8)
High risk (>9%)	696	1,046	1,742 (53.2)
Totals, NEWS2 (%)	1,904 (58.2)	1,370 (41.8)	3,274 (100)

SIRS \ MEWS	Low risk (≤9%)	High risk (>9%)	Totals, SIRS (%)
Low risk (≤9%)	1,495	37	1,532 (46.8)
High risk (>9%)	1,039	703	1,742 (53.2)
Totals, MEWS (%)	2,534 (80.9)	740 (19.1)	3,274 (100)

ARTICLE IN PRESS

ARTICLE IN PRESS

eFigure 5b: Net reclassification indexes for ICU admission

2 Admission to intensive care unit

Patients with events (n = 554; 16.9% of total)

A SIRS *vs.* qSOFA

SIRS \ qSOFA	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	129	34	163 (29.4)
High risk (>17%)	219	172	391 (70.6)
Totals, qSOFA (%)	348 (62.8)	206 (37.2)	554 (100)

	Patients with events, No.	Patients without events, No.
Correct reclassification	34	1,159
Incorrect reclassification	219	51
Net reclassification	-185	1,108

Additive NRI (95% CI) 7.34 (0.75 - 12.75) Net benifit (%, 95% CI) 1.2 (0.19 - 2.3)
Absolute NRI (%, 95% CI) 28.19 (24.19 - 32.44) Relative utility (%, 95% CI) 7.3 (1.1 - 13.9)

B SIRS *vs.* NEWS

SIRS \ NEWS	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	78	85	163 (29.4)
High risk (>17%)	61	330	391 (70.6)
Totals, NEWS (%)	139 (25.1)	415 (74.9)	554 (100)

	Patients with events, No.	Patients without events, No.
Correct reclassification	85	638
Incorrect reclassification	61	241
Net reclassification	24	397

Additive NRI (95% CI) 18.93 (14.26 - 23.40) Net benefit (%, 95% CI) 3.2 (2.4 - 4)
Absolute NRI (%, 95% CI) 12.86 (8.06 - 17.00) Relative utility (%, 95% CI) 18.9 (14.3 - 23.9)

C SIRS *vs.* NEWS2

SIRS \ NEWS2	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	76	87	163 (29.4)
High risk (>17%)	60	331	391 (70.6)
Totals, NEWS2 (%)	136 (24.5)	418 (75.5)	554 (100)

	Patients with events, No.	Patients without events, No.
Correct reclassification	87	636
Incorrect reclassification	60	237
Net reclassification	27	399

Additive NRI (95% CI) 19.54 (14.50 - 24.53) Net benefit (%, 95% CI) 3.3 (2.4 - 4.1)
Absolute NRI (%, 95% CI) 13.01 (8.34 - 16.96) Relative utility (%, 95% CI) 19.5 (14.3 - 24.4)

D SIRS *vs.* MEWS

SIRS \ MEWS	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	143	20	163 (29.4)
High risk (>17%)	186	205	391 (70.6)
Totals, MEWS (%)	329 (59.4)	122 (40.6)	554 (100)

	Patients with events, No.	Patients without events, No.
Correct reclassification	20	969
Incorrect reclassification	186	17
Net reclassification	-166	952

Additive NRI (95% CI) 5.04 (-1.07 - 9.98) Net benefit (%, 95% CI) 0.8 (-0.1 - 1.8)
Absolute NRI (%, 95% CI) 24.01 (20.95 - 27.32) Relative utility (%, 95% CI) 5 (-0.7 - 10.9)

Patients without events (n = 2,720; 83.1% of total)

SIRS \ qSOFA	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	1,386	51	1,369 (50.3)
High risk (>17%)	1,159	192	1,351 (49.7)
Totals, qSOFA (%)	2,477 (91.1)	243 (8.9)	2,720 (100)

SIRS \ NEWS	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	1,128	241	1,369 (50.3)
High risk (>17%)	638	713	1,351 (49.7)
Totals, NEWS (%)	1,766 (64.9)	954 (35.1)	2,720 (100)

SIRS \ NEWS2	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	1,132	237	1,369 (50.3)
High risk (>17%)	636	715	1,351 (49.7)
Totals, NEWS2 (%)	1,768 (65)	952 (35)	2,720 (100)

SIRS \ MEWS	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	1,352	17	1,369 (50.3)
High risk (>17%)	969	382	1,351 (49.7)
Totals, MEWS (%)	2,321 (85.3)	399 (14.7)	2,720 (100)

Overall (n = 3,287; 100% of total)

SIRS \ qSOFA	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	1,447	85	1,532 (46.8)
High risk (>17%)	1,378	364	1,742 (53.2)
Totals, qSOFA (%)	2,825 (86.3)	449 (13.7)	3,274 (100)

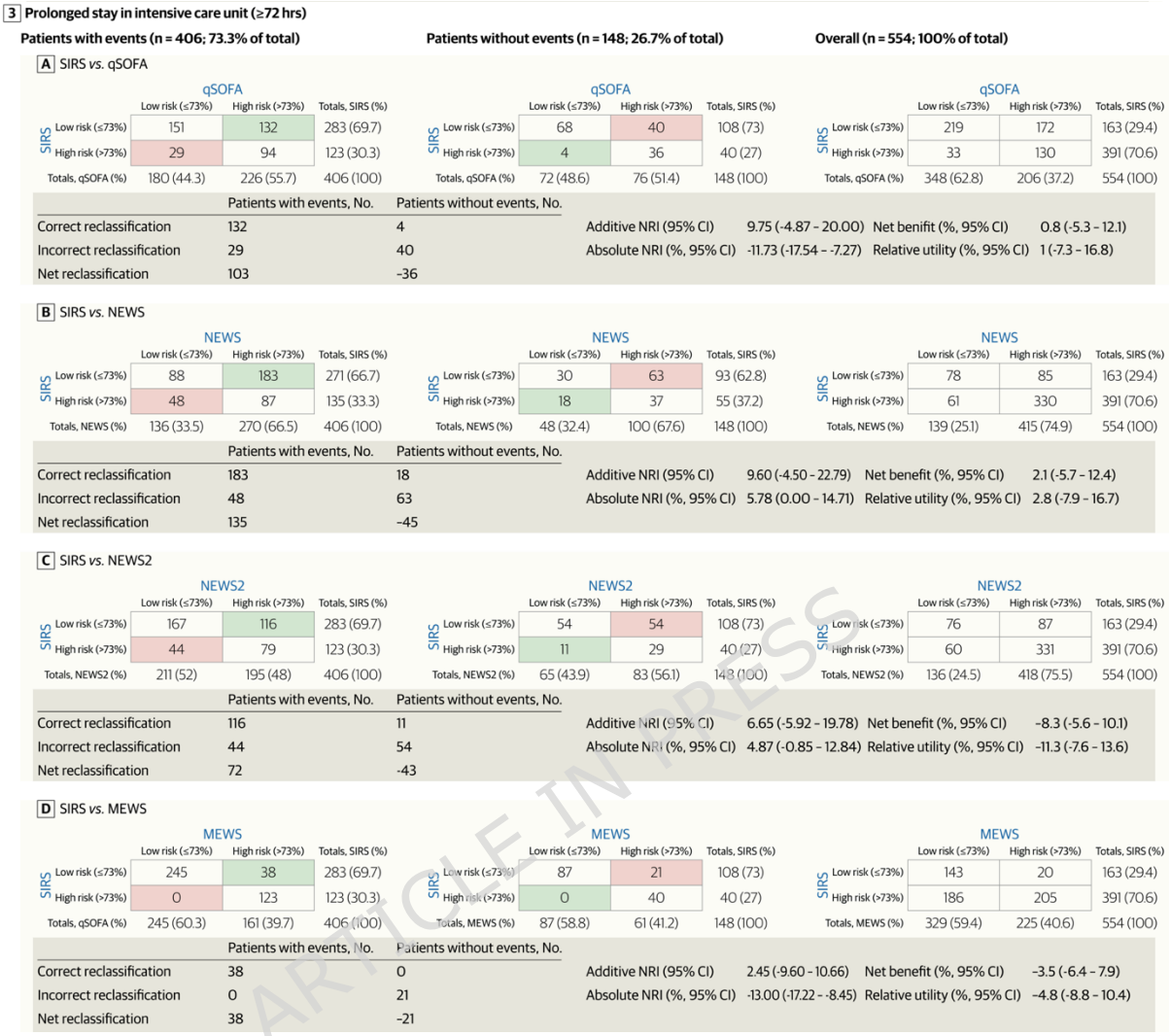
SIRS \ NEWS	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	1,206	326	1,532 (46.8)
High risk (>17%)	699	1,043	1,742 (53.2)
Totals, NEWS (%)	1,905 (58.2)	1,369 (41.8)	3,274 (100)

SIRS \ NEWS2	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	1,208	324	1,532 (46.8)
High risk (>17%)	696	1,046	1,742 (53.2)
Totals, NEWS2 (%)	1,904 (58.2)	1,370 (41.8)	3,274 (100)

SIRS \ MEWS	Low risk (≤17%)	High risk (>17%)	Totals, SIRS (%)
Low risk (≤17%)	1,495	37	1,532 (46.8)
High risk (>17%)	1,039	703	1,742 (53.2)
Totals, MEWS (%)	2,534 (80.9)	740 (19.1)	3,274 (100)

ARTICLE IN PRESS

eFigure 5c: Net reclassification indexes for prolonged ICU stay



3 Prolonged stay in intensive care unit (≥72 hrs)

	Patients with events (n = 406; 73.3% of total)			Patients without events (n = 148; 26.7% of total)			Overall (n = 554; 100% of total)		
A SIRS *vs.* qSOFA	qSOFA			qSOFA			qSOFA		
	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)
SIRS Low risk (≤73%)	151	132	283 (69.7)	68	40	108 (73)	219	172	163 (29.4)
SIRS High risk (>73%)	29	94	123 (30.3)	4	36	40 (27)	33	130	391 (70.6)
Totals, qSOFA (%)	180 (44.3)	226 (55.7)	406 (100)	72 (48.6)	76 (51.4)	148 (100)	348 (62.8)	206 (37.2)	554 (100)

	Patients with events, No.	Patients without events, No.				
Correct reclassification	132	4	Additive NRI (95% CI)	9.75 (-4.87 - 20.00)	Net benifit (%, 95% CI)	0.8 (-5.3 - 12.1)
Incorrect reclassification	29	40	Absolute NRI (%, 95% CI)	-11.73 (-17.54 - -7.27)	Relative utility (%, 95% CI)	1 (-7.3 - 16.8)
Net reclassification	103	-36				

B SIRS *vs.* NEWS	NEWS			NEWS			NEWS		
	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)
SIRS Low risk (≤73%)	88	183	271 (66.7)	30	63	93 (62.8)	78	85	163 (29.4)
SIRS High risk (>73%)	48	87	135 (33.3)	18	37	55 (37.2)	61	330	391 (70.6)
Totals, NEWS (%)	136 (33.5)	270 (66.5)	406 (100)	48 (32.4)	100 (67.6)	148 (100)	139 (25.1)	415 (74.9)	554 (100)

	Patients with events, No.	Patients without events, No.				
Correct reclassification	183	18	Additive NRI (95% CI)	9.60 (-4.50 - 22.79)	Net benefit (%, 95% CI)	2.1 (-5.7 - 12.4)
Incorrect reclassification	48	63	Absolute NRI (%, 95% CI)	5.78 (0.00 - 14.71)	Relative utility (%, 95% CI)	2.8 (-7.9 - 16.7)
Net reclassification	135	-45				

C SIRS *vs.* NEWS2	NEWS2			NEWS2			NEWS2		
	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)
SIRS Low risk (≤73%)	167	116	283 (69.7)	54	54	108 (73)	76	87	163 (29.4)
SIRS High risk (>73%)	44	79	123 (30.3)	11	29	40 (27)	60	331	391 (70.6)
Totals, NEWS2 (%)	211 (52)	195 (48)	406 (100)	65 (43.9)	83 (56.1)	148 (100)	136 (24.5)	418 (75.5)	554 (100)

	Patients with events, No.	Patients without events, No.				
Correct reclassification	116	11	Additive NRI (95% CI)	6.65 (-5.92 - 19.78)	Net benefit (%, 95% CI)	-8.3 (-5.6 - 10.1)
Incorrect reclassification	44	54	Absolute NRI (%, 95% CI)	4.87 (-0.85 - 12.84)	Relative utility (%, 95% CI)	-11.3 (-7.6 - 13.6)
Net reclassification	72	-43				

D SIRS *vs.* MEWS	MEWS			MEWS			MEWS		
	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)	Low risk (≤73%)	High risk (>73%)	Totals, SIRS (%)
SIRS Low risk (≤73%)	245	38	283 (69.7)	87	21	108 (73)	143	20	163 (29.4)
SIRS High risk (>73%)	0	123	123 (30.3)	0	40	40 (27)	186	205	391 (70.6)
Totals, qSOFA (%)	245 (60.3)	161 (39.7)	406 (100)	Totals, MEWS (%)	87 (58.8)	61 (41.2)	148 (100)	329 (59.4)	225 (40.6)	554 (100)

	Patients with events, No.	Patients without events, No.				
Correct reclassification	38	0	Additive NRI (95% CI)	2.45 (-9.60 - 10.66)	Net benefit (%, 95% CI)	-3.5 (-6.4 - 7.9)
Incorrect reclassification	0	21	Absolute NRI (%, 95% CI)	-13.00 (-17.22 - -8.45)	Relative utility (%, 95% CI)	-4.8 (-8.8 - 10.4)
Net reclassification	38	-21				

ARTICLE IN PRESS

Table 1. Baseline characteristics of the study participants

ARTICLE IN PRESS

30-days in-hospital mortality

ARTICLE IN PRESS

ARTICLE IN PRESS

Characteristics	All participants 3,274	Died 287 (8.8%)	Alive [a] 2,987 (91.2%)
Age, median (IQR), yrs	62 (40–76)	74 (62–84)	60 (39–75)
Men, No. (%)	1,817 (55.5)	142 (49.5)	1,675 (56.1)
Race, No. (%)			
Asian	89 (2.7)	4 (1.4)	85 (2.8)
Black	146 (4.5)	15 (5.2)	131 (4.4)
South Asian	238 (7.3)	13 (4.5)	225 (7.5)
White	2,801 (85.6)	255 (88.9)	2,546 (85.2)
Charlson's comorbidity index [b]			
Median (IQR)	4 (1–7)	7 (5–9)	4 (1–6)
Has score ≥3 score, No. (%)	2,111 (64.5)	263 (91.6)	1,848 (61.9)
Blood pressure, median (IQR), mm Hg			
Systolic blood pressure	122 (108–138)	110 (93–130)	122 (109–139)
Mean arterial blood pressure	88 (77–98)	80 (67–94)	88 (78–98)
Heart rate, median (IQR), beats/min	96 (82–110)	101 (85–116)	96 (82–110)
Oxygen saturation, median (IQR), %	96 (93–98)	94 (88–97)	96 (93–98)
Respiratory rate, median (IQR), breath/min	20 (19–22)	22 (20–28)	20 (19–22)
Temperature, median (IQR), °C	37.0 (36.7–37.8)	37.0 (36.6–37.6)	37.0 (36.7–37.8)
White blood cells count, median (IQR), ×10³/μL	10.6 (7.0–14.7)	12.7 (7.4–18.4)	10.4 (7.0–14.5)
SIRS [c]			
Median (IQR)	2 (1–2)	2 (1–3)	2 (1–2)
Has criteria ≥2, No. (%)	1,742 (53.2)	208 (72.5)	1,534 (51.4)
Individual component of criteria, No. (%)			
Tachycardia (HR >90 beat/min)	1,940 (59.3)	183 (63.8)	1,757 (58.8)
Tachypnea (RR ≥20 breath/min)	1,260 (38.5)	196 (68.3)	1,064 (35.6)
Fever (Temp. >38°C) or hypothermia (Temp. <36°C)	618 (18.9)	48 (16.7)	570 (19.1)
Leukocytosis (WBCs >12 x 10³/μL) or leukopenia (WBCs <4 x 10³/μL)	1,546 (47.2)	171 (59.6)	1,375 (46.0)
qSOFA [d]			
Median (IQR)	0 (0–1)	1 (1–2)	0 (0–1)
Has score ≥2, No. (%)	449 (13.7)	121 (42.2)	328 (11.0)
Individual component of score, No. (%)			
Altered level of consciousness (GCS <15)	589 (18.0)	140 (48.8)	449 (15.0)
[illegible]	[illegible]	[illegible]	[illegible]
NEW score [e]			
Median (IQR)	4 (1–7)	8 (4–11)	3 (1–6)
Has score ≥5, No. (%)[f]	1,369 (41.8)	212 (73.9)	1,157 (38.7)
NEW2 score, No. (%) [e]			
Median (IQR)	4 (1–7)	8 (4–10)	3 (1–6)
Has score ≥5, No. (%)[f]	1,370 (41.8)	211 (73.5)	1,159 (38.8)
MEW score [g]			
Median (IQR)	2 (1–4)	4 (2–6)	2 (1–4)
Has score ≥5, No. (%)	624 (19.1)	115 (40.1)	509 (17.0)
Evidence of infection, No. (%)			
Positive culture	784 (23.9)	94 (32.8)	690 (23.1)
Other microbiological methods [h]	621 (19.0)	81 (28.2)	540 (18.1)
Radiological studies	1,214 (37.1)	77 (26.8)	1,137 (38.1)
Clinical context	655 (20.0)	35 (12.2)	620 (20.8)

ARTICLE IN PRESS

Si

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 1. Baseline characteristics of the study participants

		30-days in-hospital mortality	
	All participants	**Died**	**Alive [a]**
Characteristics	**3,274**	**287 (8.8%)**	**2,987 (91.2%)**
Respiratory (excluding COVID-19)	934 (28.5)	95 (33.1)	839 (28.1)
Soft/cutaneous tissues	408 (12.5)	22 (7.7)	386 (12.9)
Urogenital	623 (19.0)	47 (16.4)	576 (19.3)
Unknown	52 (1.6)	12 (4.2)	40 (1.3)
Other [i]	44 (1.3)	0 (0.0)	44 (1.5)

COVID-19: Coronavirus disease-19; ***GCS:*** Glasgow coma scale; ***HR:*** Heart rate; ***MAP:*** Mean arterial pressure; ***MEWS:*** Modified early warning score; ***NEWS(2):*** National early warning score one and two; ***qSOFA:*** Quick sequential organ failure assessment; ***RR:*** Respiratory rate; ***SBP:*** Systolic blood pressure; ***SIRS:*** Systemic inflammatory response syndrome; ***Temp.:*** Temperature; ***WBCs:*** White blood cells.

a–Either discharged from the hospital, transferred to another hospital, or still alive by the end of the in-hospital follow-up phase (maximum 30 days).

b–Ranges from 0 to 37, with higher scores indicating a greater risk of 1-year mortality.

c– Ranges from 0 to 4, with higher scores representing a greater syndrome severity.

d–Ranges from 0 to 3, with higher scores indicating a greater likelihood of having poor outcomes.

e–Ranges from 0 to 20, with higher scores indicating a greater likelihood of clinical deterioration.

f–A score of ≥5 or <5 irrespective of whether any parameter is red; for example, the score can be <5 even if one parameter has a score of 3 (i.e., red in NEW(2) scores charts).

g– Ranges from 0 to 14, with higher scores indicating a greater likelihood of clinical deterioration.

h–For example, serological or histopathological studies.

i–Other causes of infections: Malaria (24; 0.7%), dengue fever (14; 0.42%), acute pericarditis (1; 0.03%), disseminated tuberculosis (2; 0.06%), infectious mononucleosis (1; 0.03%), lepromatous leprosy (1; 0.03%), Lyme disease (1; 0.03%), and bacterial keratitis in cancer patient (1, 0.03%).

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 2. Logistic regression models for 30-day in-hospital mortality (N = 3,207) [a]

	30-day in-hospital mortality, No.		
	Died	Alive	Adjusted OR (95% CI)
SIRS models			
SIRS criteria ≥2			
No	79	1,453	Reference
Yes	208	1,534	2.19 (1.61 – 2.97)
SIRS criteria			
0	21	516	Reference
1	58	937	1.54 (0.90 – 2.65)
2	105	893	2.76 (1.64 – 4.64)
3	82	521	3.00 (1.74 – 5.16)
4	21	120	4.52 (2.22 – 9.21)
qSOFA models			
qSOFA score ≥2			
No	166	2,659	Reference
Yes	121	328	3.99 (2.91 – 5.48)
qSOFA score			
0	53	1,831	Reference
1	113	828	3.57 (2.45 – 5.19)
2	85	270	7.68 (4.99 – 11.82)
3	36	58	12.19 (6.75 – 22.02)
NEW score ≥5			
No	75	1,830	Reference
Yes	212	1,157	3.20 (2.34 – 4.37)
NEW2 score ≥5			
No	76	1,828	Reference
Yes	211	1,159	2.89 (2.11 – 3.94)
MEW score ≥5			
No	172	2,478	Reference
Yes	115	509	2.78 (2.05 – 3.77)

MEWS: Modified early warning score; ***NEWS(2):*** National early warning score one and two; ***OR:*** odds ratio; ***qSOFA:*** Quick sequential organ failure assessment; ***SIRS:*** Systemic inflammatory response syndrome.

a—Adjusted for the age, Charlson comorbidity index, sex, race, hypertension, and infection source.

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 3. Performance of SIRS, qSOFA, NEWS(2), and MEWS for prediction of the 30-day in-hospital mortality

Outcome (95% CI) [a b]	SIRS ≥2	qSOFA ≥2	NEWS ≥5	NEWS2 ≥5
Sensitivity	72.5 (66.9 – 77.6)	42.2 (36.4 – 48.1)	73.9 (68.4 – 78.9)	73.5 (68.0 – 78.5)
Specificity	48.6 (46.8 – 50.5)	89.0 (87.8 – 90.1)	61.3 (59.5 – 63.0)	61.2 (59.4 – 63.0)
Positive predictive value	11.9 (10.5 – 13.6)	26.9 (22.9 – 31.3)	15.5 (13.6 – 17.5)	15.4 (13.5 – 17.4)
Negative predictive value	94.8 (93.6 – 95.9)	94.1 (93.2 – 95.0)	96.1 (95.1 – 96.9)	96.0 (95.0 – 96.8)
Additive NRI	Reference	10.06 (-0.75 – 17.17)	14.02 (6.44 – 20.54)	13.60 (6.61 – 20.73)
Absolute NRI	Reference	34.18 (30.10 – 38.23)	11.64 (6.25 – 15.87)	11.55 (6.60 – 15.26)

CI: Confidence interval; ***MEWS:*** Modified early warning score; ***NRI:*** Net reclassification index; ***NEWS(2):*** National early warning score one and two; ***qSOFA:*** Quick sequential organ failure assessment;

SIRS: Systemic inflammatory response syndrome.

a–Confidence intervals for sensitivity, specificity, and positive and negative predictive values are based on the exact Binomial method.

b–Confidence intervals for NRIs are based on bias-corrected and accelerated bootstrap method (1,000 replications), with recruiting hospitals being the resampling clusters.

ARTICLE IN PRESS