

ORIGINAL RESEARCH ARTICLE

Early identification of high-risk septic patients requiring critical care: Diagnostic accuracy of NEWS and qSOFA using hierarchical modeling

Muhammad Abubakar¹, Ubaid ur Rehman Khizar², Wagma Khan³, Muhammad Zeeshan⁴, Abdullah Ashfaq⁵, Misbah Shirin Memon⁶, Maheen Zahid⁷ , Abeer Murad Khamisani⁸, Muhammad Ahmad⁹ , and Ayoola Awosika^{9*} 

¹Accident and Emergency Department, Integrated Medical Care Hospital, Lahore, Punjab, Pakistan

²Accident and Emergency Department, Evercare Hospital, Lahore, Punjab, Pakistan

³Department of Pediatrics, District Headquarters Teaching Hospital, Mardan, Khyber Pakhtunkhwa, Pakistan

⁴Department of Internal Medicine, DHQ Hospital, Chiniot, Punjab, Pakistan

⁵Department of Trauma and Orthopedics, Kings Mill Hospital, Sutton in Ashfield, Nottinghamshire, United Kingdom

⁶Department of General Medicine, Pakistan Institute of Medical Sciences, Islamabad, Islamabad Pakistan

⁷Accident and Emergency Department, King Edward Medical University, Lahore, Punjab, Pakistan

⁸Accident and Emergency Department, Liaquat National Hospital and Medical College, Karachi, Sindh, Pakistan

⁹Department of Family and Community Medicine, College of Medicine Peoria, University of Illinois, Bloomington, Illinois, United States of America

*Corresponding author:

Ayoola Awosika
(ayoolaawosika@yahoo.com)

Citation: Abubakar M, Khizar UuR, Khan W, *et al.* Early identification of high-risk septic patients requiring critical care: Diagnostic accuracy of NEWS and qSOFA using hierarchical modeling. *Eurasian J Med Oncol*. doi: 10.36922/EJMO026250284

Received: June 17, 2026

Revised: August 13, 2026

Accepted: August 31, 2026

Published online: September 17, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Introduction: Early identification of septic patients at risk of deterioration is essential for care escalation and outcomes. This study evaluated the accuracy and utility of the National Early Warning Score (NEWS) and quick Sequential Organ Failure Assessment (qSOFA) for predicting mortality and identifying sepsis patients who may require critical care.

Methods: PubMed, Cochrane Library, and Google Scholar were searched for studies evaluating NEWS and qSOFA in adults with suspected infection or sepsis. Data were extracted to construct 2 × 2 tables. A bivariate random-effects model estimated sensitivity and specificity, and hierarchical summary receiver operating characteristic curves, diagnostic odds ratios (DORs), and post-test probabilities assessed utility.

Results: Ten studies involving 17,762 patients met eligibility criteria; five provided 2 × 2 data for in-hospital mortality synthesis. NEWS and qSOFA were associated with mortality (NEWS: risk ratio [RR] = 2.89, 95% confidence interval [CI]: 1.82–4.59; qSOFA: RR = 2.98, 95% CI: 1.84–4.83). NEWS had higher sensitivity (0.71) but lower specificity (0.56); qSOFA had lower sensitivity (0.44) and higher specificity (0.83). Pooled DORs were comparable (NEWS: 3.83; qSOFA: 3.82), indicating similar performance. At a pre-test probability of 15%, positive NEWS and qSOFA increased mortality to 22% and 31%, while negative results reduced it to 8% and 11%, showing complementary risk profiles.

Conclusion: NEWS and qSOFA show comparable prognostic performance but distinct diagnostic characteristics. NEWS may enable earlier detection of deterioration risk, whereas qSOFA may better identify higher-risk adverse outcomes. A complementary

sequential approach may support early stratification, timely intensive care escalation, and sepsis management decisions.

Keywords: Sepsis; Intensive care; Risk stratification; National Early Warning Score; quick Sequential Organ Failure Assessment; Mortality

1. Introduction

Sepsis is one of the most significant public health challenges in the world and remains a leading cause of death and morbidity. It is a clinical condition described as life-threatening organ dysfunction caused by a dysregulated host response to infection.¹ In spite of advances in the provision of antimicrobial therapy and supportive care for patients with sepsis, outcomes for patients with sepsis continue to be poor; particularly, among those who have delays in recognition and escalation of care. Therefore, early identification of patients at risk of clinical deterioration is critical to improving survival rates and optimizing resource use.²

Timely identification of high-risk patients upon presentation has direct implications for delivering critical care services. Delays in the identification of patients who require close or intensive monitoring, or those who need to be admitted to the intensive care unit (ICU), are associated with increased mortality, progression to multiple organ dysfunctions, and longer lengths of stay.³ Therefore, there is an increasing need for simple and reliable tools to assist in the performance of early risk stratification of high-risk patients and to assist with appropriate triage decisions within the continuum of emergency and critical care.⁴

Early warning scores are widely employed to determine those patients who are likely to develop a risk of clinical deterioration through the use of the physiological variables that are already available, including heart rate, respiratory rate, blood pressure, oxygen saturation, and state of consciousness.⁵ These measures include the National Early Warning Score (NEWS), which involves several physiological measurements and can help detect clinical deterioration early. On the other hand, the quick Sequential Organ Failure Assessment (qSOFA) measure includes signs of organ dysfunction and is thus useful for identifying high-risk patients who are at a greater risk of developing severe disease (Table 1).^{6,7}

Many studies have evaluated the predictive value of both the NEWS and qSOFA,^{8,9} but outcomes have varied widely, and it remains unclear how best to use each tool in current clinical practice. Additionally, the systematic reviews that have been conducted thus far have largely

included diverse patient populations with differing medical issues and across multiple outpatient and inpatient venues, making it hard to use those earlier studies for guidance in current decision-making situations.^{10,11} Moreover, most prior analyses of these tools' performance have not focused on the complementary use of NEWS and qSOFA to inform care escalation.

In light of the differential nature of the NEWS and qSOFA scoring systems, their joint use can offer a more holistic approach to early risk stratification. However, data on the clinical value of using NEWS and qSOFA together for risk stratification are scarce, particularly for identifying critically ill patients who would benefit from critical care services. As a result, this systematic review and diagnostic accuracy meta-analysis contributes to evaluating the potential application of NEWS and qSOFA to predict mortality and identify patients with suspected sepsis who may be at increased risk of clinical deterioration and may require transfer to higher levels of critical care. Hierarchical modeling techniques were used to develop a clinically relevant assessment of the complementary roles of these two tools in early risk stratification; thus, this study will guide clinical decision-making across the emergency-to-critical-care interface.

2. Materials and methods

A comprehensive literature review was performed using PubMed, Cochrane Library, and Google Scholar, covering the period from the inception of these databases to February 2026. The search strategy combined controlled vocabulary and free-text terms associated with sepsis, emergency department (ED) environments, early warning scores, and mortality outcomes. Boolean operators were used to combine key concepts. A representative search strategy included terms such as "sepsis," "emergency department," "early warning score," "NEWS," "qSOFA," "MEWS," and "mortality." Reference lists of relevant articles and reviews were also checked manually to identify studies that may not have been retrieved through the electronic search. The review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.

Table 1. Comparison of the clinical characteristics and diagnostic features of NEWS and qSOFA used for early risk stratification in patients with suspected sepsis

Component	NEWS	qSOFA
Primary clinical purpose	Early identification of physiological deterioration and need for escalation of care	Rapid identification of patients with suspected infection at increased risk of poor outcomes
Clinical context	Widely used in emergency departments, hospital wards, and acute care settings	Primarily used in patients with suspected sepsis outside the intensive care unit
Physiological variables included	Respiratory rate, oxygen saturation, supplemental oxygen requirement, temperature, systolic BP, heart rate, and level of consciousness	Respiratory rate, systolic BP, and altered mentation
Number of parameters	Seven physiological parameters	Three clinical parameters
Scoring method	Weighted aggregate score based on severity of physiological abnormalities	One point assigned for each abnormal criterion
Score range	0–20	0–3
Respiratory criterion	Weighted scoring based on respiratory rate abnormalities	Respiratory rate ≥ 22 breaths/min
BP criterion	Weighted scoring based on systolic BP abnormalities	Systolic BP ≤ 100 mmHg
Neurological assessment	AVPU scale or altered level of consciousness	Glasgow Coma Scale < 15 or altered mentation
Oxygen supplementation included	Yes	No
Temperature assessment included	Yes	No
Heart rate assessment included	Yes	No
Typical threshold used in sepsis studies	≥ 5 or ≥ 7	≥ 2
Principal diagnostic characteristic	Higher sensitivity for early clinical deterioration	Higher specificity for severe illness and mortality
Potential clinical advantage	Facilitates early recognition and monitoring of deteriorating patients	Identifies patients at higher risk of organ dysfunction and adverse outcomes
Potential limitation	Lower specificity may increase false-positive alerts	Lower sensitivity may fail to identify early deterioration
Relevance to critical care triage	Useful for early screening and escalation decisions	Useful for identifying patients requiring urgent critical care evaluation

Abbreviations: AVPU: Alert, Voice, Pain, Unresponsive; BP: Blood pressure; NEWS: National Early Warning Score; qSOFA: Quick Sequential Organ Failure Assessment.

2.1. Eligibility criteria

Studies were considered eligible when they met all of the following requirements: (i) enrolled adults aged ≥ 18 years who presented to the ED with suspected infection or sepsis; (ii) assessed NEWS, NEWS2, qSOFA, or another eligible early warning score at ED presentation or triage; and (iii) reported in-hospital or short-term mortality (≤ 30 days). Both prospective and retrospective observational cohort studies were eligible. Articles published in English or accompanied by an available English translation were considered.

Studies were excluded if they: (i) enrolled pediatric or disease-specific populations; (ii) were performed exclusively outside the ED, including studies restricted to

the ICU; (iii) assessed the scoring systems after hospital admission; or (iv) lacked data on eligible in-hospital or short-term mortality outcomes. ICU mortality was not considered an eligible standalone outcome because the review aimed to assess mortality prediction from the ED, whereas ICU mortality is conditional on subsequent ICU admission and may be influenced by ICU admission practices and selection. Systematic reviews, meta-analyses, editorials, case reports, and conference abstracts lacking adequate data were not eligible. However, reference lists of relevant reviews were examined to identify additional primary studies that met the inclusion criteria.

2.2. Study selection

Following the database searches, retrieved records

were imported into reference-management software for organization and duplicate removal. Two reviewers independently screened the titles and abstracts. Records considered potentially relevant at this stage underwent full-text assessment against the predefined eligibility criteria. Disagreements between the reviewers were discussed and resolved, with a third reviewer involved when consensus could not be reached.

2.3. Data extraction

Data collection was performed independently by two reviewers using a predefined extraction form. Information recorded for each study included the author, publication year, country, study design, and sample size. Patient-related variables comprised mean age, sex distribution, and the definition of sepsis applied in the respective study. Details of the evaluated early warning scores, including the threshold used and timing of assessment, were also documented.

Mortality rates and measures of prognostic performance were extracted for the relevant outcomes. These measures included sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC). Where the available findings allowed, 2×2 contingency tables were reconstructed using the reported sensitivity, specificity, and event-rate data.

2.4. Quality assessment

Risk of bias and methodological quality were assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. The assessment considered patient selection, index test, reference standard, and flow and timing. Each domain was rated as having low, high, or unclear risk of bias. Differences in assessment were resolved through reviewer consensus.

2.5. Data synthesis and statistical analysis

Analyses were carried out in R (version 4.5.2, R Foundation for Statistical Computing, Austria) and Review Manager (RevMan; version 5.4.1, The Cochrane Collaboration, UK), for diagnostic accuracy meta-analysis. The *mvmeta* (v1.0.3) and *metafor* (v5.0.1) packages were used to fit bivariate random-effects and hierarchical summary receiver operating characteristic (HSROC) models. The main analysis evaluated the prognostic performance of early warning scores for in-hospital mortality among patients presenting to the ED.

The pooled sensitivity and specificity estimates were obtained simultaneously through a bivariate random-effects model, which accounts for the correlation between the two measures as well as differences in diagnostic

thresholds across studies. Overall diagnostic performance was represented using HSROC curves.

Diagnostic odds ratios (DORs) were calculated to provide a single measure of overall discriminatory ability. These analyses, as well as sensitivity and specificity analyses, required dichotomous classification at a defined score threshold and therefore required sufficient data to construct 2×2 contingency tables. Where necessary, 2×2 tables were reconstructed from reported sensitivity, specificity, and event rates. Studies reporting AUROC were eligible for AUROC synthesis without requiring a common dichotomous threshold, allowing inclusion of studies that reported overall discrimination but did not provide sufficient data for threshold-specific 2×2 analyses. Additionally, likelihood ratios were used to estimate post-test probabilities based on a pre-test probability derived from the pooled mortality rate, allowing assessment of clinical utility.

Pooled AUROC values were calculated using a random-effects model to summarize overall discrimination. Between-study heterogeneity was assessed using the I^2 statistic and visual inspection of forest plots and HSROC space. Prespecified subgroup analyses were conducted based on type of scoring system (NEWS/NEWS2 vs. qSOFA) and mortality outcome (in-hospital vs. short-term mortality ≤ 30 days). Sensitivity analyses excluded studies using non-standard thresholds to assess the robustness of pooled estimates.

To explore potential sources of heterogeneity, an additional subgroup analysis was performed according to NEWS threshold values. Studies were categorized as using a threshold of NEWS = 5 or NEWS > 5, and pooled sensitivity and specificity estimates were compared between groups using random-effects logistic regression models. Publication bias was assessed using visual inspection of funnel plots, acknowledging the limitations of such methods in diagnostic accuracy meta-analyses.

3. Results

The database search yielded 642 records. After removal of duplicates, 629 records remained for screening. Title and abstract screening resulted in 23 articles being selected for full-text assessment. Following full-text evaluation, 10 studies fulfilled the predefined eligibility requirements and were retained for the final meta-analysis (Figure 1).

The included studies were published between 2018 and 2026 and conducted across Europe, Asia, and South America, reflecting broad geographic representation. A total of 17,762 patients were included, with individual study sample sizes ranging from 383 to 8,204. Five studies

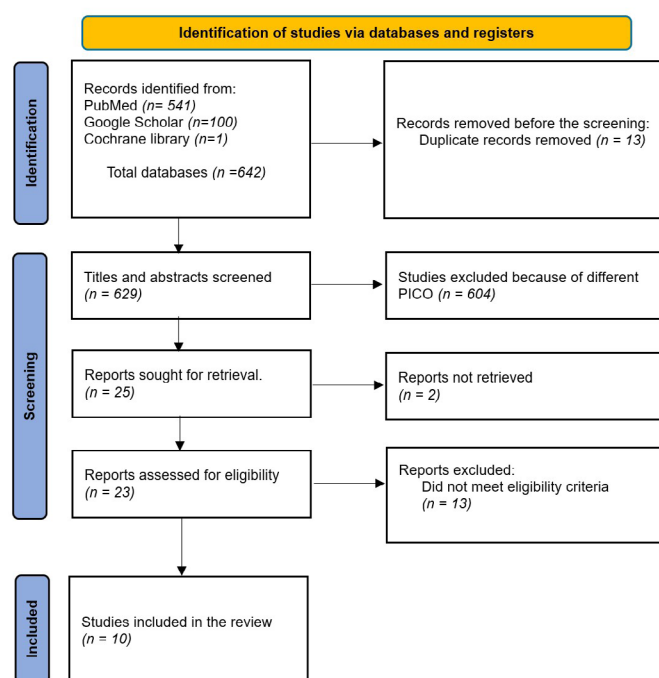


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart
Abbreviation: PICO: Population, Intervention, Comparison, and Outcome.

were retrospective, and five were prospective. All studies were conducted in ED settings, with three specifically evaluating triage-based assessment.

Study populations primarily consisted of patients with suspected infection or sepsis, although one study included patients with confirmed sepsis based on Sepsis-3 criteria. Six studies reported in-hospital mortality (primary outcome), while four reported short-term mortality (secondary outcome). All six in-hospital mortality studies contributed to the AUROC analysis; however, only five provided sufficient threshold-specific data to construct 2×2 contingency tables and were therefore included in the sensitivity/specificity and DOR analyses. The sixth study reported in-hospital mortality and discriminatory performance but did not provide a defined dichotomous threshold or sufficient data to construct a 2×2 table. The four studies reporting short-term mortality were analyzed separately as the secondary outcome.

Across studies, early warning scores evaluated included qSOFA, NEWS/NEWS2, Systemic Inflammatory Response Syndrome, and Modified Early Warning Score. Daorattanachai *et al.*¹² evaluated respiratory NEWS and was not included in the analysis. The qSOFA threshold was consistently ≥ 2 , whereas NEWS thresholds varied (≥ 5 to ≥ 8), with one study using non-integer cutoffs. Mortality rates ranged from 3.5% to 28.2%, reflecting heterogeneity

in patient severity. Detailed study characteristics are presented in [Table 2](#).

3.1. Risk of bias assessment

The aggregated findings from the current studies collectively indicate a low/moderate overall risk of bias through Quality Assessment of Diagnostic Accuracy Studies-2 analysis. The areas of patient selection and reference standard received low-risk classifications mostly because they were observed in consecutive samples of ED patients who all had very definitive mortality endpoints. Since there were very few reporting requirements on blinded assessments for the index test area, this area carried a few minor concerns. Variations in follow-up and outcome assessment in the flow and timing area denote a mix of clear and moderate risk. Applicability risk across studies is generally low, with only minor variations across patient subgroups and clinical settings.

3.2. Primary outcome: In-hospital mortality

3.2.1. Forest plot analysis

For NEWS, higher scores were significantly associated with increased in-hospital mortality (risk ratio [RR] = 2.89, 95% confidence interval [CI]: 1.82–4.59), indicating nearly a threefold increased risk. Substantial heterogeneity was observed ($I^2 = 87\%$, $p < 0.001$) ([Figure 2A](#)). Similarly,

Table 2. Characteristics of included studies ($n = 10$)

Study	Year	Country	Design	Setting	Sample Size	Population	Outcome	Threshold	Mortality (%)	Triage-based
Brink <i>et al.</i> ¹³	2019	Netherlands	Retrospective	ED	8,204	Suspected sepsis	10-day and 30-day mortality	qSOFA: ≥ 2 NEWS: ≥ 7	6.00%	No
Goulden <i>et al.</i> ¹⁴	2018	UK	Retrospective	ED admissions	1,818	Suspected infection	In-hospital mortality	qSOFA: ≥ 2 NEWS: ≥ 5	15%	No
Azijli <i>et al.</i> ¹⁵	2021	Netherlands	Prospective	ED	1,328	Suspected infection	30-day mortality	qSOFA: ≥ 2 NEWS: ≥ 5	3.50%	No
Oduncu <i>et al.</i> ¹⁶	2021	Turkey	Prospective	ED	463	Suspected infection	30-day mortality	qSOFA: ≥ 2 NEWS: ≥ 5	18.10%	No
Suttapanit <i>et al.</i> ¹⁷	2022	Thailand	Retrospective	ED triage	550	Sepsis/infection	28-day mortality	qSOFA: ≥ 2 NEWS: ≥ 5	12.40%	Yes
Ruangsomboon <i>et al.</i> ¹⁸	2021	Thailand	Retrospective	ED	1,622	Suspected sepsis	In-hospital mortality	qSOFA: ≥ 2 NEWS: ≥ 8	28.20%	No
Boekhoud <i>et al.</i> ¹⁹	2024	Netherlands	Prospective	ED	1,516	Suspected infection/sepsis	In-hospital mortality	qSOFA: ≥ 2 NEWS: ≥ 5	3.60%	No
Devia-Jaramillo <i>et al.</i> ²⁰	2025	Colombia	Retrospective	ED triage	383	Adults with confirmed sepsis (Sepsis-3)	In-hospital mortality	qSOFA: ≥ 1.5 NEWS: ≥ 7.5	20.60%	Yes
Daorattanachai <i>et al.</i> ¹²	2025	Thailand	Prospective	ED	1,348	Suspected sepsis	In-hospital mortality	–	13.40%	No
Xythalis <i>et al.</i> ²¹	2026	Greece	Prospective	ED triage	530	Suspected infection	In-hospital mortality	qSOFA: ≥ 2 NEWS: ≥ 5	8.70%	Yes

Abbreviations: ED: Emergency department; NEWS: National Early Warning Score; qSOFA: quick Sequential Organ Failure Assessment.

qSOFA demonstrated a significant association with in-hospital mortality (RR = 2.98, 95% CI: 1.84–4.83), with considerable heterogeneity ($I^2 = 92\%$, $p < 0.001$) (Figure 2B).

Leave-one-out sensitivity analysis showed that the pooled RR remained significant after exclusion of each individual study. Exclusion of Xythalis *et al.*²¹ reduced the pooled RR from 2.89 (95% CI: 1.82–4.59) to 2.36 (95% CI: 1.62–3.43) and I^2 from 87% to 82%, indicating that its influence was modest and did not fully explain the heterogeneity (Figure A1).

3.2.2. Hierarchical summary receiver operating characteristic analysis

The HSROC analysis showed that NEWS achieved a

sensitivity of 0.71 (95% CI: 0.66–0.75), accompanied by a specificity of 0.56 (95% CI: 0.45–0.66), suggesting moderate diagnostic performance. The inverse relationship between sensitivity and specificity was reflected by a moderate negative correlation ($r = -0.49$), consistent with a possible threshold effect related to differences in score cutoffs (Figure 3A). Overall, NEWS favored detecting mortality at the expense of specificity.

For qSOFA, pooled sensitivity was 0.44 (95% CI: 0.31–0.59), while specificity was 0.83 (95% CI: 0.73–0.90). Thus, qSOFA was less sensitive but more specific, indicating greater ability to correctly classify patients at high risk while potentially missing some mortality events. A similar inverse relationship between sensitivity and specificity was observed ($r = -0.31$), suggesting a threshold effect (Figure

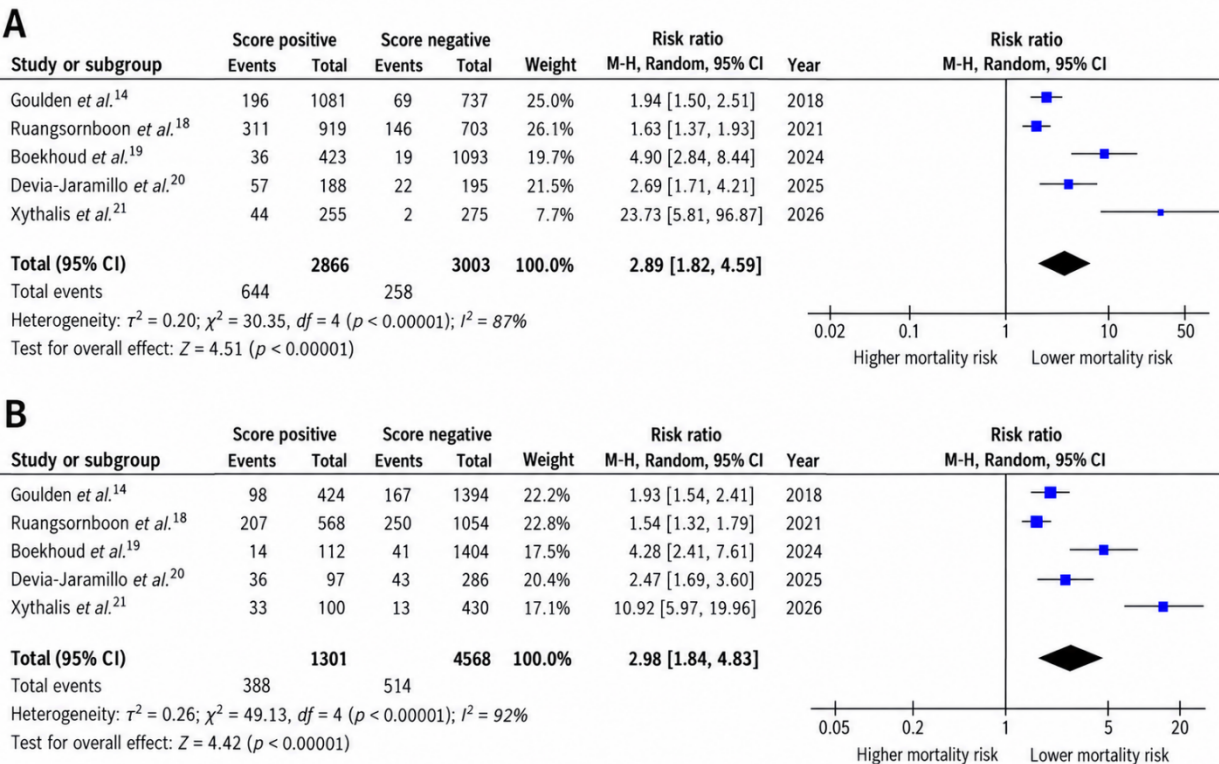


Figure 2. NEWS and qSOFA Forest plot analysis showing in-hospital mortality. (A) Forest plot illustrating the pooled risk estimates for NEWS in relation to in-hospital mortality. (B) Forest plot illustrating the pooled risk estimates for qSOFA in relation to in-hospital mortality. Five studies contributed to the RR analysis because threshold-specific data sufficient to construct 2 × 2 tables were available. Abbreviations: CI: Confidence interval; M-H: Mantel–Haenszel; NEWS: National Early Warning Score; qSOFA: quick Sequential Organ Failure Assessment; RR: Risk ratio.

3B).

3.2.3. Diagnostic odds ratio analysis

The pooled DOR for NEWS was 3.83 (95% CI: 1.86–7.90), indicating moderate overall discriminatory performance. Substantial heterogeneity was observed ($I^2 = 92.99\%$, $p < 0.001$) (Figure 4A). For qSOFA, the pooled DOR was 3.82 (95% CI: 1.85–7.88), also indicating moderate performance with similarly high heterogeneity ($I^2 = 94.09\%$, $p < 0.001$) (Figure 4B). Notably, the nearly identical DOR estimates for NEWS and qSOFA indicate comparable overall prognostic performance despite differences in sensitivity and specificity.

3.2.4. Post-test probability (Fagan analysis)

With the pre-test probability of mortality set at 15%, positive NEWS increased the estimated probability of mortality to 22%, whereas negative NEWS reduced it to 8%. For qSOFA, the probability increased to 31% following a positive result and decreased to 11% following a negative result. Overall, NEWS produced a greater reduction in

estimated mortality risk when the test result was negative, while qSOFA produced a larger increase in estimated risk when the result was positive. This pattern is consistent with the differing sensitivity and specificity profiles of the two scores.

3.2.5. Subgroup analysis

Subgroup analysis demonstrated moderate discriminatory performance for both scoring systems. The pooled AUROC was 0.69 (95% CI: 0.62–0.76) for NEWS/NEWS2 and 0.66 (95% CI: 0.59–0.74) for qSOFA, indicating moderate discrimination for both scoring systems. Both NEWS/NEWS2 ($I^2 = 91\%$) and qSOFA ($I^2 = 94.7\%$) showed considerable between-study heterogeneity. The difference in AUROC between the two scoring systems was not statistically significant ($p = 0.4977$), suggesting similar overall discriminatory performance despite their differing sensitivity and specificity profiles (Figure 5).

3.3.6. Exploration of heterogeneity

To investigate potential sources of heterogeneity, additional

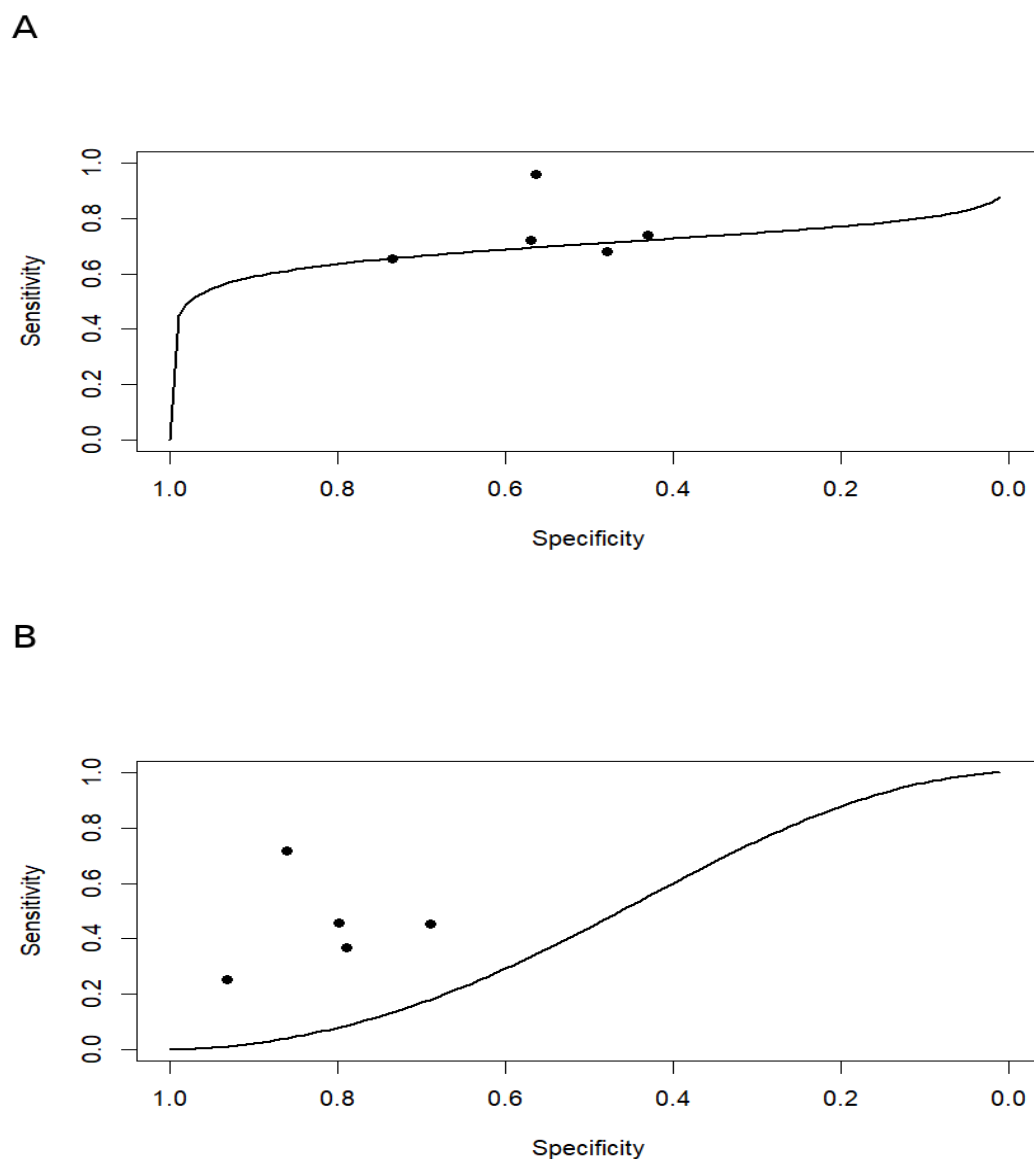


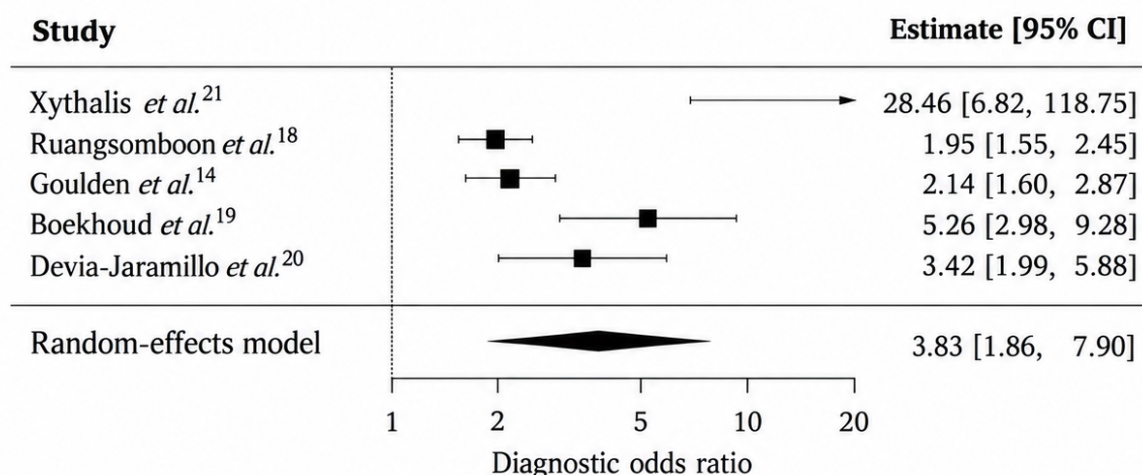
Figure 3. Hierarchical summary receiver operating characteristic predicting in-hospital mortality. (A) Hierarchical summary receiver operating characteristic (HSROC) curve showing the diagnostic performance of National Early Warning Score for in-hospital mortality in adult emergency department patients with sepsis. (B) HSROC curve showing the corresponding diagnostic performance of quick Sequential Organ Failure Assessment in the same population.

subgroup analyses were performed according to NEWS threshold and study design. For NEWS threshold, studies using a cutoff value of 5 demonstrated significantly higher pooled sensitivity than studies using higher thresholds (0.80, 95% CI: 0.71–0.87 vs. 0.68, 95% CI: 0.65–0.71) (Figure 6). The difference between subgroups was statistically significant ($Q = 5.45$, $p = 0.020$), indicating that threshold variation contributed to between-study

heterogeneity. In contrast, pooled specificity did not differ significantly between threshold groups (0.586, 95% CI: 0.494–0.673 vs. 0.574, 95% CI: 0.482–0.661) (Figure A2). No significant subgroup difference was observed ($Q = 0.04$, $p = 0.847$).

Subgroup analysis according to study design demonstrated no significant difference in pooled sensitivity between prospective and retrospective studies (0.804, 95%

A



B

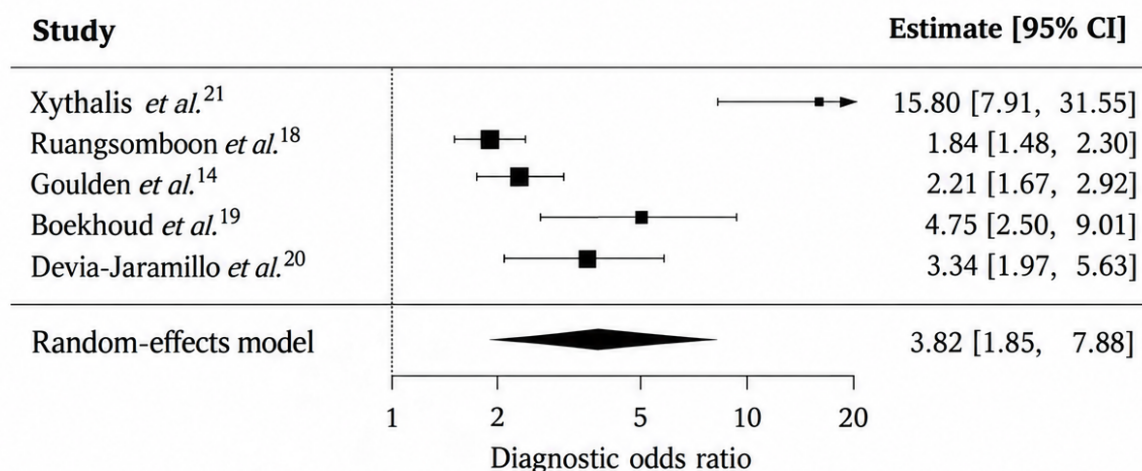


Figure 4. Pooled diagnostic odds ratios for in-hospital mortality among adult emergency department patients with sepsis. (A) Forest plot displaying the pooled DOR for NEWS in relation to in-hospital mortality among adult ED patients with sepsis. (B) Forest plot displaying the pooled DOR for qSOFA in relation to in-hospital mortality among adult ED patients with sepsis. Five studies contributed to the DOR analysis because DOR calculation required threshold-specific 2×2 data.

Abbreviations: CI: Confidence interval; DOR: Diagnostic odds ratio; ED: Emergency department; NEWS: National Early Warning Score; qSOFA: quick Sequential Organ Failure Assessment.

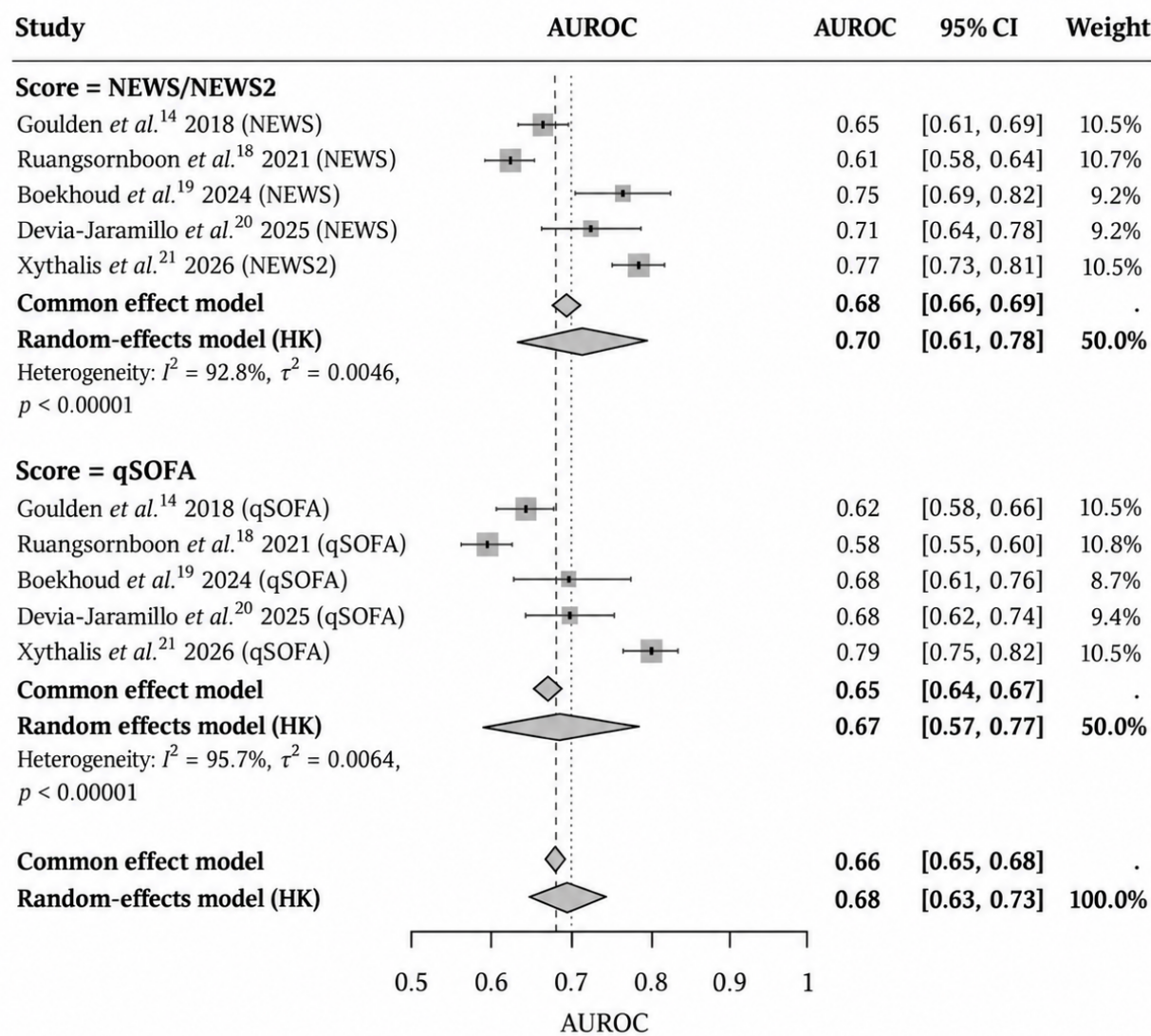
CI: 0.656–0.898 vs. 0.724, 95% CI: 0.667–0.775; $p = 0.280$) (Figure A3). However, prospective studies demonstrated significantly higher pooled specificity than retrospective studies (0.655, 95% CI: 0.592–0.713 vs. 0.522, 95% CI: 0.443–0.600; $p = 0.009$) (Figure A4).

3.3. Secondary outcome: Short-term mortality

For short-term mortality, NEWS demonstrated a strong

association with mortality (RR = 4.07, 95% CI: 3.46–4.79), with no significant heterogeneity ($I^2 = 0\%$) (Figure 7A). Similarly, qSOFA showed a strong association (RR = 4.59, 95% CI: 3.96–5.32), also with no observed heterogeneity ($I^2 = 0\%$) (Figure 7B).

Overall, the findings demonstrate comparable prognostic performance between NEWS and qSOFA, with



Heterogeneity: $I^2 = 94.1\%$, $\tau^2 = 0.0051$, $p < 0.00001$
Test for subgroup differences (common effect): $\chi^2_1 = 3.68$, $df = 1$ ($p = 0.0549$)
Test for subgroup differences (random effects): $\chi^2_1 = 0.30$, $df = 1$ ($p = 0.5816$)

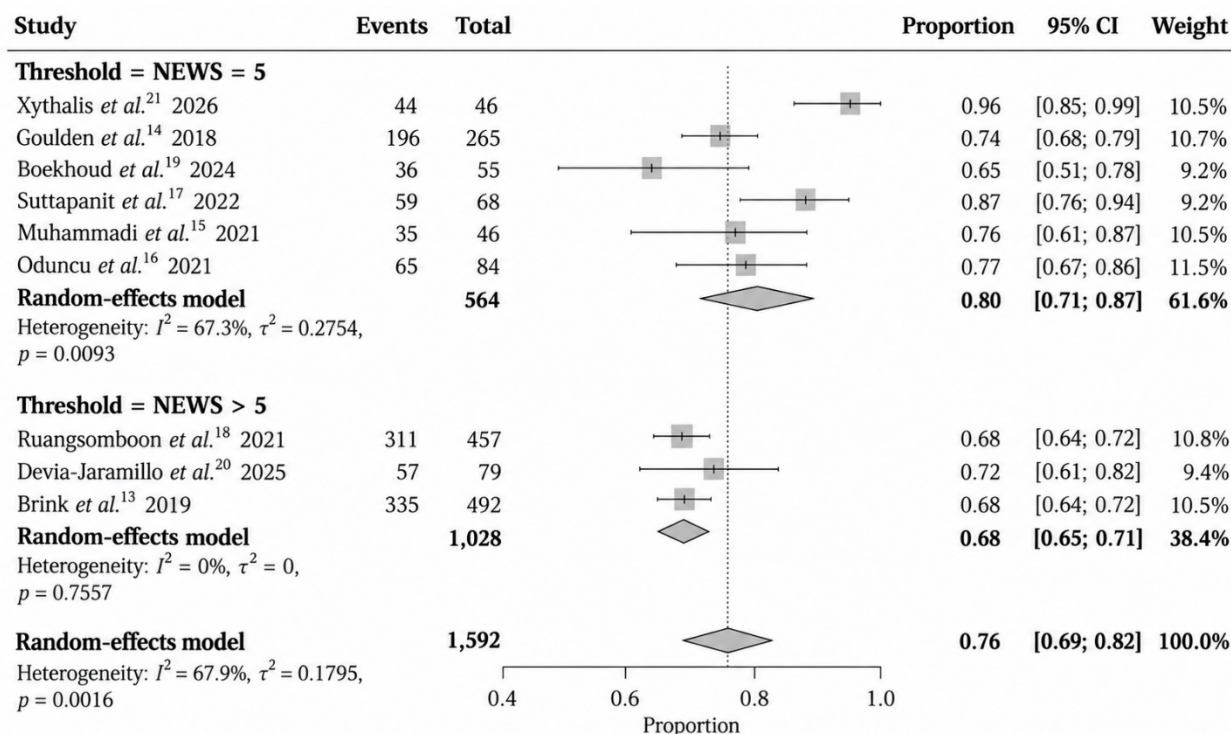
Figure 5. Forest plot comparing the AUROC of NEWS/NEWS2 and qSOFA for mortality discrimination among adult ED patients with sepsis. Six studies were included in this analysis because AUROC represents overall discriminatory ability and does not depend on a specific dichotomous threshold. Abbreviations: AUROC: Area under the receiver operating characteristic curve; CI: Confidence interval; ED: Emergency department; NEWS: National Early Warning Score; qSOFA: quick Sequential Organ Failure Assessment.

differing sensitivity and specificity profiles that may affect their clinical application.

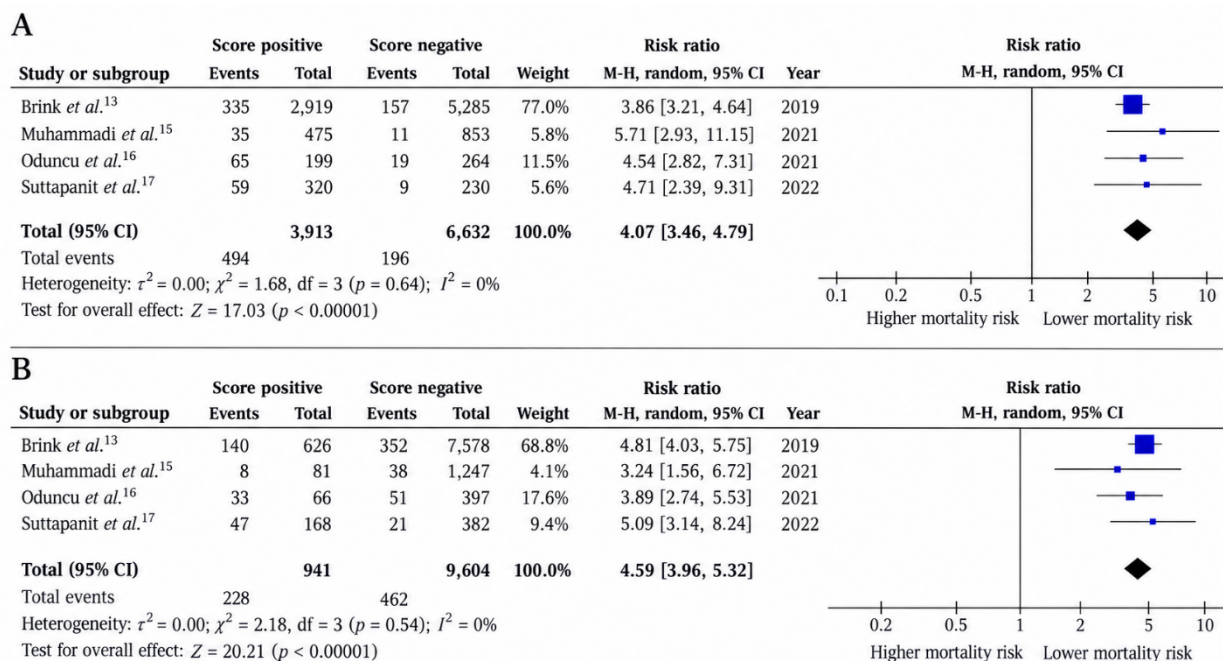
4. Discussion

This meta-analysis assessed the accuracy and usefulness of NEWS and qSOFA for predicting mortality in suspected sepsis. Both scores showed similar overall prognostic

performance based on pooled risk ratios and DORs; however, they differed in diagnostic characteristics. Specifically, NEWS had higher sensitivity than qSOFA, while qSOFA had higher specificity than NEWS. The results suggest that the two tools should not be used interchangeably but instead provide complementary information that can help facilitate timely clinical decisions.

**Figure 6.** Subgroup analysis of pooled sensitivity according to NEWS threshold

Abbreviations: CI: Confidence interval; ED: Emergency department; NEWS: National Early Warning Score.

**Figure 7.** Forest plot analysis of short-term mortality outcome. (A) Forest plot showing the association between the NEWS and short-term (28–30-day) mortality. (B) Forest plot illustrating the association between the qSOFA score and short-term (28–30-day) mortality.

Abbreviations: CI: Confidence interval; M-H: Mantel–Haenszel; NEWS: National Early Warning Score; qSOFA: quick Sequential Organ Failure Assessment.

It is speculated that higher sensitivity to NEWS would result from the combination of many physiological parameters and, in turn, would permit the diagnosis of an early physiological disturbance before the onset of overt organ dysfunction,²² making NEWS a suitable tool for screening and identifying patients most at risk of deterioration.¹³ On the other hand, qSOFA showed high specificity because it focuses on existing organ dysfunction. Thus, qSOFA may be more suitable for selecting patients at high risk given their current state and, therefore, needing intensive monitoring or ICU admission. These findings suggest that the two scores capture different stages of disease progression, supporting their complementary roles in clinical practice.

Post-test probability calculations support this conclusion. The NEWS tool was more effective at lowering the probability of mortality in cases with a negative test outcome, while the qSOFA scale had a larger impact on increasing the probability after a positive test outcome. Although the absolute difference between these two probabilities was small, combining these results can be meaningful in a sequential approach to patient management. This may help identify patients who need higher-level care without overusing critical care facilities.

Notably, these findings are particularly relevant to critical care triage and resource allocation. For many hospitals, particularly those operating under ICU capacity restrictions, it is important to identify patients who will need ICU admission efficiently. Integrating simple bedside measures such as NEWS and qSOFA into early assessment protocols would enable clinicians to prioritize patients and provide appropriate intervention and consideration for ICU admission.

These results agree with earlier studies showing higher sensitivity for NEWS and higher specificity for qSOFA in septic patients.^{23,24} However, past meta-analyses have often included a range of clinical situations, limiting their applicability to decision-making. This study advances the literature by emphasizing the complementary use of the two scores to guide escalation decisions during the critical care process.

Studies also showed noticeable discrepancies, possibly due to differences in patient characteristics, sepsis definitions, outcomes, and score thresholds. Inconsistencies in NEWS cutoffs and different uses of qSOFA led to threshold effects, reflecting the inverse relationship between sensitivity and specificity. The findings highlight the importance of interpreting scores in their clinical context.

Key methodological strengths include a geographically

broad sample population, which supports the generalizability of the conclusions. The use of bivariate random-effects modeling and HSROC analysis enables an accurate evaluation of diagnostic performance. Post-test probability analysis also makes clinical interpretation easier. Additional subgroup analyses identified threshold selection and study design as significant contributors to between-study heterogeneity. Studies using a NEWS threshold of 5 demonstrated significantly higher pooled sensitivity than studies using higher thresholds (0.80 vs. 0.68; $p = 0.020$), supporting a threshold effect. Furthermore, prospective studies demonstrated significantly higher pooled specificity than retrospective studies (0.66 vs. 0.52; $p = 0.009$), indicating that methodological differences may also contribute to variability across investigations.

The substantial heterogeneity observed across analyses may reflect clinical and methodological differences between studies, including variations in the definitions of suspected infection or sepsis, patient case-mix and representation of immunocompromised populations, mortality time horizons, score thresholds, and ED settings. These differences may limit direct comparability across studies. Furthermore, the multiple subgroup analyses involved several statistical comparisons; therefore, individual subgroup p -values should be interpreted as exploratory rather than definitive, with greater emphasis on the direction and magnitude of effects and their confidence intervals. Meta-regression was not performed because of the limited number of included studies and the risk of unstable estimates.

There are also some limitations associated with this study. First, the relatively low number of articles included in this research is notable. Heterogeneity is also rather high when analyzing the findings. The substantial heterogeneity was not attributable to Xythalis *et al.*²¹ alone, as exclusion of this study reduced I^2 only modestly from 87% to 82%; however, the reduction to 57% after excluding both Xythalis *et al.*²¹ and Boekhoud *et al.*¹⁹ suggests that study-level differences may have contributed meaningfully to the observed variability. In addition, comparing studies may be difficult because definitions of mortality and thresholds differ. Another limitation is the lack of individual patient data, which precluded more detailed patient-level subgroup analyses and formal covariate adjustment. While threshold subgroup analyses were performed, the limited number of studies using higher NEWS thresholds may have reduced the precision of subgroup estimates. Another limitation is that this analysis considered only ED settings, while ICU settings were not discussed.

Collectively, these results show the need to move away from conceptualizing NEWS and qSOFA as alternative

strategies toward a more nuanced understanding of their respective benefits within an organized risk-stratification framework. A systematic process combining both scores may help identify patients at risk of developing a critical condition, ensuring early transfer to the ICU setting. However, additional studies are needed to confirm the value of this concept.

5. Conclusion

The NEWS and qSOFA demonstrate comparable overall prognostic performance for predicting mortality among adults with suspected sepsis presenting to the ED. NEWS offers greater sensitivity and may facilitate earlier identification of patients at risk of clinical deterioration, whereas qSOFA provides greater specificity and may better identify patients at increased risk of adverse outcomes. These findings support a complementary rather than competitive role for the two tools. A sequential approach integrating NEWS and qSOFA may enhance early risk stratification, support timely escalation to critical care, and improve clinical decision-making, although prospective validation studies are required.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: Muhammad Abubakar, Ubaid ur Rehman Khizar, Abdullah Ashfaq, Misbah Shirin Memon, Maheen Zahid, Muhammad Ahmad, AAw

Formal analysis: Muhammad Abubakar, Ubaid ur Rehman Khizar, Wagma Khan, Muhammad Zeeshan, Misbah Shirin Memon, Abeer Murad Khamisani

Methodology: Ubaid ur Rehman Khizar, Wagma Khan, Muhammad Zeeshan, Abdullah Ashfaq, Misbah Shirin Memon, Maheen Zahid, Abeer Murad Khamisani

Supervision: Ayoola Awosika, Muhammad Ahmad

Writing–original draft: All authors

Writing–review & editing: Muhammad Abubakar, AS, Muhammad Zeeshan, Abdullah Ashfaq, Muhammad Ahmad, Ayoola Awosika

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

All data generated or analyzed during this study are included in this published article.

References

1. GBD 2021 Global Sepsis Collaborators. Global, regional, and national sepsis incidence and mortality, 1990–2021: a systematic analysis. *Lancet Glob Health*. 2025;13(12):e2013–e2026.
doi: 10.1016/S2214-109X(25)00356-0
2. Machado FR, Cavalcanti AB, Braga MA, *et al*. Sepsis in Brazilian emergency departments: a prospective multicenter observational study. *Intern Emerg Med*. 2023;18(2):409–421.
doi: 10.1007/s11739-022-03179-3
3. Özkaya S, Yurdakul MS. Impact of emergency department-to-intensive care unit transfer time on in-hospital mortality: A retrospective cohort study. *Medicine (Baltimore)*. 2025;104(51):e46724.
doi: 10.1097/MD.00000000000046724
4. Medeiros RP, Teixeira C. Early ICU mortality: An opportunity for proactive triage, ethical deliberation, and integrated models of care. *J Clin Anesth Intensive Care*. 2025;5(1):7–9.
doi: 10.46439/anesthesia.5.024
5. Gerry S, Bonnici T, Birks J, *et al*. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. *BMJ*. 2020;369:m1501.
doi: 10.1136/bmj.m1501
6. Wanlumkhao W, Rattanamongkolgul D, Ekpanyaskul C. Performance of Early Sepsis Screening Tools for Timely Diagnosis and Antibiotic Stewardship in a Resource-Limited Thai Community Hospital. *Antibiotics (Basel)*. 2025;14(7):708.
doi: 10.3390/antibiotics14070708
7. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the Emergency Department. *Am J Emerg Med*. 2019;37(8):1490–1497.
doi: 10.1016/j.ajem.2018.10.058
8. Sivayoham N, Hussain AN, Shabbo L, Christie D. An observational cohort study of the performance of the REDS score compared to the SIRS criteria, NEWS2, CURB65, SOFA, MEDS and PIRO scores to risk-stratify emergency department suspected sepsis. *Ann Med*. 2021;53(1):1863–1874.

- doi: 10.1080/07853890.2021.1992495
9. Lam RPK, Dai Z, Lau EHY, *et al.* Comparing 11 early warning scores and three shock indices in early sepsis prediction in the emergency department. *World J Emerg Med.* 2024;15(4):273-282.
doi: 10.5847/wjem.j.1920-8642.2024.052
 10. Guan G, Lee CMY, Begg S, Crombie A, Mnatzaganian G, Kuo YH. The use of early warning system scores in prehospital and emergency department settings to predict clinical deterioration: A systematic review and meta-analysis. *PLoS One.* 2022;17(3):e0265559.
doi: 10.1371/journal.pone.0265559
 11. Chua WL, Rusli KDB, Aitken LM. Early warning scores for sepsis identification and prediction of in-hospital mortality in adults with sepsis: A systematic review and meta-analysis. *J Clin Nurs.* 2024;33(6):2005-2018.
doi: 10.1111/jocn.17061
 12. Daorattanachai K, Maithong S, Phungoen P, Weschawalit S, Srivilaithon W. Respiratory National Early Warning Score for 28-day mortality prediction in suspected sepsis patients in the emergency department. *BMC Emerg Med.* 2025;26(1):17.
doi: 10.1186/s12873-025-01443-1
 13. Brink A, Alsma J, Verdonschot RJCG, *et al.* Predicting mortality in patients with suspected sepsis at the Emergency Department; A retrospective cohort study comparing qSOFA, SIRS and National Early Warning Score. *PLoS One.* 2019;14(1):e0211133.
doi: 10.1371/journal.pone.0211133
 14. Goulden R, Hoyle MC, Monis J, *et al.* qSOFA, SIRS and NEWS for predicting in-hospital mortality and ICU admission in emergency admissions treated as sepsis. *Emerg Med J.* 2018;35(6):345-349.
doi: 10.1136/emmermed-2017-207120
 15. Mohammadi P, Dekker R, Brown V, *et al.* A prospective, observational study of the performance of MEWS, NEWS, SIRS and qSOFA for early risk stratification for adverse outcomes in patients with suspected infections at the emergency department. *Acute Med.* 2021;20(2):116-124.
doi: 10.52964/amja.0851
 16. Oduncu AF, Kıyan GS, Yalçınlı S. Comparison of qSOFA, SIRS, and NEWS scoring systems for diagnosis, mortality, and morbidity of sepsis in emergency department. *Am J Emerg Med.* 2021;48:54-59.
doi: 10.1016/j.ajem.2021.04.006
 17. Suttapanit K, Satirachoenkul S, Sanguanwit P, Prachanukool T. The Accuracy of Sepsis Screening Score for Mortality Prediction at Emergency Department Triage. *West J Emerg Med.* 2022;23(5):698-705.
doi: 10.5811/westjem.2022.6.56754
 18. Ruangsomboon O, Boonmee P, Limsuwat C, Chakorn T, Monsomboon A. The utility of the rapid emergency medicine score (REMS) compared with SIRS, qSOFA and NEWS for Predicting in-hospital Mortality among Patients with suspicion of Sepsis in an emergency department. *BMC Emerg Med.* 2021;21(1):2.
doi: 10.1186/s12873-020-00396-x
 19. Boekhoud L, Schaap HMEA, Huizinga RL, *et al.* Predictive performance of NEWS and qSOFA in immunocompromised sepsis patients at the emergency department. *Infection.* 2024;52(5):1863-1873.
doi: 10.1007/s15010-024-02247-4
 20. Devia-Jaramillo GA, Erazo-Guerrero L, Laguado-Castro V, Alfonso-Parada JM. Evaluating Sepsis Mortality Predictions from the Emergency Department: A Retrospective Cohort Study Comparing qSOFA, the National Early Warning Score, and the International Early Warning Score. *J Clin Med.* 2025;14(14):4869.
doi: 10.3390/jcm14144869
 21. Xythalis D, Kalafati M, Mpouzika M, *et al.* A Real-Time Prospective Evaluation of the Prognostic Accuracy of SIRS, MEWS, NEWS2 and qSOFA in Predicting ICU Admission and Mortality in an Emergency Department: Implications for Nursing Practice. *Nurs Crit Care.* 2026;31(1):e70302.
doi: 10.1111/nicc.70302
 22. Marosi V, Gadda G, Timoftica C, Alberti A, Destrebecq A. National Early Warning Score 2 to Identify Sepsis in the Emergency Department. A Review of the Literature. *J Med Clin Nurs Stud.* 2023:1-5.
doi: 10.61440/jmcns.2023.v1.15
 23. Durr D, Niemi T, Despraz J, *et al.* National Early Warning Score (NEWS) Outperforms Quick Sepsis-Related Organ Failure (qSOFA) Score for Early Detection of Sepsis in the Emergency Department. *Antibiotics (Basel).* 2022;11(11):1518.
doi: 10.3390/antibiotics11111518
 24. Wang C, Xu R, Zeng Y, Zhao Y, Hu X. A comparison of qSOFA, SIRS and NEWS in predicting the accuracy of mortality in patients with suspected sepsis: A meta-analysis. *PLoS One.* 2022;17(4):e0266755.
doi: 10.1371/journal.pone.0266755

Appendix

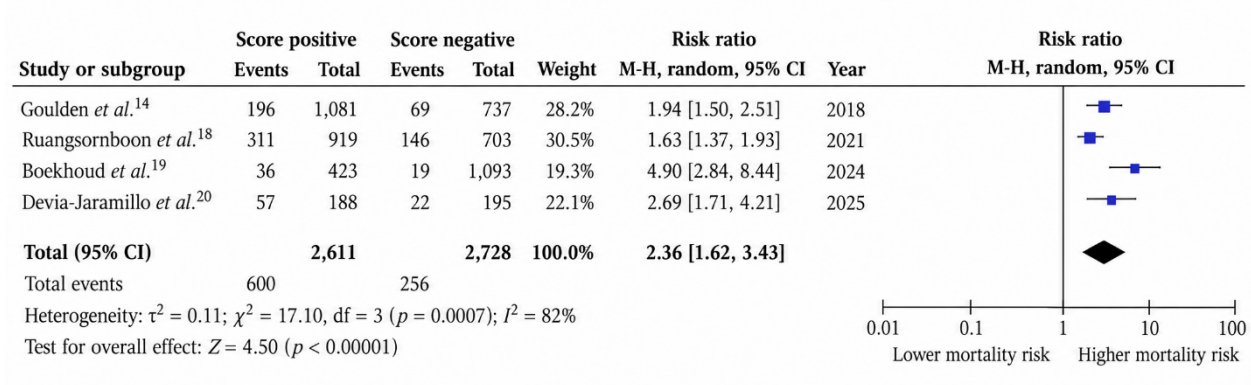


Figure A1. Leave-one-out analysis
Abbreviations: CI: Confidence interval; M-H: Mantel-Haenszel.

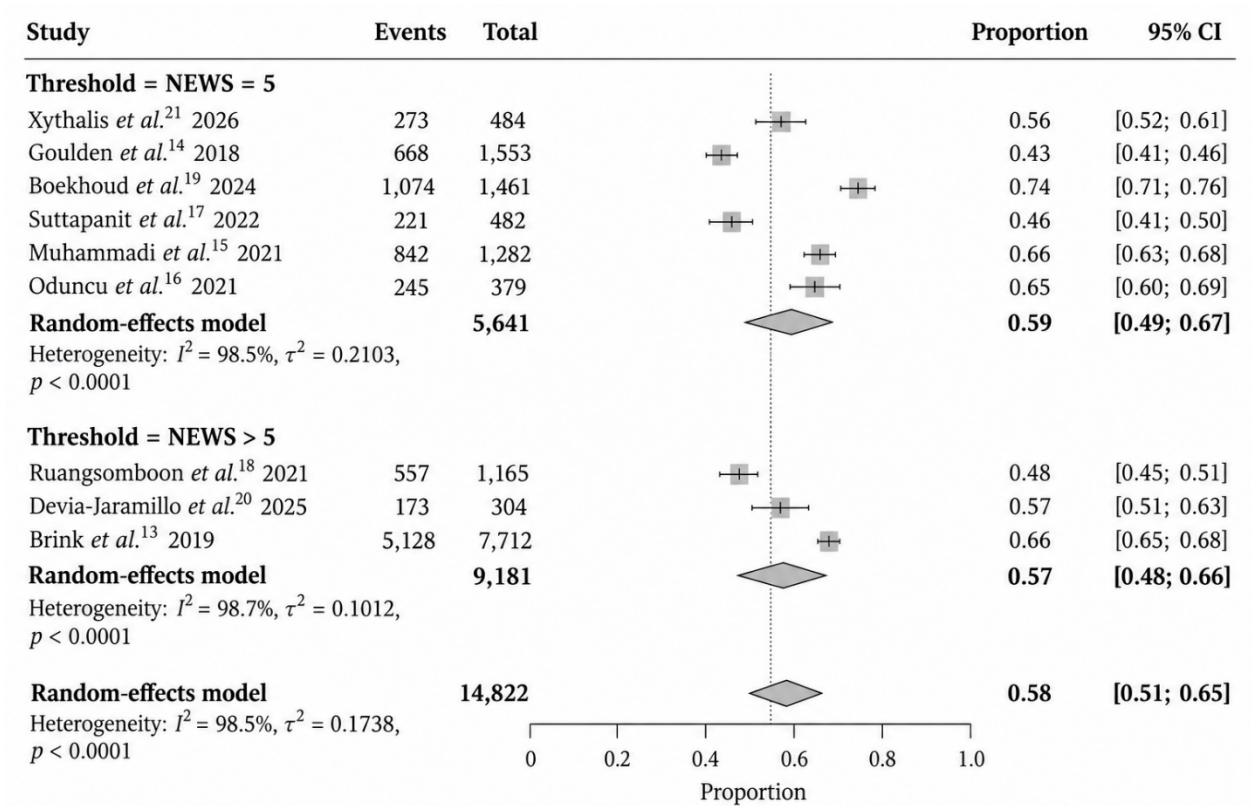


Figure A2. Threshold subgroup analysis of specificity
Abbreviations: CI: Confidence interval; NEWS: National Early Warning Score.

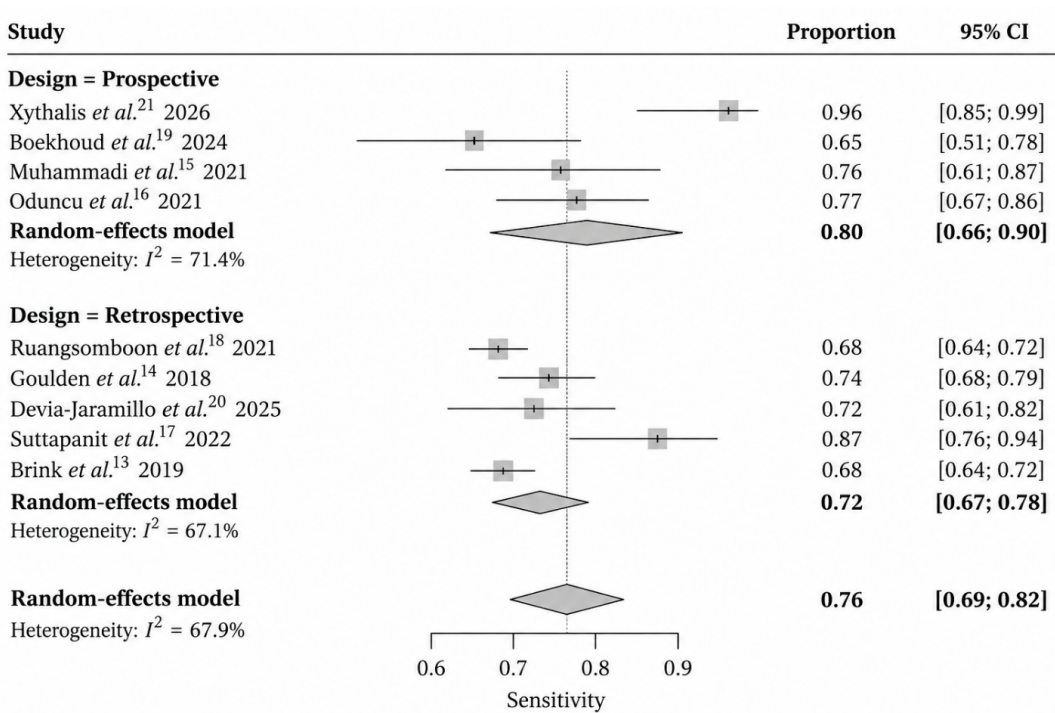


Figure A3. Subgroup analysis of pooled sensitivity according to study design. No significant difference in pooled sensitivity was observed between prospective and retrospective studies (0.804 vs. 0.724; $p = 0.280$).
Abbreviation: CI: Confidence interval.

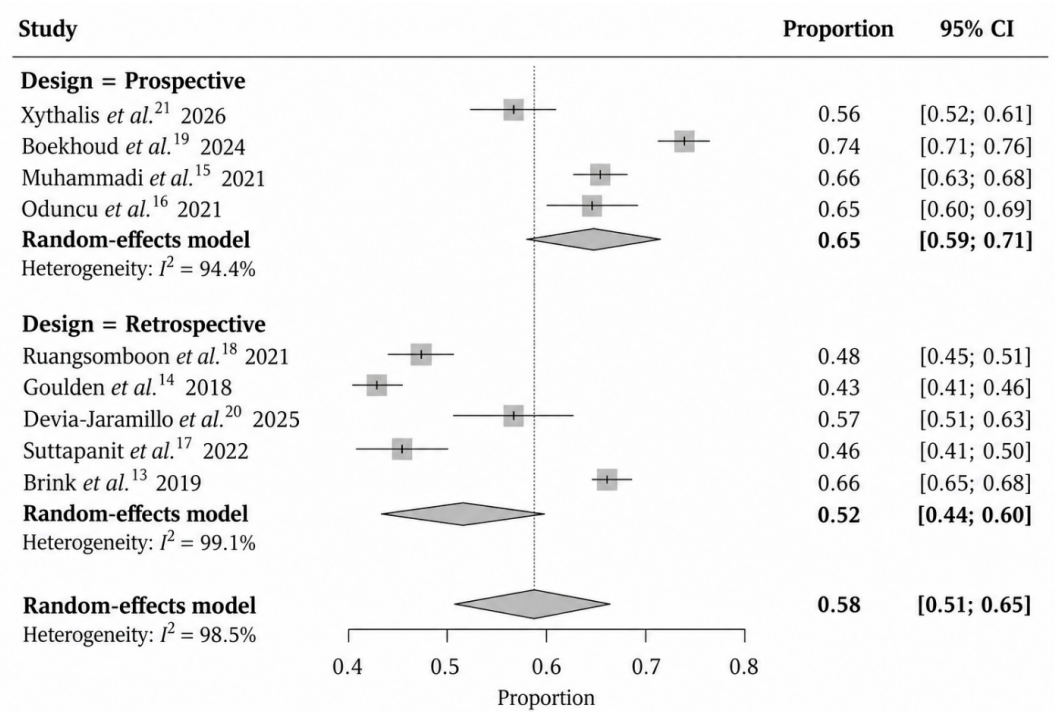


Figure A4. Subgroup analysis of pooled specificity according to study design. Prospective studies demonstrated significantly higher pooled specificity than retrospective studies (0.655 vs. 0.522; $p = 0.009$), suggesting that methodological differences contributed to between-study heterogeneity.
Abbreviation: CI: Confidence interval.