

RESEARCH

Open Access



External validation of the International Early Warning Score in non-traumatic emergency department patients: a prospective cohort study

Fatma Bayram¹, Buğra İlhan^{2*}, Zeynep Kan³, Oğuz Eroğlu² and Turgut Deniz²

Abstract

Background Emergency department (ED) overcrowding has become a global public health concern, underscoring the importance of rapid and reliable risk stratification tools. Early warning scores are widely used to identify patients at risk of deterioration and mortality. The recently developed International Early Warning Score (IEWS), which incorporates age and sex adjustments into the National Early Warning Score (NEWS) model, has shown promising results and has undergone initial external validation in a Danish cohort; however, no prospective external validation has yet been conducted, and broader international validation remains limited. This study aimed to evaluate the performance of IEWS compared with NEWS in predicting in-hospital mortality, 30-day mortality, and ICU admission among adult ED patients.

Methods This prospective observational cohort study was conducted between July and August 2024 in a tertiary university hospital ED with an annual census of ~70,000 visits. Adult patients presenting to the ED were included, while trauma cases, patients without vital signs on arrival, interhospital transfers, and cases with incomplete data were excluded. IEWS and NEWS were calculated at presentation. The primary outcome was all-cause in-hospital mortality; secondary outcomes included 30-day mortality and ICU admission.

Results A total of 8,666 patients were analyzed. The median age was 40 years (IQR: 26–58), and 51.5% were female. In-hospital mortality was 1.5% ($n = 134$), and 30-day mortality was 1.9% ($n = 163$). IEWS demonstrated excellent discriminative ability for in-hospital and 30-day mortality (AUC: 0.944 and 0.930, respectively), and good performance for ICU admission (AUC: 0.876). In contrast, NEWS showed good performance for in-hospital and 30-day mortality (AUC: 0.884 and 0.848, respectively) and moderate performance for ICU admission (AUC: 0.781). IEWS consistently outperformed NEWS across all outcomes ($p < 0.05$, DeLong's test).

Conclusion IEWS outperformed NEWS in predicting in-hospital mortality, 30-day mortality, and ICU admission among non-traumatic ED patients. Given its high sensitivity, specificity, and overall discriminative performance, IEWS

*Correspondence:
Buğra İlhan
bugra_ilhan@yahoo.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **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/>.

may serve as a reliable bedside tool for patient risk stratification in the ED. Large-scale multicenter studies are needed to confirm its generalizability across diverse populations.

Clinical trial number Not applicable.

Keywords Early warning score, International Early Warning Score, National Early Warning Score, Mortality prediction, Intensive care admission, Emergency department

Introduction

Emergency department (ED) utilization has risen markedly in recent years, transforming ED overcrowding into a pressing public health issue [1, 2]. Overcrowding contributes to adverse consequences, including prolonged waiting times, an increased risk of medical errors and adverse events, and reduced patient satisfaction [1, 3–5]. Evidence from multiple studies has linked extended ED length of stay to unfavorable outcomes across diverse patient groups [6–8]. Furthermore, timely intervention is crucial in life-threatening conditions, including sepsis, myocardial infarction, stroke, and cardiac arrest [9–11]. Collectively, these observations underscore the crucial importance of early risk stratification in distinguishing between high- and low-acuity patients, thereby enabling the rapid allocation of care to the appropriate level or facilitating timely discharge.

Especially in highly congested EDs, predicting which patients are at risk of deterioration represents a major challenge. This challenge is magnified during crises such as disasters and pandemics, where excessive demand and resource scarcity necessitate accurate discrimination between critical and non-critical patients. Early warning scores (EWS) represent “track-and-trigger” tools specifically developed to identify patients at risk of clinical deterioration outside intensive care units (ICUs), enabling early stabilization, timely escalation of care, and prevention of avoidable cardiac arrests [12]. By design, EWS provide a structured risk estimate to aid clinical decision-making. To ensure broad utility, an ideal EWS must rely on variables that are readily accessible, easily measurable, rapidly calculable, and strongly associated with patient outcomes.

Since the first proposal of EWS in the 1990s, they have been widely recognized as valuable tools for anticipating clinical deterioration [12, 13]. The development of the National Early Warning Score (NEWS) by the Royal College of Physicians in 2012 marked a significant step forward, as it established a standardized framework for risk assessment and was formally recommended for routine clinical use [12]. The adoption of EWS has subsequently been linked to a measurable reduction in adverse clinical outcomes [14]. Over the following decades, numerous validation studies have examined their performance across diverse patient populations and healthcare

settings, further supporting their role in modern emergency and acute care practice [6, 15–22].

Although widely implemented, EWS have shown inconsistent predictive performance across different patient populations and age groups [15, 18, 22]. Recognizing that the NEWS tends to underestimate risk in older adults while overestimating it in younger patients, Candel et al. proposed the International Early Warning Score (IEWs) in 2023, incorporating age- and sex-adjusted parameters [23]. Initial derivation and validation studies suggested that IEWS outperformed NEWS; however, subsequent external validation studies reported only moderate discriminative ability for mortality prediction [16, 23]. Despite these evaluations, no prospective external validation of the IEWS has yet been conducted. Given the limited evidence currently available, further external validation of IEWS is essential.

Accordingly, the present study aims to externally validate the IEWS and compare its performance with NEWS in predicting in-hospital mortality, 30-day mortality, and ICU admission among adult ED patients.

Methods

Study design

We conducted a prospective, observational cohort study from July 1 to August 31, 2024, in the ED of a tertiary university hospital with an annual census of approximately 70,000 visits. Serving as the city’s sole tertiary care facility, the hospital is also designated as a primary percutaneous coronary intervention center, a stroke center, and a level 1 trauma center. At the time of the study, no EWS, including NEWS or IEWS, was routinely used in the emergency department. This study was conducted to identify the most appropriate EWS for our clinical workflow, and the implementation of the most suitable score is planned based on the study findings. The study protocol received approval from the institutional ethics committee (Approval ID: 2024.04.11; Date: 17.04.2024) and was performed in compliance with the Declaration of Helsinki.

Patient selection and groups

All patients aged ≥ 18 years who presented to the ED between the study dates and provided informed consent were eligible for inclusion. Exclusion criteria included patients who arrived in cardiac arrest (no measurable vital signs), trauma-related presentations, incomplete

clinical data, and interfacility transfers. For patients who were unable to provide informed consent (e.g., due to altered mental status), consent was obtained from legally authorized representatives or accompanying relatives.

Trauma patients were excluded because they differ markedly from medical ED patients in terms of presentation mechanisms, clinical management processes, and expected outcomes. As these cases follow trauma-specific evaluation pathways rather than medical deterioration scoring, their inclusion would introduce heterogeneity inconsistent with the primary aim of this study.

Data on demographics, comorbid conditions, presenting vital parameters, level of consciousness, and clinical outcomes were collected from both electronic and paper-based records and entered prospectively into standardized case report forms by the research team. IEWS and NEWS were subsequently calculated by researchers blinded to patient outcomes, using their original scoring algorithms without modification.

Patients with incomplete data for IEWS/NEWS calculation were excluded during eligibility screening; thus, the final analytic cohort contained no missing values for required variables, and complete-case analysis was performed.

For outcome analysis, patients were stratified according to in-hospital mortality, 30-day mortality, and ICU admission.

Thirty-day mortality was determined using electronic hospital records for in-hospital deaths and structured telephone follow-up for patients discharged before day 30.

ICU admission decisions follow standardized institutional criteria, including hemodynamic instability, need for vasopressor or mechanical ventilatory support, or conditions requiring continuous monitoring beyond ward capability. Throughout the study period, ICU bed availability remained stable, and no capacity-related admission restrictions were reported.

Scoring systems

The National Early Warning Score (NEWS), introduced in 2012 by the Royal College of Physicians, was designed to facilitate the early detection of clinical deterioration [12]. It encompasses seven physiological parameters—respiratory rate, oxygen saturation, supplemental oxygen requirement, temperature, systolic blood pressure, heart rate, and level of consciousness—yielding a total score of 0–20.

In 2023, Candel et al. developed the International Early Warning Score (IEWs), an age- and sex-adjusted modification of NEWS [23]. By incorporating age and sex into the original framework, the IEWS consists of nine variables and generates a score between 0 and 29.

In this study, both scoring systems were applied exactly as defined in their original derivation studies. All parameters were recorded at the time of ED presentation. To avoid bias, score calculations were performed prospectively by an independent investigator blinded to clinical outcomes.

Outcomes

The primary endpoint of this study was all-cause in-hospital mortality. Secondary endpoints were defined as all-cause 30-day mortality and ICU admission.

Statistical analysis

Sample size estimation was informed by the study of Candel et al. ($n = 95,553$; in-hospital mortality: 2.4%), resulting in a calculated requirement of 5,465 patients (expected mortality: 3%, power: 80%, $\alpha = 0.05$) [23]. Calculations were performed using an online sample size calculator (www.clinicalcalc.com).

Continuous variables were summarized as mean $\pm$ standard deviation or median with interquartile range (IQR), while categorical variables were expressed as frequencies and percentages. Normality of continuous variables was evaluated with the Kolmogorov–Smirnov test, and subsequent statistical tests were selected based on these distributional assessments. Group comparisons were conducted with the Mann–Whitney U test for continuous data and the chi-square test for categorical data. The discriminative ability of each scoring system was assessed by calculating the area under the receiver operating characteristic curve (AUC). Optimal cut-off points were identified using the Youden index, and performance was expressed as AUC values with 95% confidence intervals (CIs), along with sensitivity, specificity, and likelihood ratios. Statistical analyses were conducted using SPSS® for Windows, version 23.0 (IBM, Chicago, IL, USA). Statistical significance was set at a two-sided p -value < 0.05 . Calibration for both IEWS and NEWS was assessed for all outcomes (in-hospital mortality, 30-day mortality, and ICU admission). Calibration was evaluated visually using calibration plots, in which observed event rates were plotted against predicted probabilities across deciles of risk. The 45-degree diagonal line represented perfect calibration. Deviations above the line indicated underestimation, whereas deviations below the line reflected overestimation of risk. Brier scores were calculated to quantify overall calibration performance. Decision curve analysis (DCA) was performed to evaluate the clinical utility of IEWS and NEWS by quantifying their standardized net benefit across a range of threshold probabilities (0.01–0.35). Net benefit was calculated using the standard formula proposed by Vickers et al., and decision curves for each outcome (in-hospital mortality, 30-day mortality, ICU admission) were plotted against “treat

all” and “treat none” strategies [24]. Higher net benefit indicated greater clinical usefulness at a given threshold probability. Calibration plots and decision curve analyses were performed using Python (version 3.11; Python Software Foundation, Wilmington, DE, USA).

Results

Of the 10,493 patients assessed for eligibility, 1,785 were excluded due to trauma, 21 owing to incomplete data, 11 due to inter-hospital transfer, and 10 because of absent vital signs on arrival. Thus, the final analysis comprised 8,666 patients. The study flowchart is illustrated in Fig. 1.

The cohort had a median age of 40 years (IQR: 26–58), with 4,462 (51.5%) female participants. Overall, 506 patients (5.7%) were admitted to hospital wards, and 243 (2.8%) were admitted to the ICU. The remaining patients were discharged. In-hospital mortality was 1.5% ($n=134$), while 30-day mortality was 1.9% ($n=163$). Baseline demographic and clinical characteristics are presented in Table 1.

The IEWS outperformed the NEWS across all outcomes. For in-hospital and 30-day mortality, IEWS achieved excellent discriminative ability (AUC: 0.944, 95% CI: 0.939–0.948; and AUC: 0.930, 95% CI: 0.924–0.935, respectively), whereas NEWS yielded only good performance (AUC: 0.884, 95% CI: 0.877–0.891; and AUC: 0.848, 95% CI: 0.840–0.855, respectively). Similarly, in predicting ICU admission, IEWS demonstrated good accuracy (AUC: 0.876, 95% CI: 0.869–0.883), while NEWS performed at a moderate level (AUC: 0.781, 95% CI: 0.772–0.789).

Using the Youden index to determine optimal cut-off values (in-hospital mortality: IEWS >5 ; 30-day mortality and ICU admission: IEWS >4), the IEWS achieved sensitivities of 88%, 92%, and 81% and specificities of 87%, 80%, and 81% for in-hospital mortality, 30-day mortality, and ICU admission, respectively. In contrast, for NEWS, the optimal cut-offs (in-hospital and 30-day mortality: NEWS >1 ; ICU admission: NEWS >2) yielded sensitivities of 86.6%, 82.8%, and 58.0% with specificities of 76.7%, 76.8%, and 87.8%, respectively. Comprehensive

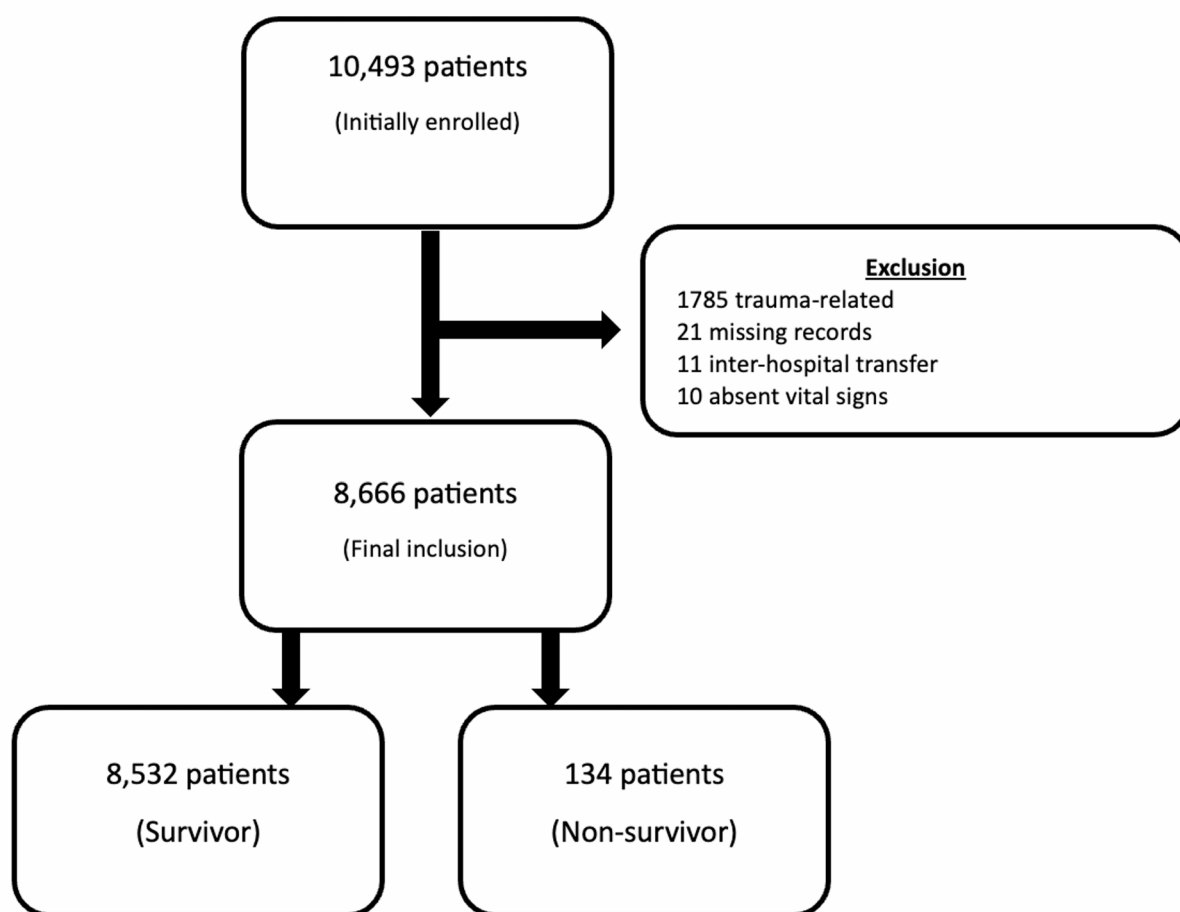


Fig. 1 Flow diagram of the study

Table 1 Baseline demographic and clinical characteristics of the patients

Variables	Total (n = 8666)	Survivor (n = 8532)	Non-survivor (n = 134)	p
Gender, female*	4462 (51.5)	4414 (51.7)	48 (35.8)	< 0.001
Arrival, ambulance*	1167 (13.5)	1137 (13.3)	30 (22.4)	0.002
Triage*				< 0.001
Red	252 (2.9)	196 (2.3)	56 (41.8)	
Yellow	5160 (59.4)	5072 (59.4)	78 (58.2)	
Green	3264 (37.7)	3264 (38.3)	0 (0.0)	
Comorbid diseases*				
COPD	400 (4.6)	371 (4.3)	29 (21.6)	< 0.001
Asthma	39 (0.5)	39 (0.5)	0 (0.0)	0.433
Diabetes mellitus	1014 (11.7)	985 (11.5)	29 (21.6)	< 0.001
Hypertension	1530 (17.7)	1473 (17.3)	57 (42.5)	< 0.001
Coronary artery disease	935 (10.8)	878 (10.3)	57 (42.5)	< 0.001
Chronic heart failure	239 (2.8)	217 (2.5)	22 (16.4)	< 0.001
Chronic kidney disease	80 (0.9)	74 (0.9)	6 (4.5)	< 0.001
Active malignancy	239 (2.8)	209 (2.4)	30 (22.4)	< 0.001
Chronic liver disease	18 (0.2)	15 (0.2)	3 (2.2)	< 0.001
Hyperthyroidism	13 (0.2)	13 (0.2)	0 (0.0)	0.651
Hypothyroidism	188 (2.2)	185 (2.2)	3 (2.2)	0.956
Benign prostate hyperplasia	92 (1.1)	86 (1.0)	6 (4.5)	< 0.001
Alzheimer disease	28 (0.3)	22 (0.3)	6 (4.5)	< 0.001
Parkinson disease	13 (0.2)	13 (0.2)	0 (0.0)	0.651
Epilepsy	30 (0.3)	29 (0.3)	1 (0.7)	0.427
Cerebrovascular disease	121 (1.4)	113 (1.3)	8 (6.0)	< 0.001
Other	276 (3.4)	274 (3.0)	2 (1.4)	< 0.001
Age**	40 (26–58)	40 (26–57)	74 (64–84)	< 0.001
Systolic blood pressure (mmHg)**	125 (118–134)	125 (118–134)	118 (103–137)	< 0.001
Diastolic blood pressure (mmHg)**	78 (71–84)	78 (71–84)	74 (59–87)	0.005
Heart rate (bpm)**	86 (78–94)	86 (78–93)	97 (81–107)	< 0.001
Respiratory rate (/min)**	16 (14–16)	16 (14–16)	16 (14–18)	< 0.001
Oxygen saturation (%)**	98 (96–98)	98 (96–98)	92 (86–96)	< 0.001
Fever (° C)**	36.6 (36.5–36.7)	36.6 (36.5–36.7)	36.6 (36.5–36.8)	0.001
Glasgow coma scale**	15 (15–15)	15 (15–15)	15 (14–15)	< 0.001
Emergency length of stay (min)**	50 (14–144)	48 (13–140)	292 (212–408)	< 0.001
IEWs**	2 (1–4)	2 (1–4)	10 (7–12)	< 0.001
NEWS**	0 (0–1)	0 (0–1)	6 (2.7–8)	< 0.001
Disposition*				< 0.001
Discharge	7643 (88.4)	7643 (89.6)	0 (0.0)	
Ward admission	506 (5.7)	459 (5.4)	47 (35.1)	
ICU admission	243 (2.8)	156 (1.8)	87 (64.9)	
Discharge against medical advice	274 (3.2)	274 (3.2)	0 (0.0)	
In-hospital mortality*	134 (1.5)			
30-day mortality*	163 (1.9)			

COPD: Chronic obstructive pulmonary disease, IEWS: International Early Warning Score, NEWS: National Early Warning Score, ICU: Intensive care unit, *n (%) - Chi-square test, **median (Q1-Q3) - Mann-Whitney U test

performance results are detailed in Table 2. Comparative analysis using the DeLong test demonstrated statistically significant differences between the IEWS and NEWS AUCs for all outcomes ($p < 0.05$ for in-hospital mortality, 30-day mortality, and ICU admission). Figure 2 provides the corresponding ROC curves for visual comparison.

For in-hospital mortality, IEWS demonstrated superior calibration compared with NEWS (Fig. 3). IEWS showed

closer alignment with the ideal 45-degree line across all deciles, particularly in mid- and high-risk strata. NEWS exhibited greater fluctuation and underestimation in intermediate predicted-risk ranges. Both scores had identical Brier scores (0.013), although the graphical calibration favored IEWS. For 30-day mortality, IEWS again showed better calibration, with smaller deviations from perfect calibration compared with NEWS. NEWS

Table 2 Performance of the scores in predicting outcomes

	Cut-off	AUC	95% CI	Sensitivity	Specificity	+LR	-LR	PPV	NPV
In-hospital mortality									
IEWS	> 5	0.944	0.939–0.948	88.81	87.10	6.88	0.13	9.7	99.8
NEWS	> 1	0.884	0.877–0.891	86.57	76.66	3.71	0.18	5.1	99.7
30-day mortality									
IEWS	> 4	0.930	0.924–0.935	92.02	80.84	4.80	0.09	8.4	99.8
NEWS	> 1	0.848	0.840–0.855	82.82	76.81	3.57	0.22	6.3	99.5
ICU admission									
IEWS	> 4	0.876	0.869–0.883	81.07	81.22	4.32	0.23	10.9	99.3
NEWS	> 2	0.781	0.772–0.789	58.02	87.75	4.74	0.22	12.1	98.7

IEWS: International Early Warning Score, NEWS: National Early Warning Score, ICU: Intensive care unit, AUC: Area under the curve, CI: Confidence interval, LR: Likelihood ratio, PPV: Positive predictive value, NPV: Negative predictive value

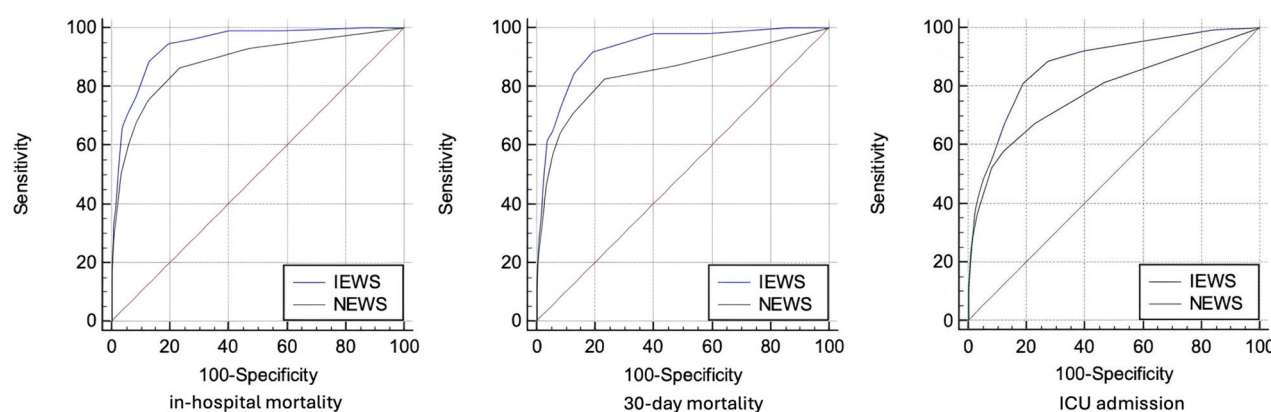
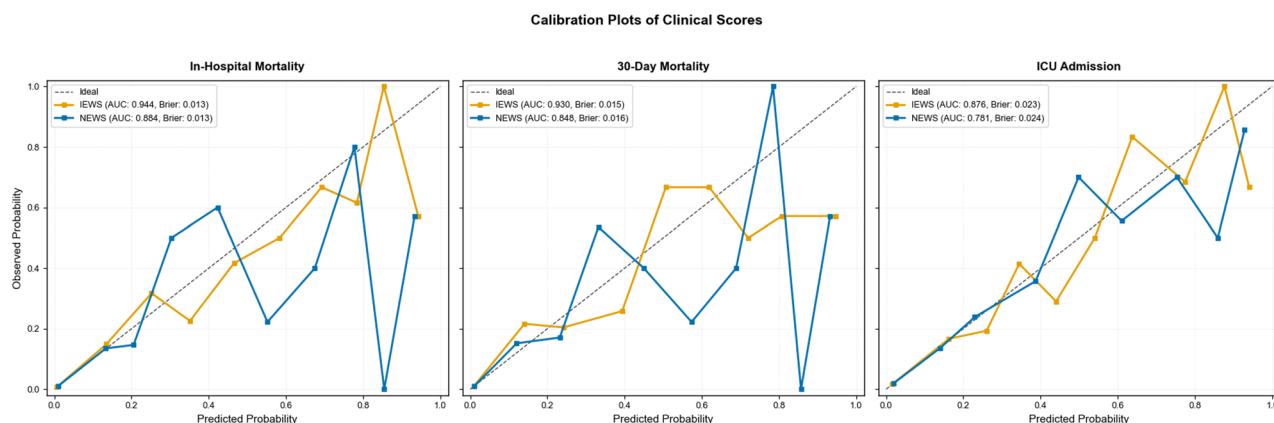
**Fig. 2** Receiver operating characteristic curves of the scores in different outcomes

Fig. 3 Calibration plots comparing the International Early Warning Score (IEWS) and the National Early Warning Score (NEWS) for in-hospital mortality, 30-day mortality, and ICU admission. Observed event rates were plotted against predicted probabilities across deciles of risk, with the dashed diagonal line representing perfect calibration. Across all outcomes, IEWS demonstrated closer alignment with the ideal line and smaller deviations than NEWS, indicating better overall calibration and more accurate estimation of absolute risk

consistently underestimated risk in several deciles, whereas IEWS demonstrated more stable agreement between predicted and observed mortality. IEWS had a slightly lower Brier score (0.015) than NEWS (0.016). For ICU admission, both scores showed generally acceptable calibration; however, IEWS displayed a closer overall fit to the ideal line, with less underestimation at higher

predicted probabilities. IEWS also had a marginally lower Brier score (0.023) compared with NEWS (0.024).

Decision curve analysis demonstrated that IEWS provided consistently higher standardized net benefit than NEWS for all three outcomes (Fig. 4). For in-hospital mortality, IEWS showed a modest but sustained net benefit advantage over NEWS between threshold

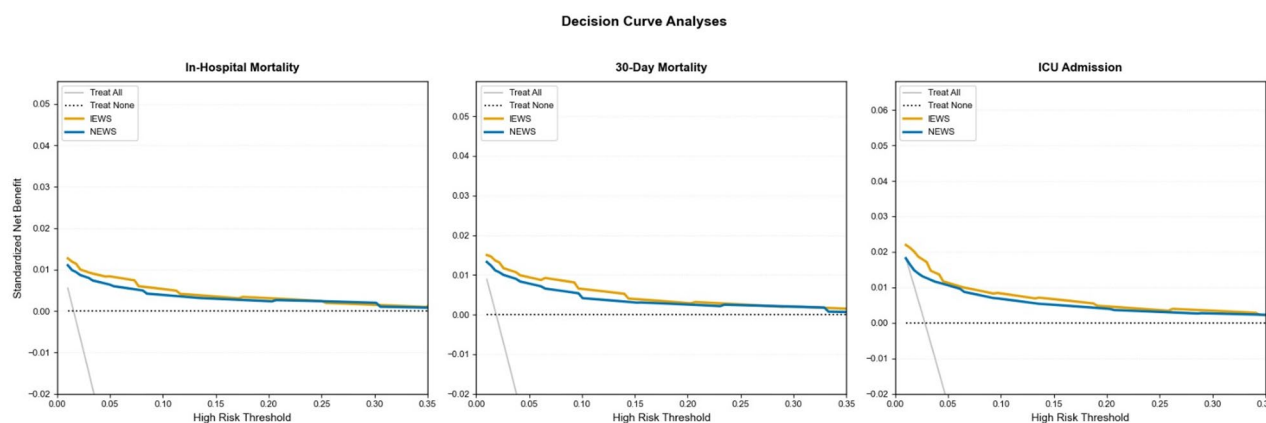


Fig. 4 Decision curve analyses comparing the standardized net benefit of the International Early Warning Score (IEWS) and the National Early Warning Score (NEWS) for in-hospital mortality, 30-day mortality, and ICU admission. Net benefit was plotted against threshold probabilities from 0.01 to 0.35, with “treat all” and “treat none” strategies shown for reference. Across all outcomes, IEWS provided consistently higher net benefit than NEWS at most clinically relevant thresholds, indicating greater potential clinical utility for risk-based decision-making in the emergency department

probabilities of 0.01–0.20. Similar patterns were observed for 30-day mortality and ICU admission, where IEWS offered greater clinical utility at lower thresholds and maintained comparable benefit at higher thresholds. Both scores performed better than “treat all” and “treat none” strategies throughout clinically relevant threshold ranges.

There were significant differences in many baseline characteristics between the two groups (Table 1). These findings underscore both clinical and demographic factors as significant determinants of in-hospital mortality.

Discussion

This study provides evidence that the IEWS, an age- and sex-adjusted refinement of NEWS, offers superior predictive accuracy compared to NEWS for in-hospital mortality, 30-day mortality, and ICU admission in adult patients presenting to the ED.

Consistent with existing evidence, our study confirms that vital signs and level of consciousness—core elements of most EWS—remain strongly associated with mortality risk [25–28]. By integrating age and sex into these established parameters, Candel et al. developed IEWS, which consistently outperformed NEWS in derivation and external validation cohorts [23]. Subsequent external validation in septic patients by Devia-Jaramillo et al. further reinforced the superiority of IEWS over NEWS [16]. However, a more recent study by Candel and Veldhuis reported substantially lower discriminative performance for IEWS (AUROC ~ 0.64), particularly in high-acuity ambulance-transported patients [29]. These discrepancies likely reflect methodological and population differences: our prospective cohort was considerably larger, included all ED presentations rather than only prehospital high-risk patients, and had lower and more balanced mortality rates, which may stabilize model estimates

and enhance discriminative accuracy. Additionally, differences in clinical workflows, timing of vital sign measurement, and outcome definitions across settings may further contribute to the variation in predictive performance observed between studies.

Our findings are in line with previous reports, showing that IEWS demonstrated higher accuracy than NEWS in predicting in-hospital mortality in our cohort (AUC: 0.944 vs. 0.884) [16, 17, 19, 23]. Whereas NEWS generally exhibits moderate-to-good discriminative performance for mortality outcomes, the excellent predictive performance of IEWS observed in this study suggests that it may offer improved risk stratification for ED patients. However, although higher discriminative accuracy is an important prerequisite for clinical usefulness, our study did not evaluate whether the application of IEWS in real-time ED practice leads to measurable improvements in clinical management, patient flow, or outcomes. Therefore, any potential clinical implications of IEWS—such as earlier recognition of high-risk patients or more appropriate allocation of monitoring resources—should be interpreted cautiously and considered hypothetical. Future implementation studies are required to determine whether these predictive improvements translate into meaningful clinical impact.

The higher age observed among non-survivors in our cohort may partly explain the strong predictive contribution of age within the IEWS model. Importantly, unlike many Western healthcare systems, Do Not Attempt Resuscitation (DNAR) orders are not legally permitted in Türkiye. Therefore, differences in resuscitation practices or treatment-limitation decisions did not influence mortality outcomes in our study. As a result, age-related differences in mortality likely reflect genuine differences in illness severity rather than variations in resuscitative intent.

To our knowledge, this is the first study to assess IEWS for predicting 30-day mortality and ICU admission. Consequently, no direct comparisons with prior reports could be made. In this prospective cohort, IEWS achieved good-to-excellent predictive performance for in-hospital mortality, 30-day mortality, and ICU admission (AUC: 0.944, 0.930, and 0.876, respectively), underscoring its promise as a reliable tool for early detection of clinical deterioration. However, as a newly developed scoring system, IEWS requires broader external validation in larger patient populations before it can be widely implemented in clinical practice.

Graham et al. and Sbiti-Rohr et al. reported poor predictive ability of NEWS for 30-day mortality (AUC: 0.61 and 0.65), whereas subsequent studies have shown moderate-to-good performance [18, 30–34]. Consistent with these latter reports, our study demonstrated good discriminative performance of NEWS in predicting 30-day mortality (AUC: 0.848). While IEWS outperformed NEWS overall, the original NEWS model nonetheless retained clinically relevant predictive value. Although several modifications of NEWS have been suggested to improve its prognostic accuracy, its original form remains a valid and practical tool. Variability in published results may be explained by differences in study populations, selected thresholds, and outcome prevalence [35].

The predictive ability of NEWS for ICU admission has been inconsistent across studies. While Graham et al. reported poor performance (AUC: 0.65), several other investigations demonstrated moderate-to-good discriminative capacity (AUC: 0.70–0.90) [6, 32, 34, 36–38]. In our cohort, NEWS achieved only moderate accuracy (AUC: 0.781). Conversely, IEWS showed superior performance with good discriminative ability (AUC: 0.876), and its higher sensitivity and specificity enabled more precise identification of patients with and without critical care needs in the ED.

In terms of calibration, an essential complement to discriminative performance, IEWS consistently showed better agreement between predicted and observed risks than NEWS across all evaluated outcomes. IEWS showed smaller deviations from the ideal 45-degree calibration line and lower or equal Brier scores for in-hospital mortality, 30-day mortality, and ICU admission. These findings suggest that IEWS not only discriminates risk more effectively but also provides more reliable absolute risk estimation, particularly in higher-risk strata. Improved calibration strengthens the clinical applicability of IEWS by ensuring more accurate prediction of absolute event probabilities.

In addition to its stronger discriminative performance and superior calibration, IEWS demonstrated slightly higher net benefit than NEWS across a wide range of clinically relevant risk thresholds. This suggests that IEWS

may offer greater practical utility in supporting early risk-based decisions in the ED, particularly when clinicians prioritize sensitivity at low-to-moderate threshold probabilities. These findings are consistent with the original derivation and validation work by Candel et al., reinforcing the potential of IEWS as an improved alternative to NEWS in real-world emergency care settings [23].

An additional strength of IEWS lies in its real-time applicability. Although it incorporates age- and sex-adjusted parameters, IEWS—like NEWS—relies solely on routinely collected physiological measurements and can therefore be calculated rapidly at the bedside without requiring laboratory data or additional diagnostics. This preserves its practicality for use in fast-paced emergency settings, enabling clinicians to identify high-risk patients promptly upon presentation.

The optimal IEWS thresholds in this study were determined using the Youden index, which statistically balances sensitivity and specificity for the best overall discrimination. However, in a fast-paced ED environment, the clinical practicality of these thresholds must also be considered. In our cohort, the Youden-derived cut-off for IEWS achieved excellent sensitivity while maintaining acceptable specificity, allowing early identification of high-risk patients without excessive false positives. This finding suggests that IEWS may facilitate earlier recognition of clinical deterioration and support timely triage and monitoring decisions. Nevertheless, further real-world validation is needed to confirm the optimal balance between diagnostic accuracy and operational feasibility across diverse healthcare settings.

Although EWSs were originally developed to detect short-term clinical deterioration—typically within hours to the first few days of presentation—we evaluated in-hospital and 30-day mortality to maintain consistency with prior NEWS and IEWS validation studies. However, these longer-term outcomes extend beyond the traditional scope of early warning score applications, as they may be influenced by inpatient management, comorbidities, and treatment decisions occurring outside the ED. Therefore, while our findings provide valuable information on overall prognostic performance, they should be interpreted with the understanding that EWSs are primarily designed to identify imminent deterioration rather than to predict long-term mortality. Future studies incorporating ED-specific outcomes, such as early clinical deterioration, in-ED cardiac arrest, or the need for time-critical interventions, would offer a more direct assessment of the real-time utility of these scores in emergency care.

Limitations

Several limitations should be acknowledged. First, the single-center design restricts the external validity of our results, and multicenter studies are warranted to confirm these findings across diverse populations. Second, the scores were derived exclusively from parameters measured at ED presentation, without accounting for dynamic changes during follow-up, which may have influenced predictive accuracy. Third, although NEWS was originally developed as a track-and-trigger system relying on repeated measurements, our evaluation was based solely on baseline ED values. This approach does not capture the dynamic monitoring advantage of NEWS and limits comparison with its intended clinical use. Fourth, the use of optimal cut-off values determined by the Youden index may limit generalizability, as different thresholds could potentially alter the predictive performance. In addition, although subgroup analyses (e.g., restricted to high-acuity triage levels or hospitalized patients) could potentially offer further clinical insights, such analyses were not prespecified in the study protocol or ethics committee approval. Performing post hoc subgroup analyses would substantially reduce the number of outcome events within each stratum, leading to underpowered estimates with limited reliability and generalizability. Therefore, subgroup analyses were not conducted, and future studies specifically designed and adequately powered for these comparisons are warranted. Finally, the study was conducted over a relatively short period (July–August 2024), which may limit temporal generalizability. Seasonal fluctuations in ED patient volume, case mix, and operational factors—such as staffing patterns during summer months—could have influenced model performance. Although no major organizational changes occurred during the study window, validation across longer and seasonally diverse periods is warranted.

Conclusion

The present study demonstrated that IEWS showed higher discriminative performance than NEWS in predicting in-hospital mortality, 30-day mortality, and ICU admission among non-traumatic ED patients. Although NEWS performed more modestly, both scores retained prognostic value for risk stratification at ED presentation. Because our study evaluated predictive accuracy rather than real-time clinical impact, these findings should be interpreted as evidence of statistical performance rather than direct indicators of improved patient outcomes or decision-making in the ED. Further multicenter studies and implementation-focused research are needed to determine whether the superior discriminative ability of IEWS translates into meaningful clinical benefits and improved operational management in diverse emergency care settings.

Abbreviations

IEWS	International Early Warning Score
NEWS	National Early Warning Score
EWS	Early Warning Score
ED	Emergency department
ICU	Intensive care unit
IQR	Interquartile range
CI	Confidence interval
AUC	Area under the curve

Acknowledgements

We would like to thank Mr. Ender İçen for his contributions to data collection and data entry, and Mr. Mert Can Yaman for his valuable support in advanced data analyses.

Author contributions

FB: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Data curation, Conceptualization. Bl: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Data curation, Conceptualization, Visualization, Validation. ZK: Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. OE: Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. TD: Writing – review & editing, Project administration, Visualization, Validation, Supervision. All authors read and approved the final submitted version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

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

Declarations

Ethical approval and consent to participate

The institutional review board of Kırıkkale University Faculty of Medicine approved the study (Approval ID: 2024.04.11; Date: Apr. 17, 2024). Informed consent was obtained from all participants. This study was carried out in accordance with the principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Emergency Medicine, Şemdinli State Hospital, Hakkari, Türkiye

²Department of Emergency Medicine, Kırıkkale University Faculty of Medicine, Kırıkkale, Türkiye

³Department of Emergency Medicine, Düzce Atatürk State Hospital, Düzce, Türkiye

Received: 15 October 2025 / Accepted: 5 December 2025

Published online: 11 December 2025

References

- Beştemir A, Aydın H. Yıllık 300 Milyon Hasta muayenesi; Türkiye'de 2. ve 3. basamak Kamu sağlık tesisleri acil Servis ve poliklinik Hizmetlerinin değerlendirilmesi. *Sakarya Tıp Dergisi*. 2022;12(3):496–502.
- Altun M, Kudu E. Overcrowding in emergency departments: A scoping literature review. *Anatol J Emerg Med*. 2025;8(2):94–100.
- Darraj A, Hudays A, Hazazi A, Hobani A, Alghamdi A. The association between emergency department overcrowding and delay in treatment: a systematic review. *Healthcare (Basel)*. 2023;11:385.

4. Pines JM, Hilton JA, Weber EJ, Alkemade AJ, Al Shabanah H, Anderson PD, et al. International perspectives on emergency department crowding. *Acad Emerg Med*. 2011;18(12):1358–70. <https://doi.org/10.1111/j.1553-2712.2011.01235.x>.
5. Sartini M, Carbone A, Demartini A, Giribone L, Oliva M, Spagnolo AM, et al. Overcrowding in emergency department: causes, consequences, and solutions—a narrative review. *Healthc (Basel)*. 2022;10(9). <https://doi.org/10.3390/healthcare10091625>.
6. İlhan B, Bozdereli Berikol G, Dogan H. The predictive value of modified risk scores in patients with acute exacerbation of COPD: a retrospective cohort study. *Intern Emerg Med*. 2022;17(7):2119–27. <https://doi.org/10.1007/s11739-022-03048-z>.
7. İlhan B, Bozdereli Berikol G, Dogan H. The prognostic value of rapid risk scores among patients with community-acquired pneumonia: A retrospective cohort study. *Wien Klin Wochenschr*. 2023;135:19–20. <https://doi.org/10.1007/s00508-023-02238-9>.
8. Demir MC, İlhan B. Performance of the pandemic medical early warning score (PMEWS), simple triage scoring system (STSS) and Confusion, Uremia, respiratory rate, blood pressure and age ≥ 65 (CURB-65) score among patients with COVID-19 pneumonia in an emergency department triage setting: a retrospective study. *Sao Paulo Med J*. 2021;139(2):170–7.
9. Bellomo R, Goldsmith D, Uchino S, Buckmaster J, Hart GK, Opdam H, et al. A prospective before-and-after trial of a medical emergency team. *Med J Aust*. 2003;179(6):283–7. <https://doi.org/10.5694/j.1326-5377.2003.tb05548.x>.
10. Buist MD, Moore GE, Bernard SA, Waxman BP, Anderson JN, Nguyen TV. Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: preliminary study. *BMJ*. 2002;324(7334):387–90. <https://doi.org/10.1136/bmj.324.7334.387>.
11. Jones SL, Ashton CM, Kiehne LB, Nicolas JC, Rose AL, Shirkey BA, et al. Outcomes and resource use of Sepsis-associated stays by presence on Admission, Severity, and hospital type. *Med Care*. 2016;54(3):303–10. <https://doi.org/10.1097/MLR.0000000000000481>.
12. Royal College of Physicians. National Early Warning Score (NEWS). Standardising the assessment of acutellness severity in the NHS. Report of a working party. London: RCP; 2012.
13. Morgan R, Williams F, Wright M. An early warning scoring system for detecting developing critical illness. *Clin Intensive Care*. 1997;8(2):100.
14. Badr MN, Khalil NS, Mukhtar AM. Effect of National early warning scoring system implementation on cardiopulmonary Arrest, unplanned ICU Admission, emergency Surgery, and acute kidney injury in an emergency Hospital, Egypt. *J Multidiscip Healthc*. 2021;14:1431–42. <https://doi.org/10.2147/JMDH.S312395>.
15. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National early warning score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation*. 2013;84(4):465–70.
16. 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). <https://doi.org/10.3390/jcm14144869>.
17. Lee YS, Choi JW, Park YH, Chung C, Park DI, Lee JE, et al. Evaluation of the efficacy of the National early warning score in predicting in-hospital mortality via the risk stratification. *J Crit Care*. 2018;47:222–6. <https://doi.org/10.1016/j.jcc.2018.07.011>.
18. van Dam P, Lievens S, Zelis N, van Doorn W, Meex SJR, Cals JWL, et al. Head-to-head comparison of 19 prediction models for short-term outcome in medical patients in the emergency department: a retrospective study. *Ann Med*. 2023;55(2):2290211. <https://doi.org/10.1080/07853890.2023.2290211>.
19. Redfern OC, Smith GB, Prytherch DR, Meredith P, Inada-Kim M, Schmidt PE. A comparison of the quick sequential (Sepsis-Related) organ failure assessment score and the National early warning score in Non-ICU patients With/Without infection. *Crit Care Med*. 2018;46(12):1923–33. <https://doi.org/10.1097/CCM.0000000000003359>.
20. Pirneskoski J, Lääperi M, Kuisma M, Oikola KT, Nurmi J. Ability of prehospital news to predict 1-day and 7-day mortality is reduced in older adult patients. *Emerg Med J*. 2021;38(12):913–8.
21. Lee J, Lee D, Lee B, No E. Association between pre-hospital National early warning score and in-hospital mortality in patients with traumatic brain injury. *Turkish J Trauma Emerg Surg*. 2023;29(3):292.
22. Patel R, Nugawela MD, Edwards HB, Richards A, Le Roux H, Pullyblank A, et al. Can early warning scores identify deteriorating patients in pre-hospital settings? A systematic review. *Resuscitation*. 2018;132:101–11.
23. Candel BGJ, Nissen SK, Nickel CH, Raven W, Thijssen W, Gaakeer MI, et al. Development and external validation of the international early warning score for improved Age- and Sex-Adjusted In-Hospital mortality prediction in the emergency department. *Crit Care Med*. 2023;51(7):881–91. <https://doi.org/10.1097/CCM.0000000000005842>.
24. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26(6):565–74. <https://doi.org/10.1177/0272989X06295361>.
25. Candel BG, Duijzer R, Gaakeer MI, Ter Avest E, Sir Ö, Lameijer H, et al. The association between vital signs and clinical outcomes in emergency department patients of different age categories. *Emerg Med J*. 2022;39(12):903–11.
26. Ljunggren M, Castrén M, Nordberg M, Kurland L. The association between vital signs and mortality in a retrospective cohort study of an unselected emergency department population. *Scand J Trauma Resusc Emerg Med*. 2016;24(1):21.
27. Candel BG, Dap S, Raven W, Lameijer H, Gaakeer MI, de Jonge E, et al. Sex differences in clinical presentation and risk stratification in the emergency department: an observational multicenter cohort study. *Eur J Intern Med*. 2022;95:74–9.
28. Nissen SK, Candel BG, Nickel CH, de Jonge E, Ryg J, Bogh SB, et al. The impact of age on predictive performance of National early warning score at arrival to emergency departments: development and external validation. *Ann Emerg Med*. 2022;79(4):354–63.
29. Candel BGJ, Veldhuis LI. Updating the international early warning score with frailty and comparing to gestalt for prediction of 3-day critical illness and mortality in emergency department patients. *Intern Emerg Med*. 2025. <https://doi.org/10.1007/s11739-025-04096-x>.
30. Brink A, Alsma J, Verdonchot RJCg, Rood PPM, Zietse R, Lingsma HF, 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.
31. Eckart A, Hauser SI, Kutz A, Haubitz S, Hausfater P, Amin D, et al. Combination of the National early warning score (NEWS) and inflammatory biomarkers for early risk stratification in emergency department patients: results of a multinational, observational study. *BMJ Open*. 2019;9(1):e024636. <https://doi.org/10.1136/bmjopen-2018-024636>.
32. Graham CA, Leung LY, Lo RSL, Yeung CY, Chan SY, Hung KKC. NEWS and qSIRS superior to qSOFA in the prediction of 30-day mortality in emergency department patients in Hong Kong. *Ann Med*. 2020;52(7):403–12. <https://doi.org/10.1080/07853890.2020.1782462>.
33. Lee SB, Kim DH, Kim T, Kang C, Lee SH, Jeong JH, et al. Emergency department triage early warning score (TREWS) predicts in-hospital mortality in the emergency department. *Am J Emerg Med*. 2020;38(2):203–10. <https://doi.org/10.1016/j.ajem.2019.02.004>.
34. Sbiri-Rohr D, Kutz A, Christ-Crain M, Thomann R, Zimmerli W, Hoess C, et al. The National early warning score (NEWS) for outcome prediction in emergency department patients with community-acquired pneumonia: results from a 6-year prospective cohort study. *BMJ Open*. 2016;6(9):e011021. <https://doi.org/10.1136/bmjopen-2015-011021>.
35. Kudu E, İlhan B. Optimizing diagnostic precision: the role of cut-off selection in predictive performance studies. *Am J Emerg Med*. 2025;95:285–6. <https://doi.org/10.1016/j.ajem.2025.06.043>.
36. Dadeh A-a, Kulparat M. Predictive performance of the news–lactate and news towards mortality or need for critical care among patients with suspicion of sepsis in the emergency department: a prospective observational study. *Open Access Emerg Med*. 2022:619–31.
37. Jo S, Yoon J, Lee JB, Jin Y, Jeong T, Park B. Predictive value of the National early warning Score–Lactate for mortality and the need for critical care among general emergency department patients. *J Crit Care*. 2016;36:60–8.
38. Oduncu AF, Kiyani GS, Yalcinli S. Comparison of qSOFA, SIRS, and NEWS scoring systems for diagnosis, mortality, and morbidity of sepsis in emergency department. *Am J Emerg Med*. 2021;48:54–9. <https://doi.org/10.1016/j.ajem.2021.04.006>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.