BMJ Open

Performance of universal early warning scores in different patient subgroups and clinical settings: a systematic review

Baneen Alhmoud,[1] Timothy Bonnici,[1,2] Riyaz Patel ,[2,3,4] Daniel Melley,[4] Bryan Williams ,[2,3] Amitava Banerjee [1,2,4]

To cite: Alhmoud B, Bonnici T, Patel R, *et al*. Performance of universal early warning scores in different patient subgroups and clinical settings: a systematic review. *BMJ Open* 2021;**11**:e045849. doi:10.1136/bmjopen-2020-045849

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-045849).

Received 13 October 2020
Revised 01 March 2021
Accepted 04 March 2021

[1]Institute of Health Informatics, University College London, London, UK
[2]University College London Hospitals NHS Trust, London, UK
[3]Institute of Cardiovascular Science, University College London, London, UK
[4]Barts Health NHS Trust, London, UK

Correspondence to
Dr Amitava Banerjee;
ami.banerjee@ucl.ac.uk

ABSTRACT

Objective To assess predictive performance of universal early warning scores (EWS) in disease subgroups and clinical settings.

Design Systematic review.

Data sources Medline, CINAHL, Embase and Cochrane database of systematic reviews from 1997 to 2019.

Inclusion criteria Randomised trials and observational studies of internal or external validation of EWS to predict deterioration (mortality, intensive care unit (ICU) transfer and cardiac arrest) in disease subgroups or clinical settings.

Results We identified 770 studies, of which 103 were included. Study designs and methods were inconsistent, with significant risk of bias (high: n=16 and unclear: n=64 and low risk: n=28). There were only two randomised trials. There was a high degree of heterogeneity in all subgroups and in national early warning score (I^2=72%–99%). Predictive accuracy (mean area under the curve; 95% CI) was highest in medical (0.74; 0.74 to 0.75) and surgical (0.77; 0.75 to 0.80) settings and respiratory diseases (0.77; 0.75 to 0.80). Few studies evaluated EWS in specific diseases, for example, cardiology (n=1) and respiratory (n=7). Mortality and ICU transfer were most frequently studied outcomes, and cardiac arrest was least examined (n=8). Integration with electronic health records was uncommon (n=9).

Conclusion Methodology and quality of validation studies of EWS are insufficient to recommend their use in all diseases and all clinical settings despite good performance of EWS in some subgroups. There is urgent need for consistency in methods and study design, following consensus guidelines for predictive risk scores. Further research should consider specific diseases and settings, using electronic health record data, prior to large-scale implementation.

PROSPERO registration number PROSPERO CRD42019143141.

Strengths and limitations of this study

- ► The first systematic review to investigate the performance of early warning scores (EWS) in different patient disease subgroups and clinical settings.
- ► Meta-analysis was performed for different EWS and national EWS validation studies in different disease and clinical setting subgroups.
- ► This study is limited to specific diseases and settings and does not consider the use of EWS in the general population.
- ► Analysis of predictive accuracy of EWS is based on area under the curve, not other validation measures.
- ► During the study period 1997–2019, approaches to EWS and their validation have changed.

INTRODUCTION

Across diseases, patient deterioration can range from critical care review and sepsis to cardiorespiratory arrest and death.[1 2] Delays or failures in timely detection of deterioration adversely affect prognosis, morbidity, mortality and healthcare utilisation.[3] For example, the 20 000 in-hospital cardiac arrests per year in England are associated with costs of £50 million for resuscitation and postarrest care.[4] Around the world, earlier recognition and prevention of deterioration in unwell patients has far-reaching implications for reduction in mortality and morbidity, reduction in the cost of healthcare and allocation of scarce high dependency and critical care resources. Preventive interventions are needed to overcome these challenges.[5]

Specific characteristics have long been known to be associated with deteriorating patient health,[2 5–8] including physiological parameters, such as heart rate and blood pressure.[5 9–12] Early warning scores (EWS), widely used in high-income countries, were borne out of the need for early detection of patient deterioration. EWS are tools derived from prediction models that assess patient characteristics and physiological parameters to stratify the risk of developing a worsening event or need for medical attention.[13] The algorithms underlying EWS can be 'aggregate-weighted' to sum up a set of parameters to produce a score or use more advanced statistical modelling.[14] EWS inform clinical decision making, enabling escalation of attention and care when required. Universal tools, such

Box 1 Definitions

- Universal early warning scores (EWS): EWS that are globally adopted and applicable in every setting and for any disease subgroup.
- Standardised EWS: EWS model with a set of parameters used in a unified approach to predict deterioration in any patient subgroup.[8 23]
- External validation: evaluation of the model's predictive accuracy with data different than the one sued for model development.[27]
- Internal validation: evaluation of a model's predictive accuracy with the same data set used for the development or in a population in which the model is intended for use.[27]
- Discrimination: the ability of a model to distinguish between the patients who will develop an outcome of interest and the ones who will not.[26]
- Calibration: the accuracy of risk estimates in relation to the observed number of events.[73]

as the modified early warning score (MEWS),[15] were developed for use across different hospital settings, but specialised, non-standard EWS are also designed for particular subgroups, for example, Rapid Emergency Medicine Score[16] and Quick Sequential Organ Failure Assessment (qSOFA)[17] for patients with infections. In recognising different settings, EWS may have compromised simplicity and timeliness of assessment.[13] For example, a number of EWS rely on parameters that do not exist in the first hours of assessment, such as blood investigations and imaging.[1 18 19]

From fragmented implementation and inadequate early assessment via specialised tools, EWS have shifted back to universal prediction models, particularly, the national early warning score (NEWS),[20] followed by NEWS2.[21] NEWS was designed to produce a universal assessment of acute illness severity across the National Health Service (NHS).[22] While showing good discrimination compared with other EWS, especially in predicting mortality, there was a need to accommodate additional clinical parameters in the score. The updated NEWS2, emphasising appropriate scoring for type 2 respiratory failure, confusion and severe sepsis,[21] was formally endorsed by NHS England[23] to be the EWS used in acute care. However, there have been concerns regarding excessive calls to clinicians, administrative workload and variable symptoms across diseases and settings.[24] The effectiveness of the universal EWS (box 1) with standardised use across all settings is not clear in specific disease populations[25] and requires validation to estimate discrimination and calibration, like other clinical prediction models.[26] While internal validation is useful, generalisability and reproducibility needs external validation.[27]

Systematic reviews have evaluated EWS in prehospital, intensive care unit (ICU) and general settings,[3 28 29] and sepsis,[15] with narrow inclusion criteria and inadequate assessment of study quality. A recent systematic review evaluated development and validation of EWS in general

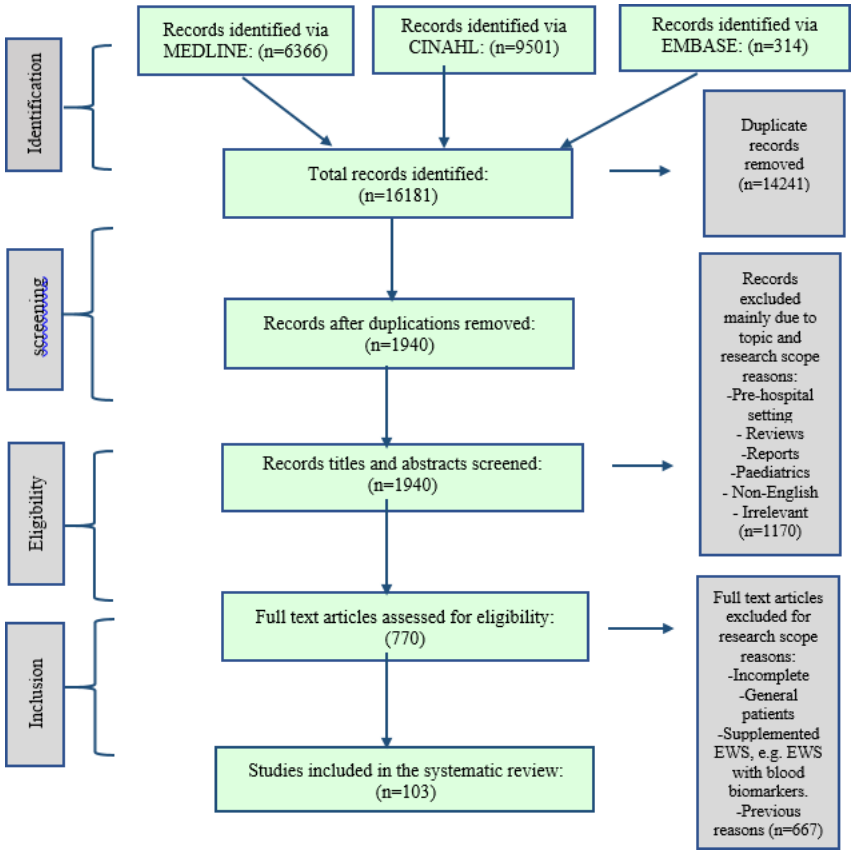

Figure 1 Search strategy and included studies regarding universal early warning scores (EWS) in different disease subgroups and clinical settings.

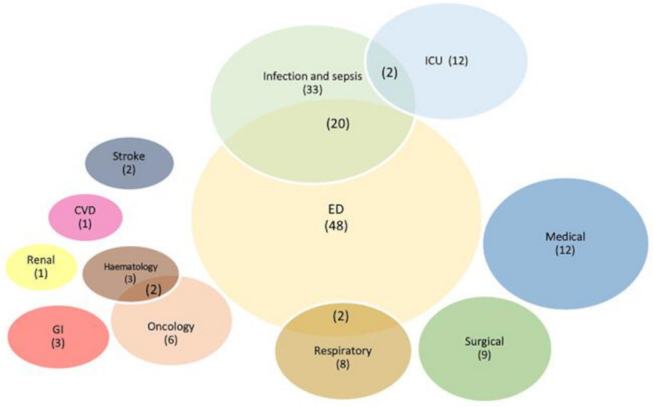

Figure 2 Number of studies regarding performance of early warning scores in different disease subgroups and clinical settings. Each bubble represents the disease subgroup and/or setting where different early warning scores were examined. The size of the bubble represents the number of studies (n), and overlapping bubbles show studies where disease subgroup and settings overlap. CVD, cardiovascular diseases; ED, emergency department; GiI, gastrointestinal diseases; ICU, intensive care unit.

patients but did not include studies in specific disease subgroups or settings.[30]

OBJECTIVE

In a systematic review, we will assess performance of universal EWS in particular diseases and clinical settings in predicting mortality, transfer to ICU and cardiac arrest.

METHODS

Search strategy

The protocol adhered to Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols guidelines.[31] Published articles were identified in MEDLINE, CINHAL and Embase, between 1997 (initial development of EWS) and 2019. The Cochrane database was searched for systematic reviews (CDSR) and trials (CENTRAL). For grey literature, Google Scholar was searched. During the screening procedure, studies were added from references in review articles and studies. Search strategies were developed by two authors (BA and AB) and reviewed by a third author (TB). Terms used for searching databases include terms for early warning or track and trigger scores and acronyms, identified subgroups and settings (eg, Medical Subject Heading (MeSH)) and free-text search terms (figure 1; online supplemental methods).

Inclusion and exclusion criteria

Patient subgroups were identified according to disease categories and clinical settings (online supplemental methods). Studies were included if: (1) validation of a universal EWS with standardised prediction model in adult patients; (2) EWS validation was in a specific setting or disease; (3) the performance of the EWS, or the impact on mortality, transfer to ITU and cardiac arrest, was examined; and (4) they were prospective or retrospective cohort, cross-sectional, case–control design or trials.

Studies were excluded if: (1) patients were less than 16 years of age; (2) EWS performance was only examined in derivation, not validation; (3) non-universal EWS was developed for a specific subgroup, for example, obstetric early warning score for obstetric patients or qSOFA for patients with infections; or (4) EWS validation was performed in a general patient dataset or setting, for example, validation in a general hospital without consideration of hospital subgroups.

Data extraction

Articles were screened by title and abstract by one author (BA), then full-text screening was by two reviewers (BA and AB). Data were extracted independently by two reviewers (BA and AB) using a standardised and piloted data form. A third reviewer (TB) resolved any disagreements. Items for extraction for studies examining predictive accuracy were based on the CHARMS[32] checklist, except for tool derivation, which was excluded.

Quality assessment

Risk of biases in validation studies was assessed using Prediction model Risk Of Bias ASsessment Tool (PROBAST),[33] which classifies studies as low, unclear or high risk of bias in four aspects: participant selection, predictors, outcomes and analysis within the overall risk of bias and the study applicability domains.

Evidence synthesis

We conducted the analysis using MS Excel and R programs. We summarised the results using descriptive statistics and graphical plots. Meta-analysis was performed, in different subgroups, using area under the curve (AUC) for identified universal EWS and for NEWS in studies. Fisher-Z transformation for correlation coefficients was conducted for AUC into normally distributed Z with 95% CI to evaluate the effect size and test for the heterogeneity. Where applicable, narrative synthesis was conducted.

Patient and public involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS

Study characteristics

Of the 16 181 articles identified by our search, we screened 1355 articles by title and abstract, assessing 770 articles in full for eligibility. We included 103 studies, published between 2006 and 2019, in the final stage. These studies were predominantly observational (retrospective=65, prospective=36 and RCT=2). Emergency department (ED) (n=48) was the most common clinical setting, followed by medical (n=12), ICU (n=12) and surgical (n=9) settings. Sepsis (n=33) was the most common disease subgroup. Other subgroups ranged

Table 1 Studysetting and design for included studies of predictive performance for early warning scores in patient subgroups and settings

		Subgroups								Settings				Study design				
Author, year	Country	CVD	GI	Haematology	Renal	Stroke	Oncology	Respiratory	Infection/ sepsis	ICU	ED	Surgical	Medical	Retrospective	Prospective	RCT	Case–control	Number of patients
Kellett, 2012[46]	Canada	●	○	○	●	●	●	○	○	●	○	●	●	●	○	○	○	10007
Kim, 2017	Korea	○	●	○	○	○	○	○	○	○	○	○	○	●	○	○	○	2172
Bozkurt, 2015	Turkey	○	●	○	○	○	○	○	○	○	○	○	○	○	●	○	○	202
Seak, 2017	Taiwan	○	●	○	○	○	○	○	○	○	○	○	○	●	○	○	○	66
Hu, 2016[66]	USA	○	○	●	○	○	●	○	○	○	○	○	○	●	○	○	○	565
Liljehult, 2016[53]	Denmark	○	○	○	○	●	○	○	○	○	○	○	○	●	○	○	○	274
Mulligan,2010[60]	UK	○	○	●	○	○	○	○	○	○	○	○	○	○	●	○	○	71
Cooksley, 2012[61]	UK	○	○	○	○	○	●	○	○	○	○	○	○	●	○	○	○	840
Vaughn, 2018[39]	USA	○	○	○	○	○	●	○	○	○	○	○	○	●	○	○	○	504
Young, 2014	USA	○	○	●	○	○	●	○	○	○	○	○	○	●	○	○	○	61
Von, 2007	UK	○	○	○	○	○	●	○	○	○	○	○	○	●	○	○	○	43
Pedersen, 2018	Denmark	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	○	11266
Forster, 2018	UK	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	○	8812
Pimentel, 2018[41]	UK	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	○	1394
Sbiti-Rohr, 2016[52]	Switzerland	○	○	○	○	○	○	●	○	○	●	○	○	○	○	●	○	925
Brabrand, 2017	Denmark	○	○	○	○	○	○	●	○	○	○	○	●	○	●	○	○	570
Jo, 2016	Korea	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	○	553
Barlow, 2007	UK	○	○	○	○	○	○	●	○	○	○	○	○	○	●	○	○	419
Bilben, 2016[55]	Norway	○	○	○	○	○	○	●	○	○	●	○	○	○	●	○	○	246
Delahanty, 2019	USA	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	920026
Redfern, 2018	UK	○	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	241996
Churpek, 2017[35]	USA	○	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	53849
Faisal, 2019	UK	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	36161
Churpek,2017[35]	USA	○	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	18523
Henry, 2015[40]	USA	○	○	○	○	○	○	○	●	●	○	○	○	●	○	○	○	13014
Brink,2019[34]	TheNetherlands	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	8204
De groot, 2017	TheNetherlands	○	○	○	○	○	○	○	●	○	●	○	○	○	●	○	○	2280
Corfield, 2014	UK	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	2003
Goulden, 2018[56]	UK	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	1818
Khwannimit, 2019[38]	Thailand	○	○	○	○	○	○	○	●	●	○	○	○	●	○	○	○	1589
Ghanem-oubi, 2011[63]	Israel	○	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	1072

Continued

Table 1 Continued

		Subgroups								Settings				Study design				
Author, year	Country	CVD	GI	Haematology	Renal	Stroke	Oncology	Respiratory	Infection/ sepsis	ICU	ED	Surgical	Medical	Retrospective	Prospective	RCT	Case–control	Number of patients
Saeed, 2019	UK, France, Italy, Sweden &; Spain	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	1058
Innocenti, 2018[65]	Italy	○	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	742
Camm, 2018	UK	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	533
Tirotta, 2017	Italy	○	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	526
Pong, 2019	Malaysia	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	364
Prabhakar, 2019	Malaysia	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	343
Martino, 2018	Italy	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	310
Vorwerk, 2009[19]	UK	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	308
Qin, 2017[51]	China	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	292
Schmedding, 2019	Gabon	○	○	○	○	○	○	○	●	○	●	○	○	○	●	○	○	277
Albur, 2016[64]	UK	○	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	245
çildir, 2013[36]	Turkey	○	○	○	○	○	○	○	●	○	●	○	○	○	●	○	○	230
Chiew, 2019[54]	Malaysia	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	214
Samsudin, 2018	Malaysia	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	214
Chang, 2018	China	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	152
Geier, 2013	Germany	○	○	○	○	○	○	○	●	○	●	○	○	○	●	○	○	151
Asiimwe, 2015	Uganda	○	○	○	○	○	○	○	●	○	○	○	○	○	●	○	○	150
Hung, 2017	Taiwan	○	○	○	○	○	○	○	●	○	●	○	○	●	○	○	○	114
Garcea, 2006	UK	○	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	110
Yoo, 2015	Korea	○	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	100
Siddiqui, 2017[37]	Malaysia	○	○	○	○	○	○	○	●	●	○	○	○	●	○	○	○	58

Studies are ranked according to sample size from largest to smallest in each subgroup.
Note: black dots in the subgroup column represent the disease or the settings where the sample was studied and brown dots in the study by Kellet (2012) represent different samples for each subgroup.
CVD, cardiovascular disease; ED, emergency department; GI, gastrointestinal diseases; ICU, intensive care unit.

from respiratory (n=8) to renal (n=1) (figures 1 and 2). Mortality was the main studied outcome. Cardiac arrest was infrequently studied (n=8).

Quality assessment

There was a significant risk of bias found in majority of studies (high risk=16; unclear risk=64) and low risk in only 28 studies. In terms of applicability, narrow inclusion of conditions in a certain disease group was commonly related to risk of bias, while in general settings, biases were often due to low sample size or unspecified timing of EWS assessment. There was a wide variation in sample size (median: 551 and range: 43–920 029). There was variation in defining study population by number of patients, hospital admissions or not specifying the particular study sample. Almost half of the studies (n=49; 48%) validated in <500 patients with either multiple observations or a single observation set (tables 1–4). External validation was more common (n=83) than internal validation (n=18), and two studies included internal and external validations (online supplemental table S1).

EWS validation in patient subgroups

Subgroups and EWS

In the studies validating EWS, there was heterogeneity in subgroup definitions, models and methods of predictive accuracy. There was overlap between diseases and settings commonly between studies of patients with infections receiving care in ED[34–36] and patients with sepsis admitted to ICU.[37 38] EWS models that were integrated with electronic health records (EHRs) were examined in recent studies (n=9). Research on datasets using EWS-embedded EHRs had larger sample sizes, ranging from 504[39] to 13 014 patients[40] (tables 1–4), with moderate to high predictive ability (AUC: 0.65–0.85). Several studies included comparison between different EWS in the same cohort (n=21)[35 38 41] (online supplemental table S2).

Methodology

There was significant heterogeneity in methods across studies. The majority of studies were observational. Evaluation of predictive accuracy of different EWS in the same study was common.[22 42–44] To measure accuracy of EWS, AUC was most commonly used (n=94), especially when comparing different EWS in the same study.[22 45] Presentation of results was variable; for example, confidence intervals were missing in many studies. Other measures, such as analysing sensitivity and specificity, prognostic index and ORs, were found in only eight studies (tables 1–4). Consequently, it was only feasible to analyse predictive accuracy in studies where AUC was the selected measure.

Timing from EWS assessment to endpoints was variable. Many studies included (n=43) AUC within 24–48 hours, while 11 studies had endpoints more than 48 hours after EWS. However, the majority (n=65; 63%) did not specify time horizon or in-hospital outcome.

Predictive performance of EWS

Outcomes were most commonly mortality, transfer to ICU, developing sepsis (in patients with infections) and cardiac arrest. Few studies examined other outcomes, for example, respiratory arrest (n=1) and organ failure (n=4). Mortality, ICU admission and cardiac arrest were best predicted in medical (AUC mean: 0.74, 0.75 and 0.74)[46–48] and surgical settings (0.80, 0.79 and 0.75),[49 50] and respiratory diseases (0.75, 0.80 and 0.75), respectively. EWS prediction of sepsis had reasonable predictive performance in all subgroups (AUC: 0.71–0.79) and infectious diseases in particular (AUC: 0.79). Certain outcomes related to specific disease groups were not studied, for example, cardiac arrest was not studied in cardiac patients[22]; respiratory arrest was not tested in respiratory patients.[46–52]

The best predictive performance was found in studies examining cardiac,[46] stroke[46 53] and renal[46] diseases (AUC: 0.93, 0.88 and 0.87, respectively). In emergency settings, predictive accuracy was variable (AUC: 0.56–0.91).[54–58] In haematology and oncology diseases, EWS predictive accuracy was suboptimal in mortality (online supplemental figure S1), cardiac arrest and ICU transfer (AUC: 0.52–0.69; figures 3 and 4).[59–61] EWS prediction of ICU transfer was reasonable in ED,[57 62] infectious diseases,[63 64] and where both groups overlap,[42 65] but not in gastroenterology and haematology (AUC: 0.64 and 0.60)[60 66] (online supplemental figure S2) Cardiac arrest was the least examined outcome among the three endpoints (n=8) and unstudied in cardiac diseases (figures 3 and 4, online supplemental figure S3).

For mortality prediction, meta-analysis of included EWS showed high degree of statistical heterogeneity across all subgroups (I^2=72%–99%) (figure 5). In validation studies of NEWS in different disease subgroups, there was also significant heterogeneity (I^2=99%; figure 6).

DISCUSSION

In this comprehensive review of universal EWS across all diseases and settings, we had three main findings. First, EWS studies in different diseases and clinical settings were heterogeneous in methodology, predictive performance measures and number of studies in each subgroup. Second, validation of EWS is limited in specialised settings, including cardiac disease. Third, despite widespread EHR and EWS integration, few studies have explored EHR-based EWS.

Inconsistency in evaluation and the lack of high-quality validation make the evidence of validity questionable, ultimately affects how EWS can and should be used in clinical practice as a risk score for deterioration prediction. Heterogeneity across studies in all subgroups challenges implementation of EWS in all diseases and all settings.

Table 2 Early warning scores, predictive measures and outcomes for included studies in patient subgroups and settings

		EWS										Predictive measure	Outcomes studied				
Author, year	EHR	VIEWS	MEWS	EWS	NEWS	NEWS2	SOS	Worthing	HOTEL	TREWS	HEWS		Mortality	ICU	CA	RA	Sepsis
Kellett, 2012[46]	✗	●	○	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Kim, 2017	✓	●	○	○	○	○	○	○	○	○	○	AUC	✗	✓	✗	✗	✗
Bozkurt 2015	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Seak, 2017	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Hu, 2016[66]	✓	○	○	●	○	○	○	○	○	○	○	AUC	✓	✓	✓	✗	✗
Liljehult, 2016[53]	✗	○	○	●	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Mulligan, 2010[60]	✗	○	○	●	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Cooksley, 2012[61]	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✓	✓	✗	✗
Vaughn, 2018[39]	✓	○	●	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Young, 2014	✗	○	●	○	○	○	○	○	○	○	○	Sens & Spec	✓	✗	✗	✗	✗
Von, 2007	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Pedersen, 2018	✓	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Forster, 2018	✓	○	○	○	●	○	○	○	○	○	○	Sens & Spec	✓	✗	✗	✗	✗
Pimentel 2018[41]	✓	○	○	○	●	●	○	○	○	○	○	AUC	✓	✓	✓	✗	✗
Sbiti-Rohr, 2016[52]	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Brabrand, 2017	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Jo, 2016	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Barlow, 2007	✗	○	○	●	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Bilben, 2016[55]	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Delahanty *et al*, 2019	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✓
Redfern, 2018	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Churpek, 2017[35]	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Faisal, 2019	✗	○	○	○	●	○	○	○	○	○	○	AUC	✗	✗	✗	✗	✓
Churpek, 2017	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Henry, 2015[40]	✓	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Brink, 2019[34]	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
de Groot, 2017[43]	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Corfield, 2014	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Goulden, 2018[56]	✓	○	○	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Khwannimit, 2019[38]	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Ghanem-Zoubi, 2011[63]	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Saeed, 2019	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗

Continued

Table 2 Continued

Author, year	EHR	EWS										Predictive measure	Outcomes studied				
		VIEWS	MEWS	EWS	NEWS	NEWS2	SOS	Worthing	HOTEL	TREWS	HEWS		Mortality	ICU	CA	RA	Sepsis
Innocenti, 2018[65]	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Camm, 2018	✗	○	○	○	●	○	○	○	○	○	○	Sens & Spec	✓	✓	✗	✗	✗
Tirotta, 2017	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Pong, 2019	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Prabhakar, 2019	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Martino, 2018	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Vorwerk, 2009[19]	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Qin, 2017[51]	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Schmedding, 2019	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Albur, 2016[64]	✗	○	○	●	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Cildir, 2013[36]	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Chiew, 2019[54]	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Samsudin, 2018	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Chang, 2018	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Geier, 2013	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✓
Asiimwe, 2015	✗	○	●	○	○	○	○	○	○	○	○	Prognostic index	✓	✗	✗	✗	✗
Hung, 2017	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Garcea, 2006	✗	○	○	●	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Yoo, 2015	✗	○	●	○	○	○	○	○	○	○	○	OR	✓	✓	✗	✗	✗
Siddiqui, 2017[37]	✗	○	○	●	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗

Studies are ranked according to sample size from largest to smallest in each subgroup.
AUC, area under the curve; CA, cardiac arrest; EHR, electronic health records; EWS, early warning score; HEWS, Hamilton early warning score; HOTEL, Hypotension, Oxygen saturation, Temperature, ECG abnormality, Loss of independence score; ICU, intensive care unit; MEWS, modified early warning score; NEWS, national early warning score; RA, respiratory arrest; Sens & Spec, sensitivity and specificity; SOS, search out severity score.; TREWS, triage in emergency department early warning score; VIEWS, vitalpac early warning score; Worthing, worthing physiological scoring system.

Table 3 Study design and setting forincluded studies of predictive performance for early warning scores in clinical settings

		Settings				Study design				
Author, year	Country	ICU	ED	Surgical	Medical	Retrospective	Prospective	RCT	Case–control	Number of patients
Calvert, 2016	Israel	●	○	○	○	●	○	○	○	29 083
Awad, 2017	UK	●	○	○	○	●	○	○	○	11 722
Reini, 2012	Sweden	●	○	○	○	○	●	○	○	518
Chen, 2019	Taiwan	●	○	○	○	●	○	○	○	370
Baker, 2015	Tanzania	●	○	○	○	○	●	○	○	269
Gök, 2019	Turkey	●	○	○	○	●	○	○	○	250
Moseson, 2014	USA	●	○	○	○	○	●	○	○	227
Jo, 2013	South Korea	●	○	○	○	●	○	○	○	151
Kown, 2018	Korea	○	●	○	○	●	○	○	○	1 986 334
Usman, 2019	USA	○	●	○	○	●	○	○	○	115 734
Jang, 2019	Korea	○	●	○	○	●	○	○	○	56 368
Wei, 2019	China	○	●	○	○	●	○	○	○	39 977
Lee, 2019	Korea	○	●	○	○	●	○	○	○	27 173
Singer, 2017	USA	○	●	○	○	●	○	○	○	22 530
Eick, 2015	Germany	○	●	○	○	○	●	○	○	5730
Bulut, 2014[58]	Turkey	○	●	○	○	○	●	○	○	2000
Kivipuro, 2018	Finland	○	●	○	○	○	●	○	○	1354
Eckart, 2019[62]	USA	○	●	○	○	●	○	○	○	1303
Ho, 2013	Malaysia	○	●	○	○	●	○	○	○	1024
Skitch, 2018	Canada	○	●	○	○	●	○	○	○	845
Liu, 2014	Malaysia	○	●	○	○	○	●	○	○	702
Dundar, 2016[57]	Turkey	○	●	○	○	○	●	○	○	671
Yuan, 2018	China	○	●	○	○	○	●	○	○	621
Naidoo, 2014	South Africa	○	●	○	○	●	○	○	○	590
Liu, 2015	China	○	●	○	○	○	●	○	○	551
So, 2015	China	○	●	○	○	○	●	○	○	544
Dundar, 2019	Turkey	○	●	○	○	●	○	○	○	455
Lam, 2006	China	○	●	○	○	○	●	○	○	425
Xie, 2018	China	○	●	○	○	○	●	○	○	383
Cattermole, 2009	China	○	●	○	○	○	●	○	○	330
Heitz, 2010	USA	○	●	○	○	●	○	○	○	280

Continued

Table 3 Continued

Author, year	Country	Settings				Study design				Number of patients
		ICU	ED	Surgical	Medical	Retrospective	Prospective	RCT	Case–control	
Sirivilaithon, 2019	Thailand	○	●	○	○	○	○	○	●	250
Cattermole, 2014	China	○	●	○	○	○	●	○	○	230
Najafi, 2018	Iran	○	●	○	○	○	●	○	○	185
Bartkowiak, 2019[50]	USA	○	○	●	○	●	○	○	○	32 537
Kovacs, 2016	UK	○	○	●	●	●	○	○	○	20 626
Plate, 2018	The Netherlands	○	○	●	○	○	●	○	○	1782
Sarani, 2012	The Netherlands	○	○	●	○	○	●	○	○	572
Hollis, 2016	USA	○	○	●	○	●	○	○	○	522
Gardner-Thorpe 2006	UK	○	○	●	○	○	●	○	○	334
Garcea, 2010	UK	○	○	●	○	●	○	○	○	280
Cuthbertson, 2007[49]	UK	○	○	●	○	●	○	○	○	136
Prytherch, 2010	UK	○	○	○	●	●	○	○	○	35 585
Smith, 2013[22]	UK	○	○	○	●	●	○	○	○	35 585
Rasmussen, 2018	Denmark	○	○	○	●	●	○	○	○	17 312
Ghosh, 2018	USA	○	○	○	●	●	○	○	○	2097
Duckitt, 2007[47]	UK	○	○	○	●	○	●	○	○	1102
Colombo, 2017	Italy	○	○	○	●	●	○	○	○	471
Abbot, 2016[48]	UK	○	○	○	●	○	●	○	○	322
Wheeler, 2013	Malawi	○	○	○	●	○	●	○	○	302
Graziadio, 2019	UK	○	○	○	●	○	●	○	○	292

Studies are ranked according to sample size from largest to smallest in each subgroup.
Black dots in the subgroup column represent the disease or the settings where the sample was studied, and brown dots in the study by Kellet represent different samples for each subgroup.
CVD, cardiovascular disease; ED, emergency department; GI, gastrointestinal diseases; ICU, intensive care unit.

Table 4 Early warning scores, predictive measures and outcomes for included studies of predictive performance for early warning scores in clinical settings

Author, year	EHR	EWS										Predictive measure	Outcomes studied				
		VIEWS	MEWS	EWS	NEWS	NEWS2	SOS	Worthing	HOTEL	TREWS	HEWS		Mortality	ICU	CA	RA	Sepsis
Calvert, 2016	✗	○	●	○	○	○	○	○	○	○	○	AUC	✗	✗	✗	✗	✓
Awad, 2017	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Reini, 2012	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Chen, 2019	✗	○	○	○	●	○	○	○	○	○	○	AUC	✗	✗	✗	✓	✗
Baker, 2015	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Gök, 2019	✗	○	●	○	○	○	○	○	○	○	○	AUC	✗	✗	✗	✗	✓
Moseson, 2014	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Jo, 2013	✗	●	●	○	○	○	○	○	●	○	○	AUC	✓	✗	✗	✗	✗
Kown, 2018	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Usman, 2019	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✓
Jang, 2019	✗	○	●	○	○	○	○	○	○	○	○	AUC	✗	✗	✓	✗	✗
Wei, 2019	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Lee, 2019	✗	○	●	○	●	○	○	○	○	●	○	AUC	✓	✗	✗	✗	✗
Singer, 2017	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Eick, 2015	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Bulut, 2014[58]	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Kivipuro, 2018	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Eckart, 2019[62]	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
, 2013	✗	○	●	○	○	○	○	○	○	○	○	AUC	✗	✓	✗	✗	✗
Skitch, 2018	✗	○	○	○	●	○	○	○	○	○	●	AUC	✗	✗	✗	✗	✓
Liu, 2014	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✓	✗	✗
Dundar, 2016[57]	✗	●	●	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Yuan, 2018	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Naidoo, 2014	✗	○	○	○	○	○	○	○	○	●	○	Sens & Spec	✓	✗	✗	✗	✗
Liu, 2015	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
So, 2015	✗	○	●	○	○	○	○	○	○	○	○	Sens & Spec	✓	✗	✗	✗	✗
Dundar, 2019	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Lam, 2006	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Xie, 2018	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Cattermole, 2009	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Heitz, 2010	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Sirivilaithon, 2019	✗	○	○	○	●	○	○	○	○	○	○	AUC	✗	✗	✗	✗	✗
Cattermole, 2014	✗	○	●	○	●	○	○	●	○	○	○	AUC	✓	✗	✗	✗	✗

Continued

Table 4 Continued

Author, year	EHR	EWS										Predictive measure	Outcomes studied				Sepsis
		VIEWS	MEWS	EWS	NEWS	NEWS2	SOS	Worthing	HOTEL	TREWS	HEWS		Mortality	ICU	CA	RA	
Najafi, 2018	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Bartkowiak, 2019[50]	✗	○	●	○	●	○	○	○	○	○	○	AUC	✓	✓	✓	✗	✗
Kovacs, 2016	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✓	✓	✗	✗
Plate, 2018	✗	●	○	○	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Sarani, 2012	✗	○	●	○	○	○	○	○	○	○	○	Sens & Spec	✓	✓	✗	✗	✗
Hollis, 2016	✗	○	○	●	○	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗
Gardner-Thorpe, 2006	✗	○	●	○	○	○	○	○	○	○	○	Sens & Spec	✓	✓	✗	✗	✗
Garcea *et al*, 2010	✗	○	○	●	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Cuthbertson, 2007[49]	✗	○	●	●	○	○	○	○	○	○	○	AUC	✗	✓	✗	✗	✗
Prytherch, 2010	✗	●	○	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Smith, 2013[22]	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✓	✓	✗	✗
Rasmussen, 2018	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Ghosh, 2018	✓	○	●	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Duckitt, 2007[47]	✗	○	○	●	○	○	○	●	○	○	○	AUC	✓	✓	✗	✗	✗
Colombo, 2017	✗	○	●	○	○	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Abbot, 2016[48]	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✗	✗	✗	✗
Wheeler, 2013	✗	○	●	○	○	○	○	○	●	○	○	AUC	✓	✗	✗	✗	✗
Graziadio, 2019	✗	○	○	○	●	○	○	○	○	○	○	AUC	✓	✓	✗	✗	✗

Studies are ranked according to sample size from largest to smallest in each subgroup.
AUC, area under the curve; CA, cardiac arrest; EHR, electronic health records; EWS, early warning score; HEWS, Hamilton early warning score; HOTEL, Hypotension, Oxygen saturation, Temperature, ECG abnormality, Loss of independence score; MEWS, modified early warning score; NEWS, national early warning score; RA, respiratory arrest; Sens and Spec, sensitivity and specificity; SOS, search out severity score.; TREWS, triage in emergency department early warning score; VIEWS, vitalpac early warning score; Worthing, worthing physiological scoring system.

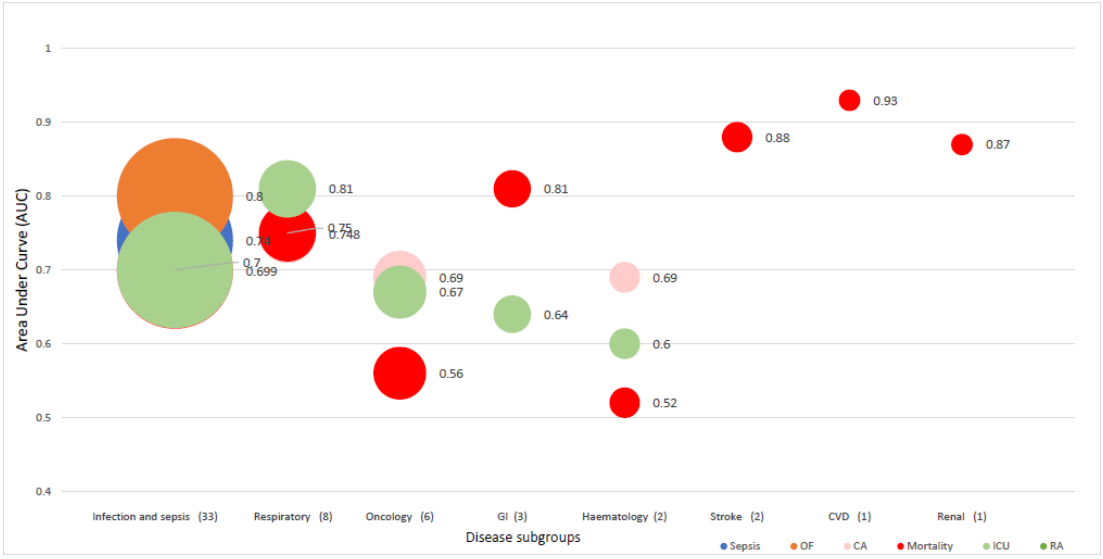

Figure 3 Early warning score performance in different disease subgroups. Each bubble represents critical events predicted by early warning scores for each disease subgroup with average AUC of studies beside each event type. The size of the bubble represents the number of studies in each subgroup. CA, cardiac arrest; CVD, cardiovascular diseases; GI, gastrointestinal diseases; ICU, intensive care unit; OF, organ failure; RA, respiratory arrest.

In methodology, observations selections method, time horizon between EWS score and event and the metric used in assessment were inconsistent. Choosing multiple observations or a single observation prior the outcome may not significantly affect the ranking of EWS.[67] Yet, selecting a single observation is generally associated with high AUC compared with multiple observations,[46 67] supporting the use of multiple observations for each episode. Moreover, AUC, the most commonly used measure of predictive performance, has limitations and other metrics, including positive predictive value, should also be assessed.[68] Recording observations at an agreed threshold point before events in a standardised method is necessary to evaluate EWS effectively.

The universal EWS with standardised models were primarily designed for general patient populations in wards and EDs and remain underevaluated in specific diseases and settings. In medical and ED contexts, EWS perform well, suggesting the role of EWS in general settings, or at the early stage of clinical assessment. Our positive findings in respiratory disease may indicate the emphasis of several EWS, such as NEWS2, on respiratory changes when patients are deteriorating. Specific disease areas may show unique alarm signs when critical events

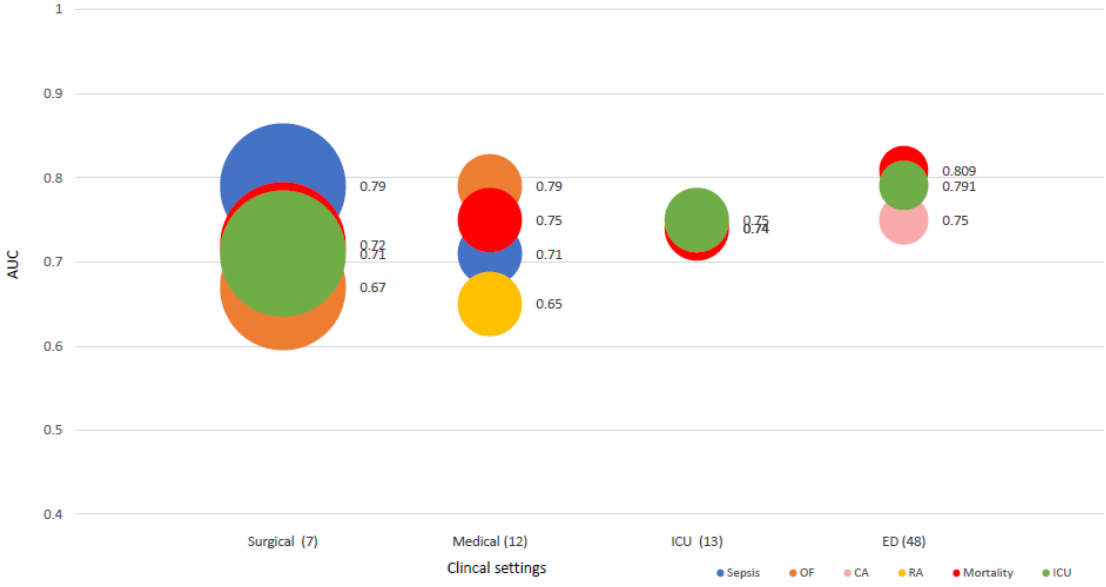

Figure 4 Early warning score performance in different clinical settings. Each bubble represents critical events predicted by early warning scores for each disease subgroup with average AUC of studies beside each event type. The size of the bubble represents the number of studies in each subgroup. CA, cardiac arrest; ED, emergency department; ICU, intensive care units; OF, organ failure; RA, respiratory arrest.

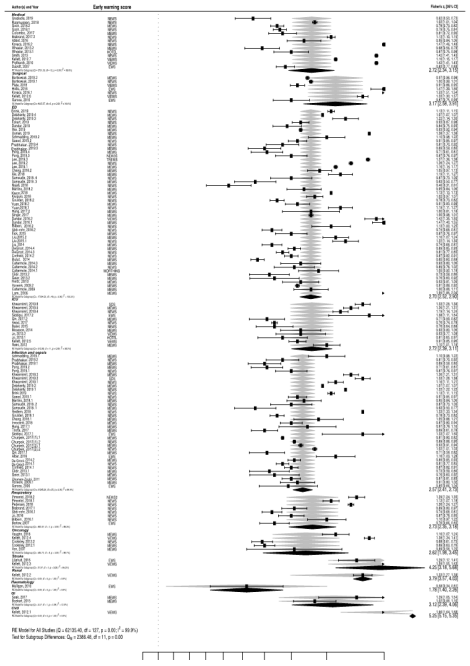

Figure 5 Forest plot of predictive accuracy of universal early warning scores (EWS) for mortality in different disease subgroups and clinical settings. CVD, cardiovascular diseases; ED, emergency department; GI, gastrointestinal diseases; Hem, haematological diseases; ICU, intensive care units; Infec, infectious diseases; Med, medical settings; Onco, oncology diseases; Renal, renal diseases; renal diseases; Resp, respiratory diseases; stroke, patients who had a stroke; Surg, surgical settings . Note: number following author(s) and year indicate more than one EWS evaluated in the study.

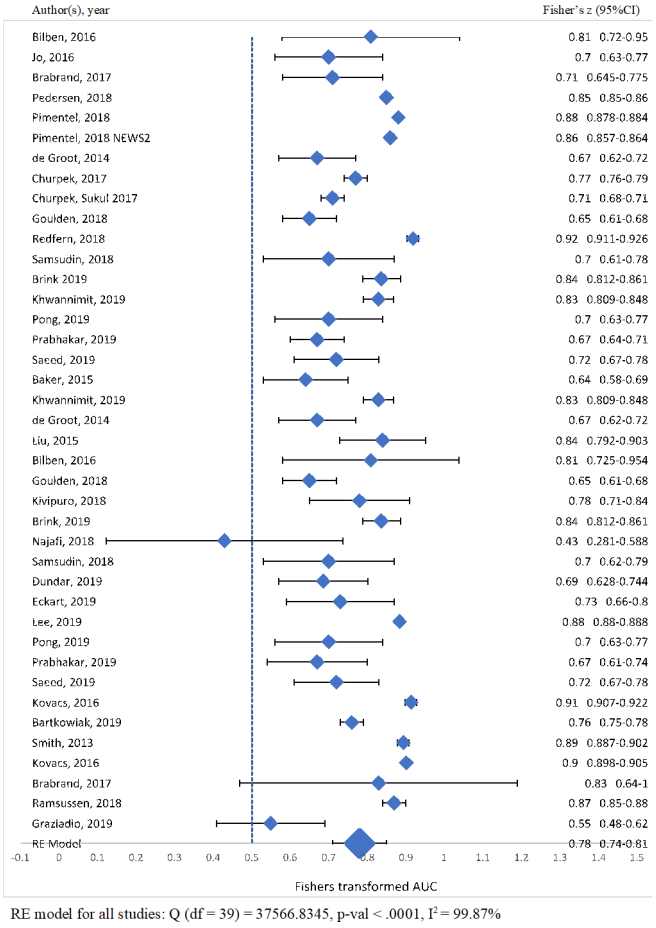

Figure 6 Forest plot of predictive accuracy of NEWS for mortality. AUC, area under the curve; NEWS, national early warning score.

are anticipated, which may not be captured by universal EWS, such as NEWS2, where prediction of deterioration is based on predefined thresholds in all patients.[23] Critical events are commonly associated with CVD. With CVD being a leading cause of mortality globally, and the significant impact of morbidity on health and social care, early detection of deterioration is necessary.[69] However, EWS are poorly validated in CVD, some of the parameters may not be applicable and EWS may be unrepresentative.[25] A recent study of NEWS2 in patients with COVID-19 infection found poor performance in severity prediction,[70] despite pre-existing conditions being common and predictive in patients with severe outcomes. EWS may need to take account of disease-specific risk factors and comorbidities.

Widespread uptake of EHR and digitisation of patient observations are expected to contribute to efficient use of EWS by reducing human errors in documentation and calculation, as well as delays in escalation of care. However, relatively few studies have considered EHR-based EWS, and those studies have not analysed whether predictive performance of EWS is related to EHR use, diseases or settings. Investigating implementation and adoption of EWS is necessary to understand the application and performance of EWS. Predictive algorithms derived by machine learning have been successfully used in developing and validating EWS[41 71] but will require robust evaluation. Studying the implementation process of EWS within EHR will provide opportunities for qualitative and quantitative insights into escalation of care, as well as facilitators and barriers to use of EWS in routine practice.

There are several limitations in this review and in included studies. We aimed for a comprehensive investigation of all EWS developing since 1997, but this long study period may lead to bias in comparing studies with old and new validation approaches statistically and technically. We excluded EWS specifically derived and validated for particular disease populations or settings and excluded studies considering a general patient population. Meta-analysis was only done for studies using AUC, excluding other methods for assessing performance of EWS. The distinction between general patient settings and specific disease or patient subgroups is dependent on hospital, healthcare system and country, and there is inevitably overlap between patients and settings at different stages in patient pathways. It was only feasible to include studies with a clear disease or setting identified to avoid confusion.

Validation of EWS in disease subgroups should consider similarities and differences across diseases, sample size

and include measures of model discrimination and calibration. Further research should adhere to established guidelines on clinical outcomes and predictive clinical scoring for decision making, such as the PROGRESS framework.[72]

CONCLUSION

Universal EWS in specific disease subgroups and settings require further validation of their performance in detecting worsening outcomes. Despite good performance in respiratory patients and medical and surgical settings in studies to date, the predictive accuracy of EWS in all disease subgroups and all clinical settings remains unknown. The current evidence base does not necessarily support use of standard EWS in all patients in all settings. Future research should include validation of EWS in particular patient subgroups and settings, with standardised methodology following established guidelines. Going towards the utilisation of EHR for EWS development, validation and implementation within EHR should be considered for improved EWS systems.

Twitter Baneen Alhmoud @BaneenAlhmoud, Riyaz Patel @DrRiyazPatel and Amitava Banerjee @amibanerjee1

Contributors AB conceived the study. BA, AB and TB conducted the search, data extraction and data analysis. BA wrote the initial draft of the manuscript. All authors contributed to interpretation of findings, critical review and revisions of the manuscript.

Funding BA has received PhD funding from the Saudi Arabian Cultural Bureau.

Competing interests AB has received research grants from AstraZeneca.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs

Riyaz Patel http://orcid.org/0000-0003-4603-2393
Bryan Williams http://orcid.org/0000-0002-8094-1841
Amitava Banerjee http://orcid.org/0000-0001-8741-3411

REFERENCES

1 Cetınkaya HB, Koksal O, Sigirli D, *et al*. The predictive value of the modified early warning score with rapid lactate level (ViEWS-L) for mortality in patients of age 65 or older visiting the emergency department. *Intern Emerg Med* 2017;12:1253–7.
2 Cei M, Bartolomei C, Mumoli N. In-hospital mortality and morbidity of elderly medical patients can be predicted at admission by the modified early warning score: a prospective study. *Int J Clin Pract* 2009;63:591–5.
3 Alam N, Hobbelink EL, van Tienhoven AJ, *et al*. The impact of the use of the early warning score (EWS) on patient outcomes: a systematic review. *Resuscitation* 2014;85:587–94.
4 Hogan H, Hutchings A, Wulff J, *et al*. Interventions to reduce mortality from in-hospital cardiac arrest: a mixed-methods study. *Health Serv Deliv Res* 2019;7:1–110.
5 Adhikari NKJ, Fowler RA, Bhagwanjee S, *et al*. Critical care and the global burden of critical illness in adults. *Lancet* 2010;376:1339–46.
6 Hogan H, Healey F, Neale G, *et al*. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual Saf* 2012;21:737–45.
7 De Meester K, Das T, Hellemans K, *et al*. Impact of a standardized nurse observation protocol including MEWS after intensive care unit discharge. *Resuscitation* 2013;84:184–8.
8 Paterson R, MacLeod DC, Thetford D, *et al*. Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit. *Clin Med* 2006;6:281–4.
9 Moon A, Cosgrove JF, Lea D, *et al*. An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR. *Resuscitation* 2011;82:150–4.
10 Kause J, Smith G, Prytherch D, *et al*. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom--the ACADEMIA study. *Resuscitation* 2004;62:275–82.
11 Hillman KM, Bristow PJ, Chey T, *et al*. Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med* 2002;28:1629–34.
12 Wilkinson K, Martin IC, Gough MJ. National confidential enquiry into patient outcome and death. An age old problem. In: *A review of the care received by elderly patients undergoing surgery*. London: NCEPOD, 2011.
13 Morgan RJM, Williams F, Wright MM. An early warning scoring system for detecting developing critical illness. *Clin Intensive Care* 1997;8:100.
14 Linnen DT, Escobar GJ, Hu X, *et al*. Statistical modeling and aggregate-weighted scoring systems in prediction of mortality and ICU transfer: a systematic review. *J Hosp Med* 2019;14:161–9.
15 Hamilton F, Arnold D, Baird A, *et al*. Early warning scores do not accurately predict mortality in sepsis: a meta-analysis and systematic review of the literature. *J Infect* 2018;76:241–8.
16 Wuytack F, Meskell P, Conway A, *et al*. The effectiveness of physiologically based early warning or track and trigger systems after triage in adult patients presenting to emergency departments: a systematic review. *BMC Emerg Med* 2017;17:38.
17 Plevin R, Callcut R. Update in sepsis guidelines: what is really new? *Trauma Surg Acute Care Open* 2017;2:e000088.
18 Mohammed MA, Rudge G, Watson D, *et al*. Index blood tests and national early warning scores within 24 hours of emergency admission can predict the risk of in-hospital mortality: a model development and validation study. *PLoS One* 2013;8:e64340.
19 Vorwerk C, Loryman B, Coats TJ, *et al*. Prediction of mortality in adult emergency department patients with sepsis. *Emerg Med J* 2009;26:254–8.
20 Royal College of Physicians of London. National early warning score (news): standardising the assessment of acute-illness severity in the NHS. *R Coll Physician* 2012.
21 Royal College of Physicians of London. Nhs England approves use of national early warning score (news) 2 to improve detection of acutely ill patients. *R Coll Physician* 2017.
22 Smith GB, Prytherch DR, Meredith P, *et al*. 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;84:465–70.
23 Inada-Kim M, Nsutebu E. News 2: an opportunity to standardise the management of deterioration and sepsis. *BMJ* 2018;360:k1260.
24 Direkze S, Jain S. Time to intervene? lessons from the NCEPOD cardiopulmonary resuscitation report 2012. *Br J Hosp Med* 2012;73:585–7.
25 Badreldin AMA, Doerr F, Bender EM, *et al*. Rapid clinical evaluation: an early warning cardiac surgical scoring system for hand-held digital devices. *Eur J Cardiothorac Surg* 2013;44:992–8.
26 Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453–73.
27 Debray TPA, Vergouwe Y, Koffijberg H, *et al*. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;68:279–89.

28 Smith MEB, Chiovaro JC, O'Neil M, O'Neil M, *et al*. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. *Ann Am Thorac Soc* 2014;11:1454–65.
29 Williams TA, Tohira H, Finn J, *et al*. The ability of early warning scores (EWS) to detect critical illness in the prehospital setting: a systematic review. *Resuscitation* 2016;102:35–43.
30 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.
31 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
32 Moons KGM, de Groot JAH, Bouwmeester W, *et al*. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the charms checklist. *PLoS Med* 2014;11:e1001744.
33 Wolff RF, Moons KGM, Riley RD, *et al*. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51.
34 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:e0211133.
35 Churpek MM, Snyder A, Sokol S, *et al*. Investigating the impact of different suspicion of infection criteria on the accuracy of quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores. *Crit Care Med* 2017;45:1805–12.
36 Çıldır E, Bulut M, Akalın H, *et al*. Evaluation of the modified MEDS, MEWS score and Charlson comorbidity index in patients with community acquired sepsis in the emergency department. *Intern Emerg Med* 2013;8:255–60.
37 Siddiqui S, Chua M, Kumaresh V, *et al*. A comparison of pre ICU admission SIRS, EWS and Q SOFA scores for predicting mortality and length of stay in ICU. *J Crit Care* 2017;41:191–3.
38 Khwannimit B, Bhurayanontachai R, Vattanavanit V. Comparison of the accuracy of three early warning scores with SOFA score for predicting mortality in adult sepsis and septic shock patients admitted to intensive care unit. *Heart Lung* 2019;48:240–4.
39 Vaughn JL, Kline D, Denlinger NM, *et al*. Predictive performance of early warning scores in acute leukemia patients receiving induction chemotherapy. *Leuk Lymphoma* 2018;59:1498–500.
40 Henry KE, Hager DN, Pronovost PJ, *et al*. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015;7:299ra122–299.
41 Pimentel MAF, Redfern OC, Gerry S, *et al*. A comparison of the ability of the National early warning score and the National early warning score 2 to identify patients at risk of in-hospital mortality: a multi-centre database study. *Resuscitation* 2019;134:147–56.
42 Churpek MM, Snyder A, Han X, *et al*. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. *Am J Respir Crit Care Med* 2017;195:906–11.
43 de Groot B, Stolwijk F, Warmerdam M, *et al*. The most commonly used disease severity scores are inappropriate for risk stratification of older emergency department sepsis patients: an observational multi-centre study. *Scand J Trauma Resusc Emerg Med* 2017;25:91.
44 Smith GB, Prytherch DR, Schmidt PE, *et al*. Review and performance evaluation of aggregate weighted 'track and trigger' systems. *Resuscitation* 2008;77:170–9.
45 Smith GB, Prytherch DR, Schmidt PE, *et al*. A review, and performance evaluation, of single-parameter "track and trigger" systems. *Resuscitation* 2008;79:11–21.
46 Kellett J, Kim A. Validation of an abbreviated Vitalpac™ early warning score (views) in 75,419 consecutive admissions to a Canadian regional hospital. *Resuscitation* 2012;83:297–302.
47 Duckitt RW, Buxton-Thomas R, Walker J, *et al*. Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study. *Br J Anaesth* 2007;98:769–74.
48 Abbott TEF, Torrance HDT, Cron N, *et al*. A single-centre cohort study of national early warning score (news) and near patient testing in acute medical admissions. *Eur J Intern Med* 2016;35:78–82.
49 Cuthbertson BH, Boroujerdi M, McKie L, *et al*. Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient? *Crit Care Med* 2007;35:402–9.
50 Bartkowiak B, Snyder AM, Benjamin A, *et al*. Validating the electronic cardiac arrest risk triage (eCART) score for risk stratification of surgical inpatients in the postoperative setting: retrospective cohort study. *Ann Surg* 2019;269:1059–63.
51 Qin Q, Xia Y, Cao Y. Clinical study of a new modified early warning system scoring system for rapidly evaluating shock in adults. *J Crit Care* 2017;37:50–5.
52 Sbiti-Rohr D, Kutz A, Christ-Crain M, *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:e011021.
53 Liljehult J, Christensen T. Early warning score predicts acute mortality in stroke patients. *Acta Neurol Scand* 2016;133:261–7.
54 Chiew CJ, Liu N, Tagami T, *et al*. Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. *Medicine* 2019;98:e14197.
55 Bilben B, Grandal L, Søvik S. National Early Warning Score (NEWS) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient - a prospective observational study. *Scand J Trauma Resusc Emerg Med* 2016;24:80.
56 Goulden R, Hoyle M-C, Monis J, *et al*. qSOFA, SIRS and news for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. *Emerg Med J* 2018;35:345–9.
57 Dundar ZD, Ergin M, Karamercan MA, *et al*. Modified early warning score and VitalPac early warning score in geriatric patients admitted to emergency department. *Eur J Emerg Med* 2016;23:406–12.
58 Bulut M, Cebicci H, Sigirli D, *et al*. The comparison of modified early warning score with rapid emergency medicine score: a prospective multicentre observational cohort study on medical and surgical patients presenting to emergency department. *Emerg Med J* 2014;31:476–81.
59 Smith GB, Prytherch DR, Schmidt PE, *et al*. Review and performance evaluation of aggregate weighted 'track and trigger' system. *Resuscitation* 2008;77:170–9.
60 Mulligan A. Validation of a physiological track and trigger score to identify developing critical illness in haematology patients. *Intensive Crit Care Nurs* 2010;26:196–206.
61 Cooksley T, Kitlowski E, Haji-Michael P. Effectiveness of modified early warning score in predicting outcomes in oncology patients. *QJM* 2012;105:1083–8.
62 Eckart A, Hauser SI, Kutz A, *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:e024636.
63 Ghanem-Zoubi NO, Vardi M, Laor A, *et al*. Assessment of disease-severity scoring systems for patients with sepsis in general internal medicine departments. *Crit Care* 2011;15:R95.
64 Albur M, Hamilton F, MacGowan AP. Early warning score: a dynamic marker of severity and prognosis in patients with gram-negative bacteraemia and sepsis. *Ann Clin Microbiol Antimicrob* 2016;15:23.
65 Innocenti F, Tozzi C, Donnini C, *et al*. SOFA score in septic patients: incremental prognostic value over age, comorbidities, and parameters of sepsis severity. *Intern Emerg Med* 2018;13:405–12.
66 Hu SB, Wong DJL, Correa A, *et al*. Prediction of clinical deterioration in hospitalized adult patients with hematologic malignancies using a neural network model. *PLoS One* 2016;11:e0161401.
67 Jarvis SW, Kovacs C, Briggs J, *et al*. Are observation selection methods important when comparing early warning score performance? *Resuscitation* 2015;90:1–6.
68 Romero-Brufau S, Huddleston JM, Naessens JM, *et al*. Widely used track and trigger scores: are they ready for automation in practice? *Resuscitation* 2014;85:549–52.
69 Mozaffarian D. Global scourge of cardiovascular disease. *J Am Coll Cardiol* 2017;70:26–8.
70 Carr E, Bendayan R, Bean D. Supplementing the National early warning score (NEWS2) for Anticipating early deterioration among patients with COVID-19 infection. *medRxiv* 2020.
71 Churpek MM, Yuen TC, Park SY, *et al*. Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*. *Crit Care Med* 2014;42:841–8.
72 Hemingway H, Croft P, Perel P, *et al*. Prognosis research strategy (progress) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
73 Van Calster B, McLernon DJ, van Smeden M, *et al*. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;17:230.