# 

# Vital Sign Charts & Early Warning Systems

# A Scoping Literature Review

May 2024

Report prepared for

A logo with black and white text

Description automatically generated

Contents

[Abbreviations 3](#_Toc166762254)

[Executive Summary 5](#_Toc166762255)

[Introduction 6](#_Toc166762256)

[Aim and Scope 7](#_Toc166762257)

[Methods 8](#_Toc166762258)

[Studies of Adult Inpatient EWS/Vital Sign Charts 10](#_Toc166762259)

[Review Articles 10](#_Toc166762260)

[Individual Studies 16](#_Toc166762261)

[Additional Studies 25](#_Toc166762262)

[Studies of Maternity Inpatient EWS/Vital Sign Charts 29](#_Toc166762263)

[Background: Maternity EWS Tools 29](#_Toc166762264)

[Review Articles 30](#_Toc166762265)

[Head-to-head Studies of Maternity EWS 32](#_Toc166762266)

[Studies of Physiology in Pregnancy 34](#_Toc166762267)

[Additional Studies 35](#_Toc166762268)

[Maternity EWS Test Performance 37](#_Toc166762269)

[Design of EWS Vital Sign Charts 39](#_Toc166762270)

[Discussion 41](#_Toc166762271)

[Adult 41](#_Toc166762272)

[Maternity 46](#_Toc166762273)

[Recommendations 50](#_Toc166762274)

[Answers to the 5 Research Questions 50](#_Toc166762275)

[Additional Recommendations 51](#_Toc166762276)

[Limitations of this Study 52](#_Toc166762277)

[References 53](#_Toc166762278)

[Appendix: Core Papers of the Review 58](#_Toc166762279)

[Table A1: Adult (Reviews) 58](#_Toc166762280)

[Table A2: Adult (Other Studies) 62](#_Toc166762281)

[Table A3: Maternity (Reviews) 68](#_Toc166762282)

[Table A4: Maternity (Other Studies) 70](#_Toc166762283)

## Abbreviations

10-SOV Ten signs of vitality

AAM Advanced alert monitor system

AI Artificial intelligence

ASEWS Age-specific early warning score

AUC (or AUROC) Area under the receiver operating characteristic curve (a measure of discrimination ability, with 1 being perfect discrimination)

AVPU Alert, voice, pain, unresponsive (neurological response assessment)

BTF Between the flags (NSW early warning system)

CDC Centers for Disease Control and Prevention

CI Confidence interval

CPR Cardiopulmonary resuscitation

CRT Capillary refill time

DBP Diastolic blood pressure

eCART Electronic Cardiac Arrest Triage Score

ED Emergency department

EHR Electronic health record

EWS Early warning system(s)

FiO2 Fraction of inspired oxygen

HR Heart rate

ICD-10 International Classification of Disease (10th Edition)

ICU Intensive care unit

IHCA In-hospital cardiac arrest

LoC Level of consciousness

LR Likelihood ratio

MCEWS Manual centile based early warning score

MET Medical emergency team

MEOWS Modified Early Obstetric Warning Score

MERC Modified Early Obstetric Warning Criteria

MEWC Modified early warning criteria

MEWS Modified Early Warning System[[1]](#footnote-2)

MEWT Modified Early Warning Tool

ML Machine learning

NEWS National Early Warning Score

NEWS2 National Early Warning Score (updated version)

NHS National Health Service (UK)

NICE National Institute for Health and Care Excellence (UK)

NNE Number needed to evaluate

NPV Negative predictive value

NSW New South Wales

NZEWS New Zealand Early Warning System

NZMEWS New Zealand Maternity Early Warning System

ObsEWS Obstetric early warning score

OEWS Obstetric early warning system

OR Odds ratio

PPV Positive predictive value

Q-ADDS Queensland Adult Deterioration Detection System

qSOFA Quick sequential organ failure assessment

RCP Royal College of Physicians

RCT Randomised controlled trial

ROC Receiver operating characteristic curve

RR Respiratory rate

RRS Rapid response system

RRT Rapid response team

SBAR Situation, background, assessment, recommendation (communication tool)

SBP Systolic blood pressure

SD Standard deviation

SIRS Systemic inflammatory response syndrome criteria

SOB Shortness of breath

SpO2 (or O2 sat) Blood oxygen saturation

T2RF Type-II respiratory failure

TRIPOD Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

TTS Track-and-trigger systems

UK United Kingdom

US United States

ViEWS VitalPAC Early Warning Score

## Executive Summary

A considerable challenge for adult and maternity inpatient care providers is recognising clinical deterioration early. Physiological track-and-trigger systems (TTS) or early warning systems (EWS), are one means that may help care providers recognise actual or potential clinical deterioration early.

Te Tāhū Hauora Health Quality & Safety Commission (Te Tāhū Hauora) New Zealand undertook a national patient deterioration programme between 2016 and 2021, including testing and roll out of adult and maternity vital sign charts. Now Te Tāhū Hauora is undertaking a review of the national adult and maternity vital signs charts to ensure they remain aligned to best practice. This report presents the findings of a scoping literature review to inform the wider review. Findings are based on 42 ‘adult’ and 15 ‘maternity’ studies included following searches and screening for relevance.

Results indicate that EWS can perform well in the general adult inpatient population, across most diagnostic groups, particularly for predicting mortality in the shorter term (<24 hours). Aggregate weighted EWS appear to perform better than activation criteria tools. The NEWS tool is widely studied and has been found in review studies to consistently outperform other manual or paper based EWS. However, electronic EWS and those employing advanced analytics such as AI and ML appear more promising.

Outcome benefits of EWS included associations with reduced in-hospital cardiac arrests and in-hospital mortality as well as evidence of a dose-response relationship (with NEWS). However, there is a lack of high-quality evidence and little data on the net benefit of EWS. The New Zealand national vital sign EWS chart incorporates key features of NEWS and had comparable performance to NEWS in a multi-site study.

The picture in maternity is less clear due to fewer head-to-head comparisons of EWS tools, fewer external validation studies, and fewer consistently reported settings, populations, and outcomes of interest. Therefore, inferring the optimal EWS strategy for maternity/obstetric populations is more challenging. However, some tools are associated with improved maternal outcomes and can probably be recommended. Tool choice will depend on the goal of implementation, and therefore test characteristics desired. However, the review found that general electronic tools such as eCART outperformed maternity specific manual paper-based tools at certain thresholds when compared head-to-head.

The review identified variables additional to those used in NZEWS and NZMEWS, which might enhance prediction of deterioration in adult and in maternity populations. These are catalogued in the ‘Discussion’ below. Any amendments to the NZEWS or NZMEWS would require re-validation study, ideally head-to-head with the existing tool and other high performing EWS. Net benefit or cost-effectiveness analysis at different activation thresholds would also be useful for guiding practice.

In the report below, description of the individual studies is followed by ‘Discussion’ and ‘Recommendation’ sections, which identify the relevant underlying evidence base and provide details allowing the reader to weigh the various considerations guiding EWS choice, inclusion of variables, activation thresholds, and chart design.

## Introduction

A considerable challenge for adult and maternity inpatient care providers is recognising clinical deterioration early. Physiological track-and-trigger systems (TTS) or early warning systems (EWS), are one means that may help care providers recognise actual or potential clinical deterioration early. TTS/EWS are clinician-administered bedside physiological assessment protocols, charts or tools designed to record routinely assessed clinical parameters. Common parameters recorded include respiratory rate, blood pressure, temperature, heart rate, urine output and mental/neurological alertness. Many systems involve the application of scores or alert indicators to the observed physiological parameters based on prespecified limits. The overall system score or alert limit is then used to assist the care provider identify a need to escalate care. This, in turn, may allow for earlier intervention(s) to alter the course of the emerging critical illness and ultimately reduce or avoid mortality and morbidity sequelae.

Te Tāhū Hauora Health Quality & Safety Commission (Te Tāhū Hauora) New Zealand undertook a national patient deterioration programme between 2016 and 2021. The national adult early warning system (EWS) including the adult vital signs charts was tested in six hospital sites between November 2016-May 2017 and implemented across Aotearoa New Zealand in 2018 as the first stage of the five-year patient deterioration programme. This was followed by the development and testing of a maternity vital signs chart and supporting documents across three maternity hospitals in Aotearoa New Zealand, led by the then Maternal Morbidity Working Group. An evaluation and finalisation of the national maternity early warning system (NZMEWS) and implementation by Te Tāhū Hauora occurred in 2020 throughout maternity services, with hospital wide implementation continuing into 2021. The NZMEWS includes standardised maternity vital sign charts that should be used for those pregnant and recently pregnant anywhere in hospital, not only within maternity services.

The NZEWS and NZMEWS both combine aggregate scoring (derived from NEWS) and single parameter escalation (used in ‘Medical Emergency Team’ calling systems common in Australia and Aotearoa New Zealand). The aggregate scores facilitate a graded response to patient deterioration supplemented by single extreme vital sign derangement mandating rapid escalation. The vital sign charts are only one part of a wider system that includes an effective escalation pathway, clinical governance, education and ongoing measurement. Following the implementation of NZEWS/NZMEWS across the motu, Te Tāhū Hauora is undertaking a review of the national adult and maternity vital signs charts to ensure they remain aligned to best practice.

The purpose of this literature review was to determine the latest evidence for the processes and clinical parameters that should be included in vital sign charts that form part of early warning systems. Simultaneously reviewing the literature for both pregnant and non-pregnant adults will standardise any necessary changes. Findings of this literature review will inform any recommendations for changes to the vital sign charts.

The literature review was guided by the following five questions:

1. What vital sign parameters are validated to predict acute deterioration (death, cardiac arrest or admission to a higher level of care) in adult hospital inpatients?
2. What vital sign parameters are validated to predict acute deterioration (death, cardiac arrest or admission to a higher level of care) in hospital inpatients who are currently pregnant or post-partum?
3. What other clinical parameters should be documented on vital sign charts?
4. What are the features of good vital sign chart design?
5. What are the key elements of effective escalation pathways and response systems?

### Aim and Scope

The literature review was a scoping review with the aim of identifying academic literature, published since 2016, relevant to the five questions above. The focus was adult hospital inpatients and maternity inpatients.

Scoping reviews are used to survey the literature to establish a broad understanding of research on a topic. Usually, a five-stage methodology includes defining the questions, searching the literature, selecting studies, visually organizing the information, and summarizing the information.

Few studies examine single physiological variables as predictors of patient deterioration and guides to response, so it was expected that most literature would examine various multi-variable EWS including single variable threshold tools and aggregate weighted scoring tools.

To answer the questions about validity, it was important to seek studies not just of single tools in single cohorts, but research that compared various tools in the same cohort at the same time, and the same tool across different sites and populations. Therefore, less focus was given to single centre studies and single tool studies (unless comparison with other tools, or external validation) as these data should probably not change practice without further comparative evidence.

## Methods

The search was a time-limited search of PubMed, Cochrane Library, Google Scholar, and Elicit (multiple databases), using a combination of controlled vocabulary (eg, MeSH [Medical Subject Headings] terms), free text keywords, and natural language queries for the AI search tool Elicit.

Important search terms included: Recognition system, trigger, early warning score, early warning system, vital signs score, predictive system, predictive score, warning score, outcome, validity, validation, performance, composite endpoint, mortality, morbidity, cardiac arrest, ICU admission.

Several thousands of search results were identified in multiple searches and results were then restricted by publication date (after 2016), language (English only), and study type (meta-analysis, systematic review, review, randomised controlled trial, clinical trial, observation study, validation study). Inspection of titles was performed to include/exclude studies as follows:

Included:

* Primary literature/academic studies
* Of in-hospital vital sign recognition (trigger/early warning) systems (whether response was studied) used to recognise patient deterioration
* In developed & Western settings
* In general patient populations
  + Adult 16+, medical & surgical, inpatient
  + Maternity, inpatient
* That report on validation of the recognition systems
* Also included studies if they report vital sign parameters that predict patient deterioration
* Also included studies reporting on vital sign chart design and human factors

Excluded:

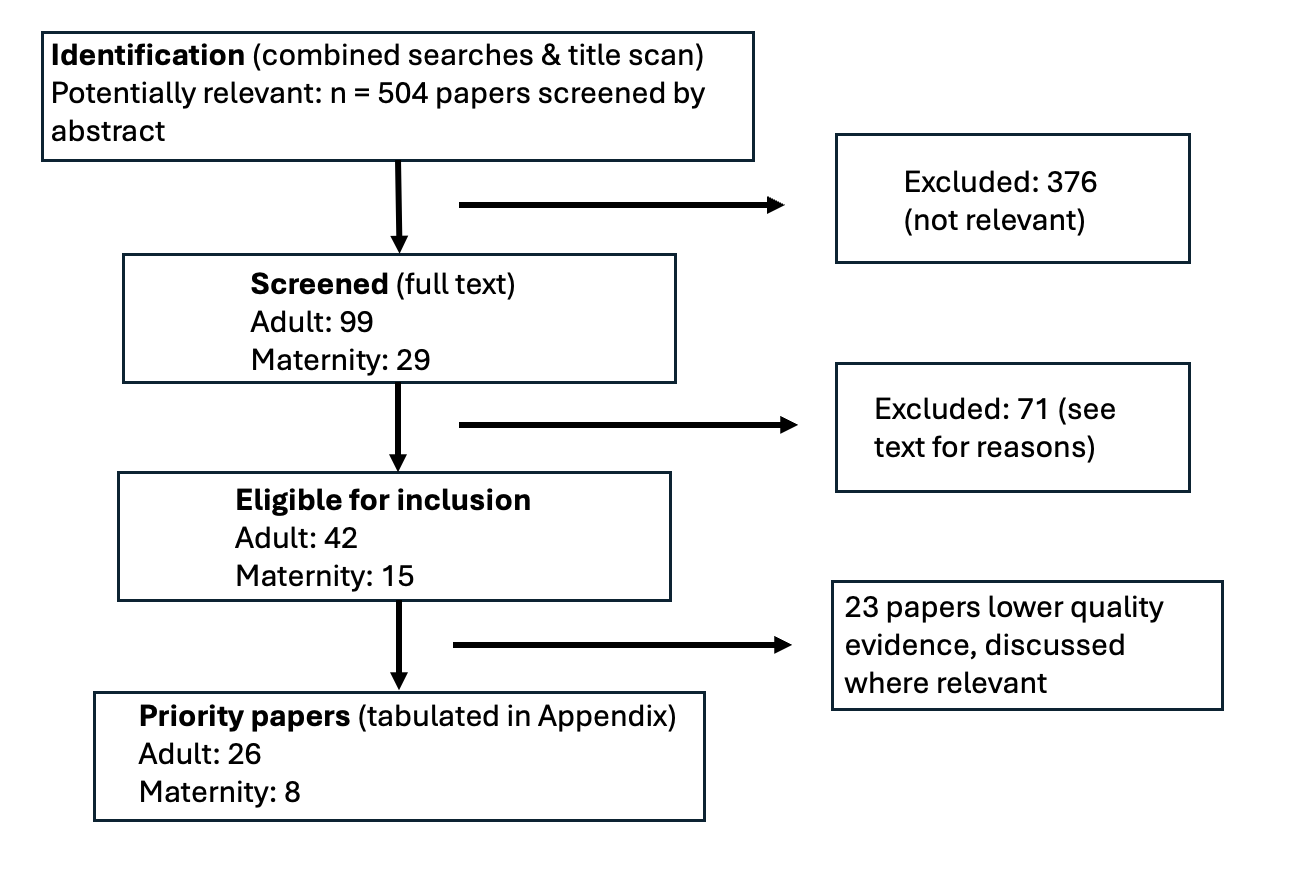
* Studies only of tools relying on advanced analytics, eg machine learning
* Studies of digital-only tools (without manual/paper-based comparison)
* Paediatric or neonatal early warning scores/systems
* Studies only in emergency departments
* Out of hospital studies (unless receiving acute hospital-level care at home), eg exclude primary care, aged residential care, or pre-hospital settings
* Disease specific tools or modifications (eg Covid-19; Sepsis)
* Studies of general EWS only in a specific patient group eg pancreatitis, haematological malignancy
* Interventions requiring wireless or continuous physiological monitoring
* Grey literature, eg NICE, ACSQHC
* Studies only of response (efferent arm) or response team structure
* Governance of EWS
* Studies focusing only on user experience
* Qualitative studies
* Studies of patient/family/whanau escalation processes
* Studies describing only education or knowledge

In total 504 results remained to be screened by abstract, and 99 ‘adult’ and 29 ‘maternity’ full-text papers were obtained.

On examination of full text another 57 ‘adult’ studies, and 14, ‘maternity’ studies were excluded for the following reasons: setting was non-Western, non-developed nation, study limited to specific condition eg sepsis, pneumonia, trauma, pancreatitis, or specific location eg ICU, ED only, qualitative study only, EWS including vital signs not studied, opinion article only, AI tool or automated EWS only, no validation of EWS, methodological study only, study protocol only, superseded literature review. The search process is illustrated in Figure 1.

The 42 ‘adult’ and 15 ‘maternity’ studies remaining comprised the review. Given the time-limited nature of the scoping review, these 57 studies were coded on a 1 to 5 scale estimating their quality and relevance. Studies coded ‘1’ were systematic reviews or meta-analyses, ‘2’ was for randomised studies or large population studies, ‘3’ was for head-to-head comparisons of multiple EWS, or for external validations or multiple locations, ‘4’ was generally for single site or single tool studies, and ‘5’ was for other studies that may have some relevance to the research questions at hand.

In the first instance key details of studies coded 1, 2, or 3 were tabulated (n = 34; see Appendix). Studies coded 4 or 5 (n = 23; 16 adult, 7 maternity) are surveyed in less detail, later in the review.

**Figure 1**: Screening and inclusion of studies (Adapt Research, 2024)

## Studies of Adult Inpatient EWS/Vital Sign Charts

### Review Articles

**See Appendix Table A1 for summary details of studies described below.**

#### Systematic Reviews

**McGuaghey et al. (2021)**

A Cochrane systematic review by McGaughey et al. (2021) updated an earlier 2007 Cochrane review. It looked at EWS and RRS and obtained studies up to 2020. Included studies were four RCTs and seven non-randomised studies. Two RCTs pre-dated 2016, and the setting of another one was Iran, not a usual comparator health system to New Zealand. The remaining RCT (Haegdorens et al. 2018) was already obtained from the searches of the present study and is discussed below. The RCTs were judged to be of high risk of bias, the seven non-randomised studies were judged to be of critical to serious risk of bias. The authors concluded that EWS and RRS may lead to little or no difference in hospital mortality, unplanned ICU admissions, length of hospital stay or adverse events. The review assessed outcomes only and did not assess predictive validity of EWS.

**Alhmoud et al. (2021)**

The systematic review by Alhmoud et al. (2021) examined randomised trials and observational studies of internal or external validation of EWS to predict deterioration (mortality, ICU transfer and cardiac arrest) in various disease subgroups and clinical settings. The review included 103 studies published between 2006–2019. Twenty-one studies compared different EWS head-to-head in the same cohort (these are listed in Alhmoud et al. 2021, Supplementary Table S2). Only two RCTs were included. Key findings include:

* Meta-analysis revealed that mortality, ICU admission and cardiac arrest were best predicted in medical (AUC mean: 0.74, 0.75 and 0.74) and surgical settings (0.80, 0.79 and 0.75), and respiratory diseases (0.75, 0.80 and 0.75).
* The best predictive performance was found in studies examining cardiac, stroke and renal diseases (AUC: 0.93, 0.88 and 0.87, respectively).
* In ED, predictive accuracy was variable (AUC: 0.56–0.91).
* In haematology, oncology, and gastrointestinal diseases, EWS predictive accuracy for various outcomes was suboptimal.
* Cardiac arrest was the least examined outcome among the three endpoints (n=8) and unstudied in cardiac diseases.

The authors conclude that EWS studies in different disease and clinical settings were heterogenous. The review found that validation of general EWS is limited in specialised settings. Furthermore, the methodology and quality of validation studies of EWS (significant risk of bias in most studies) are insufficient to recommend their use in all diseases and all clinical settings. However, the authors note that in medical and ED contexts, EWS can perform well, suggesting the role of EWS in general settings, or at the early stage of clinical assessment.

**Holland & Kellett (2022)**

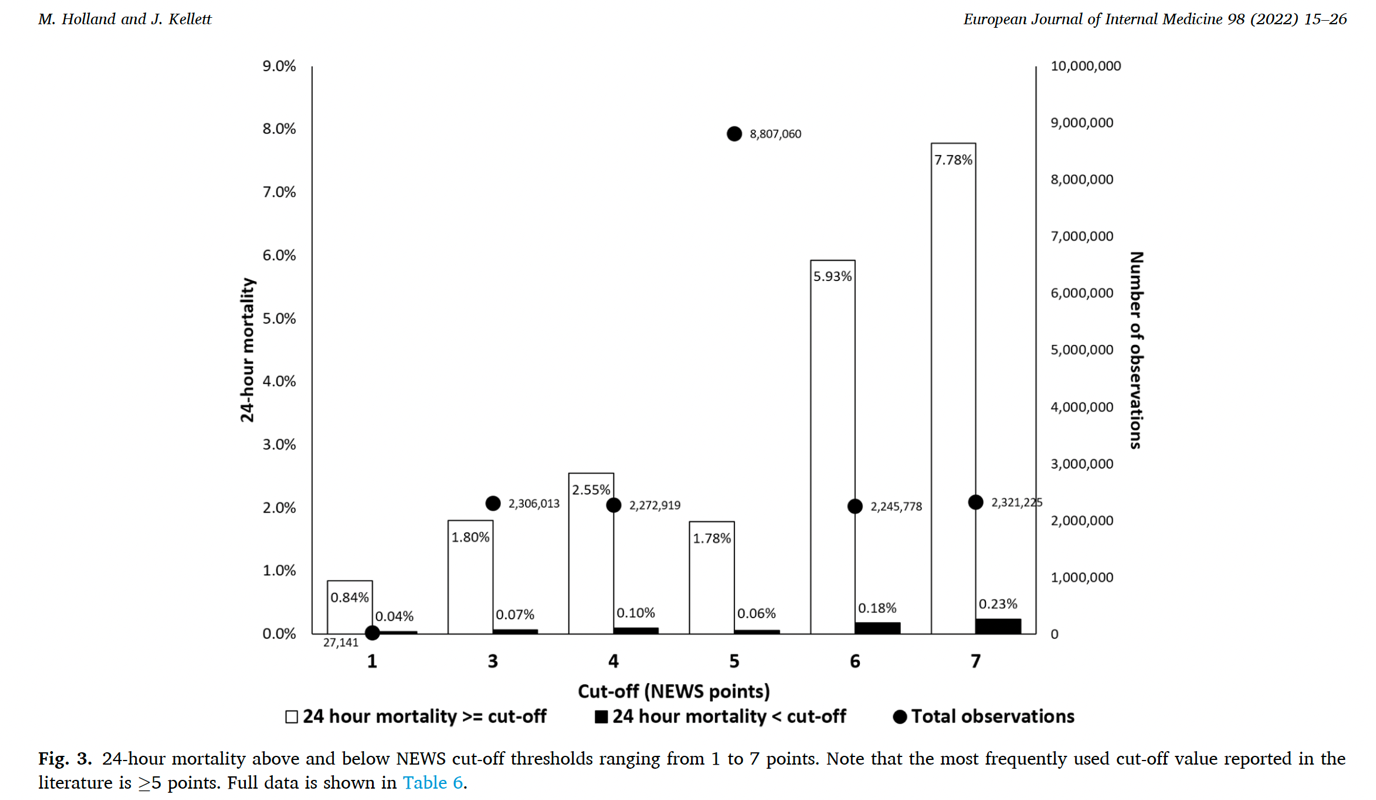
This systematic review obtained 121 studies published between 2012–2021 that examined the ability of ViEWS, NEWS, and NEWS2 to predict mortality by timeframe. Settings studied also included pre-hospital and ED and different diseases including sepsis and Covid-19. NEWS had good predictive validity for mortality within 24 hours. All studies with more than 50,000 observations had an AUC for 24-hour mortality of ≥0.880. This suggests that it may not matter which of these three EWS is used (although the only larger studies of NEWS2 were pre-hospital settings). In a meta-analysis the mean AUC for predicting 24-hour mortality was 0.897 for NEWS, 0.848 for NEWS2, and 0.886 for ViEWS. The authors noted that:

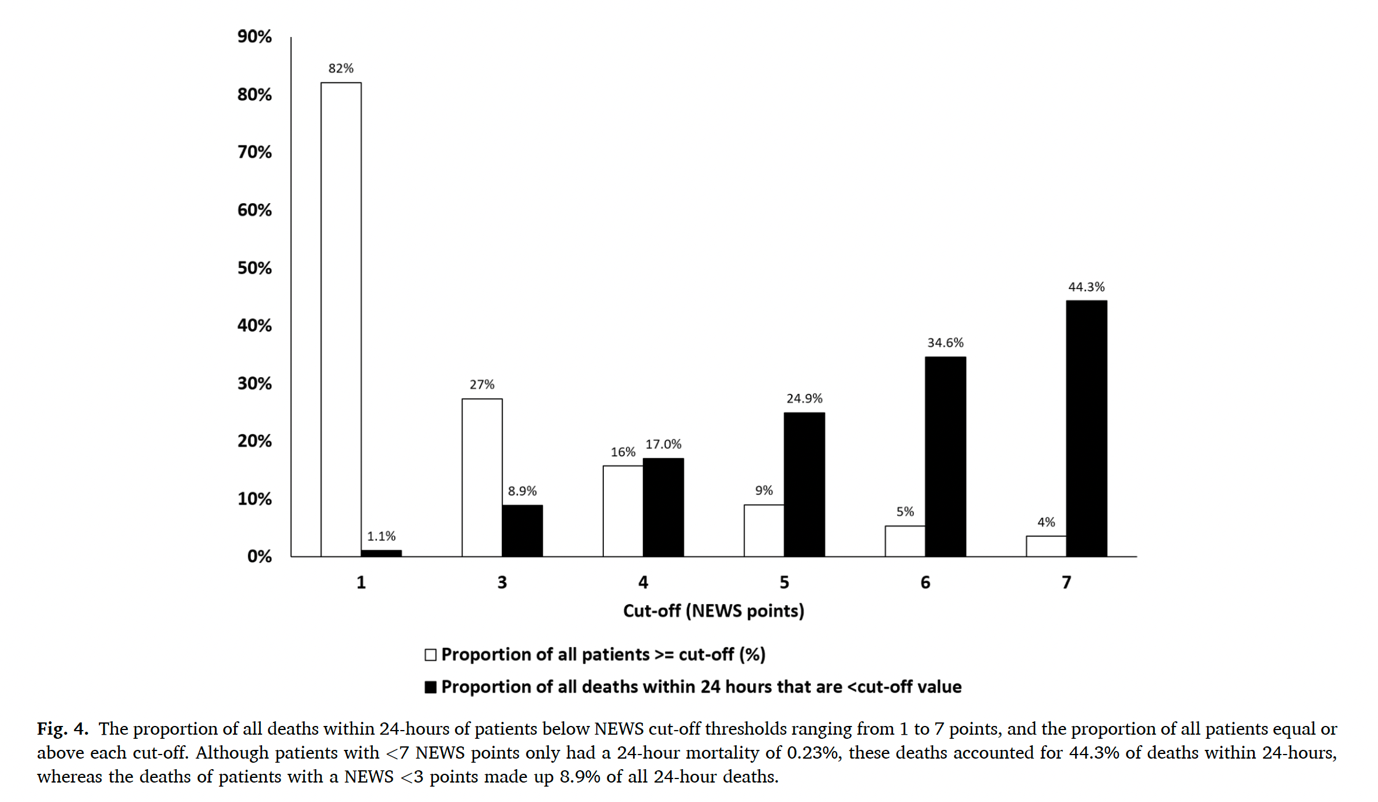
* Studies of general ward patients had a higher AUC for 24-hour mortality compared to patients seen earlier in their treatment.
* 24-hour mortality AUC may be lower for observations prior to hospital, in the ED, or patients admitted as emergencies, or in ICU.
* The AUC for in-hospital mortality of patients with respiratory illness, sepsis or suspected sepsis was significantly lower (*p <*0.0001) than for other patients.
* However, despite these confounders, only two studies reported an AUC for 24-hour mortality <0.83.
* The AUC deteriorated beyond 24 hours. It was noted that lots of patients die with low NEWS. Hence NEWS is specific but not sensitive.
* Table 3 in the paper lists the AUC across many settings and studies, along with number of observations.

These findings might imply that any of these three EWS are suited to use with routine admissions or in post-ED settings. NEWS reliably identifies patients most and least likely to die within 24-hours, but NEWS mortality predictions beyond 24-hours are unreliable.

The study additionally notes the 24-hour mortality at different NEWS thresholds and the proportion of patients that would trigger a response. Figure 3 and Figure 4 in Holland & Kellett (2022) provide a visual summary and are reproduced as Figure 2 below.

**Figure 2**: NEWS cut-offs vs deaths (as illustrated in Fig 3. & Fig 4. reproduced from Holland & Kellett 2022, *European Journal of Internal Medicine*)





**Fu et al. (2020)**

This systematic review evaluated 29 EWS validation studies (29 different EWS) according to the TRIPOD criteria. Two studies used unweighted activation criteria, 13 studies utilized aggregated weighted scores, and 14 studies applied complex computerized scores. EWS included between 3 and 72 predictor variables.

Predictor variables frequently included: heart rate, respiratory rate, SBP, DBP, and temperature. Mental status, pulse oximetry, and age were also common. Some models tracked vital sign trends, eg mean, SD, maximum, minimum, or range over time. Eleven studies utilized lab data for model derivation, but these varied greatly across studies. More complex algorithms incorporated the comorbidity index, length of stay, history of ICU stays, care directive status, physician orders, and patient demographic data.

Twenty-eight studies reported AUC, but only 7 studies reported PPV. Yet, EWS with poor PPV and high sensitivity potentially lead to alert fatigue and poor clinical usability.

The authors note a paradigm shift in EWS development over the past two decades from clinical consensus to data-driven approaches. Data-driven models rely on statistical methods for feature selection and model derivation and their performance is therefore strongly influenced by the database from which they are derived. For instance, omission of variables may be caused by low prevalence in the study cohort. The authors note that multiple studies have shown that aggregated weighted scores and computerized scores perform better in discriminating patients with higher risk of clinical deterioration than activation criteria.

**Gerry et al. 2020**

This systematic review included 95 studies (to June 2019) of the validation of EWS for use in adult inpatient populations, with nearly all studies using death, ICU transfer, cardiac arrest, or composite of these as outcome variables. 84 studies externally validated an EWS and these are summarised in Table 2 in the paper. Relevant to the present review, 12 development studies were published since 2016.

Predictor variables most used in EWS were respiratory rate (n=30/34, 88%), heart rate (n=28/34, 83%), oxygen saturation, temperature, and SBP (all n=24/34, 71%), and level of consciousness (56%). Age (n=13/34, 38%) and sex (n=3/34, 9%) were less frequently included.A median 12 predictors were considered in development of EWS, with a median of 7 included. The most common method to decide on inclusion was backwards elimination.

MEWS and NEWS were by far the most studied EWS (see Table 1 in Gerry et al. 2020) but the authors state that claims of ‘extensive validation’ might be misleading given risk of bias across studies. Also, NEWS was developed by consensus rather than statistical methods, which would be normal for predictive models. Electronic implementation of EWS might be an opportunity to introduce better scoring systems. Finally, simple bedside vital sign scoring systems often make the ‘unlikely assumption’ that each vital sign has the same predictive value. This is not the case, eg respiratory rate is often the best predictor and temperature may add little discrimination (as found in studies described below).

Gerry et al.’s review doesn’t provide details of validation outcomes and is overall critical of all research on these tools in the papers reviewed, with all included studies being rated at high risk of bias. Problems were observed across all aspects of study design and analysis. No study assessed clinical utility using net benefit methods. The authors noted that due to heterogeneity of outcomes and time horizons used in the validation studies, and the relative lack of head-to-head comparisons, they did not quantitatively synthesise performance metrics for specific EWSs. Some of these issues were addressed in subsequent studies (as below).

**Brekke et al (2019)**

This systematic review aimed to evaluate the ability of *trends* in intermittently monitored vital signs to predict clinical deterioration in patients hospitalized with acute illness. It included just two retrospective cohort studies, both without controls. One study was published pre-2016. The other (Churpek et al. 2016), retrospectively analysed vital sign data in an EHR. Data included: respiratory rate, heart rate, SBP, DBP, temperature and oxygen saturation, and included both medical and surgical ward patients. Univariate analysis found respiratory rate to be the best predictor of deterioration when using the current value, AUC 0.70 (95% CI 0.70–0.70). A model including the current respiratory rate and the maximum rate prior to current was the most the accurate predictor (AUC 0.73). However, the review authors report that although trends statistically improved the predictive power of all vital signs the improvements were considered minor.

**Credland et al. (2018)**

In a narrative review that included 7 studies (UK 4, Denmark 2, Netherlands 1), Credland et al. found poor compliance with the Early Warning Score (EWS) protocol, significant scoring inaccuracy with omitted EWS, missing elements of the EWS and incorrectly calculated EWS. Adherence to monitoring frequency was poor and higher EWS was associated with reduced compliance with the escalation protocol. These findings are notable in the context of the ‘dose-response’ relationship between EWS protocol compliance and outcomes found by Haegdorens et al. (2019) and described below.

**Mann et al. (2021)**

This systematic review of EHR and digital EWS tools included 46 published studies (2008-2020) on the development, validation, and implementation of tools for predicting patient deterioration in general wards in hospitals. Although digital EWS tools and advanced analytics are beyond the scope of the present report, this study is included to highlight their potential. A core finding of Mann et al.’s review is that machine learning approaches may be better at identifying patients at risk of deterioration than approaches using logistic regression, and advanced analytics tools are probably better than standard manual EWS. However, the authors also note that many studies are at risk of bias, including the many before and after studies, and there are also risks of model overfitting to specific datasets, which may not be predictive in future cohorts. There is also evidence of inconsistency in efferent limb performance (including patient assessment, intervention, transfer, monitoring, and follow-up). The authors conclude that despite relative progress in the development of algorithms to predict patient deterioration, the literature has not shown that the deployment or implementation of such algorithms is reproducibly associated with improvements in patient outcomes, but this remains understudied rather than unproved.

**Fang et al (2019)**

This systematic review included 48 studies of the methodologies and metrics used in studies to validate EWS. The key finding was that methodologies and performance metrics used in studies were heterogeneous hence making it difficult to interpret and compare EWS performance. Variation in patient episode definition (eg 24 hours, 30 days), and observation set definition (eg first set, worst score) means that metrics such as AUC cannot necessarily be compared across studies. These findings emphasise the importance of looking to head-to-head studies of EWS.

#### Non-Systematic Reviews

**Holland & Kellett (2023)**

In a wide-ranging review of NEWS, Holland & Kellett (2023) show that NEWS is well validated across many settings in unselected and disease specific patients (eg NEWS ≥5 is a better predictor of sepsis than SIRS or qSOFA). These authors note the modifications introduced in NEWS2, and find it is not as sensitive, has less predictive performance, and offers no benefits to type-2 respiratory failure patients compared with NEWS. The main problem now is determining how and when to respond to various NEWS scores, rather than what NEWS measures and how. A NEWS of ≥ 3 would generate a trigger in 27% of observations (see Figure 2, above), which would detect 88% of all deaths within 24 hours. In contrast, a NEWS of ≥ 5 would generate a trigger in 9% of observations, which would detect 73% of all deaths within 24 hours and a NEWS of ≥ 7 would generate a trigger in 4% of observations, which would detect 56% of all deaths within 24 hours. However, the authors also note the arbitrary nature of the Royal College of Physicians escalation pathway (graded response also recommended by NICE), which lacks an evidence base to support who should respond, at what threshold, and how. Response will be determined by resource constraints. Agreed context-specific protocols of how to respond to the score, and when to override it, must be developed and subsequently modified in the light of experience.

Any trigger threshold selected would depend on required management of conditions likely to be present for each score in that context, also on the patient population and their clinical setting, and available resources. Deaths in patients with higher scores may be less preventable and, therefore, a lower cut-off score may save more lives. Holland and Kellett propose that the number needed to evaluate (NNE) may be the most useful measure of clinical utility and cost-efficiency as it provides the number of patients that need to be evaluated further to detect one adverse outcome (ie the reciprocal of PPV).

The following additional weaknesses of NEWS were noted:

* NEWS may miss conditions such as acute kidney injury (urine output not measured), early distributive shock (DBP not scored), or delirium (AVPU not adequate).
* NEWS does not consider the patient’s usual blood pressure, heart rate, respiratory rate or oxygen saturation potentially leading to undertreatment of relative hypotension or overuse of supplemental oxygen in chronic lung disease.
* Respiratory rate may most accurately predict outcome, yet it is poorly measured.

**Williams (2022)**

Williams (2022) provides a narrative overview of the first 10 years’ experience with NEWS in the UK, raising two issues relevant to the present research questions:

* **Vital sign chart design:** The vital signs chart was updated for NEWS2 to overcome problems for those with red–green colour-blindness and the recording of vital signs were reordered to reflect the Resuscitation Council UK ABCDE (airway, breathing, circulation, disability, and exposure) format.
* **Neurological assessment:** Consciousness in the original chart used the familiar AVPU (alert, verbal, pain and unresponsive) scale. In the updated NEWS2 chart, ‘new-onset confusion’ was added, recognising that this is an important sign of physiological decompensation, even if patients are alert, especially in older patients.

Despite these changes, most studies comparing NEWS and NEWS2 find little difference in performance of the two tools, with NEWS2 possibly performing slightly worse (see below).

### Individual Studies

**See Appendix Table A2 for summary details of studies described below.**

#### Randomised Controlled Trials

**Haegdorens et al. (2018)**

This study was a stepped wedge cluster randomised controlled trial including 14 Belgian acute care hospitals with two medical and two surgical wards each. The intervention was implementation of a standardised observation (NEWS) and communication protocol (SBAR). NEWS was implemented based on evidence of its superiority to 33 other EWS. NEWS was measured every 12 hours with changes to frequency based on the NICE guidelines. Deviations from this protocol had to be signed by the attending physician. Half of the original 14 hospitals discontinued the study, which left the study possibly underpowered to detect any change in the outcomes of interest. Twenty-eight wards of seven hospitals studied from October 2013 until May 2015 were included in the final analysis. The control group contained 34,267 patient admissions and the intervention group 35,389. Comorbidity and nurse staff levels were collected as potential confounders. When adjusted for clustering and study time, there was no significant difference between the control and intervention group in unexpected death rates (1.5 vs 0.7/1000, OR 0.82, 95%CI 0.34–1.95), cardiac arrest rates (1.3 vs 1.0/1000, OR 0.71, 95%CI 0.33–1.52) or unplanned ICU admissions (6.5 vs 10.3/1000, OR 1.23, 95%CI 0.91–1.65).

**Haegdorens et al. (2019)**

This was a post-hoc analysis of the above RCT including 60,956 patients of which 32,722 were in the intervention group. Comorbidity scores were sampled in 3,600 patients and vital signs in 2,951 patients. In 668 patients, vital signs were collected before a serious adverse event. The mean number of vital signs per observation increased significantly in the intervention group. The observation frequency increased in patients with a serious adverse event and decreased in patients without a serious adverse event. Protocol compliance was negatively correlated with patient mortality adjusted for comorbidity and age, with Pearson product moment = -0.451 (p < 0.05*).* The authors state that only wards in the two highest comorbidity quartiles possibly benefitted from adhering to the protocol. The authors also state that their study is, to their knowledge, the only study confirming a dose–response relationship between NEWS protocol compliance and patient outcomes.

**Petersen et al 2016**

This was a pragmatic, ward-level randomised, non-blinded, controlled trial at an urban University hospital. The study included 1346 patients acutely admitted to a surgical or medical ward with an initial EWS of 0–1. The intervention was 12-hourly recording of vital signs vs 8-hourly recording, using an EWS based on NEWS. However, 60% of patients were lost to follow up, mainly due to early discharge. No significant reduction was observed in the proportion of clinical deterioration with monitoring frequencies of 12 vs 8-hourly.

#### Observational Studies

**Hogan et al 2020**

In an analysis of retrospective data from 106 NHS hospitals, Hogan et al. compared the use of NEWS versus other track and trigger systems and electronic implementation versus paper-based systems. The National Cardiac Arrest Audit provided up to 6 years’ historical data on 21,595 patients having had 22,057 in hospital cardiac arrests (IHCA) based on 13,059,865 hospital admissions. Data about interventions was at organisational not patient level. A random effects Poisson regression analysis adjusted for case mix, temporal trend and seasonality was conducted. Introduction or use of NEWS in a hospital was associated with a reduction of 9.4% in the rate of ward based IHCA compared to non-NEWS systems (incidence rate ratio 0.906, p < 0.001). The use of an electronic TTS was also associated with a reduction of 9.8% in the rate of IHCA compared with paper-based TTS (incidence rate ratio 0.902, p = 0.009). There was no change in hospital survival.The interpretation is that electronic implementation seems better than paper based EWS; and NEWS seems better than non-NEWS systems. However, there could be other wrap-around factors associated with EHR and NEWS implementation that confound the relationships.

**Bhonagiri et al 2021**

This study reports analysis of the NSW Between the Flags (BTF) system. BTF is a standardised single trigger system employing colour coded charts. The study was a prospective observational analysis of data from the BTF registry (2010–16), and the Australia and NZ Intensive Care Society Adult Patient Database (2008–16). Between August 2010 and June 2016, 5506 IHCA occurred in 35 sites among 7,068,849 hospital admissions. The cardiac arrest rate per 1000 hospital admissions declined from 0.91 in the BTF implementation period to 0.70 post-implementation. Propensity score analysis showed significant declines in ICU and hospital mortality and length of stay for cardiac arrest patients admitted to the ICU (all p < 0.001). It was estimated 912 cardiac arrests were averted due to programme implementation (911.5, 95% CI: 738.3–1075.0). There was also a significant decline in all-cause mortality, with an average 6.2% reduction in the post-implementation period (January 2013 to June 2016) versus the implementation period (August 2010 to December 2012) (P < 0.001). In the same period the incidence of RRT activations increased from fewer than 10 per 1000 hospital admissions to more than 25 per 1000 hospital admissions. The BTF implementation included governance and education initiatives, so it is not clear from this before and after study the degree to which the BTF vital signs chart was effective. There was also no comparison with any aggregate-weighted score EWS. For reference, Table 1 below compares the BTF vital sign thresholds to those used as single variable triggers in the NZEWS and the ‘red’ zone on NEWS.

**Table 1**: BTF vital sign thresholds versus single variable MET triggers in the NZEWS, with NEWS for reference. Given the more permissive mandatory RRT thresholds in NZEWS compared to BTF for RR, SBP, and LoC, it seems that NZEWS will trigger fewer RRT based on single triggers than BTF. This is likely to be desirable given BTF’s higher sensitivity and lower specificity when compared with NEWS, (B = blue, R = red, Y = yellow, N = normal).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Vital sign | NSW BTF  (single trigger tool; any ‘red’ requires RRT) | NEWS  (not a single trigger tool, but any ‘red’ requires medical team notified) | NZEWS  (hybrid tool, any ‘blue’ requires RRT) | Comment |
| Respiratory rate | R: <5/>30  Y: <10/>25  N: 10–25 | R: ≤8/≥25  Y: 21–24  N: 12-20 | B: ≤4/≥36  R: <9/>25  Y: 9–11  N: 12–20 | NZ more permissive of high RR than BTF before RRT; note Campbell et al. (2020) found no difference in Q-ADDS when altering RR trigger from 30 to 36. |
| O2 saturation | R: <90  Y: 90–95  N: >95 | R: ≤91  Y: 92–93  N: ≥96 | B: N/A  R: ≤91  Y: 94–95  N: ≥96 | NZ higher O2 sats threshold than BTF, but O2 sats do not trigger RRT |
| Supplemental O2 | No red/yellow | R: N/A  Y: yes  N: no | No red/blue | Supplemental O2 is not an RRT trigger in any of the three tools. |
| Systolic BP | R: <90/>200  Y: <100/>180  N: 100–180 | R: ≤90/≥220  Y: 91–100  N: 111–219 | B: <70  R: <90/≥220  N: 110–219 | NZ more permissive of low and high SBP than BTF; note Campbell et al. (2020) found reducing Q-ADDS SBP threshold from 90 to 80 reduced RRT calls by 45% with small drop in sensitivity. |
| Heart rate | R: <40/>140  Y: <50/>120  N: 50–120 | R: ≤40/≥131  Y: 111–130  N: 51–90 | B: <40/≥140  R: >130  N: 50–90 | NZ RRT trigger for HR same as BTF. |
| Temperature | R: N/A  Y: >38.5/<35.5  N: 35.5–38.5 | R: ≤35.0  Y: ≥39.1  N: 36.1–38.0 | B: N/A  R: N/A  Y: ≤34/≥39  N: 36–38 | NZ more permissive of low/high temperature, but BTF does not have ‘red’ temperature zones; note Luis et al. (2018) found temperature didn’t contribute to model discrimination; note Sebat et al. (2020) found increased temperature associated with improved survival. So, temperature trigger may not be doing much work in adult EWS. |
| Level of consciousness | R: P,U  Y: V,C  N: A | R: V,P,U  N: A | B: U  R: V,P  N: A | NZ more permissive of neurological disturbance than BTF, requiring ‘U’ for RRT |
| Pain | R: N/A  Y: 7–10  N: 0–6 | N/A | N/A | No vital sign chart triggers RRT for pain |

**Bedoya et al 2019**

Bedoya et al. collected data at a US tertiary care academic hospital and a community hospital on 85,322 patients, including 42,402 patients pre-NEWS and 42,920 patients post-NEWS implementation. NEWS was integrated into the EHR and an automated best practice advisory was notified at a threshold of NEWS ≥7. Following implementation there was no change in the primary outcome of ICU transfer or death. The AUC was 0.72–0.80 and 0.74–0.90 for the academic and community facilities respectively, and re-training NEWS with newly generated hospital-specific coefficients improved model performance. However, the authors state that the NEWS alert was ignored in 86% of cases. Additionally, Holland & Kellett (above) contest that deaths in patients with NEWS ≥7 are less likely to be preventable. Also, using just an automatic trigger at NEWS ≥7 leaves no room for a graded response at lower scores.

**Brabrand et al (2017)**

This retrospective single centre cohort study tested multiple EWS head-to-head in the same population, where the vital sign score of interest was that obtained in the ED. Tools studied included ED-specific tools evaluated head-to-head with the NEWS. Tools evaluated were: Goodacre, Groarke, Worthing, Rapid Acute Physiology Score, Rapid Emergency Medicine Score, and NEWS. Results showed that discriminatory power (AUC) for 24-hour mortality was above 0.8 for all EWS (except the Groarke score, 0.587 – noting that 0.5 defines no discriminatory ability at all) and was highest for the Worthing score (0.847). The discriminatory power for predicting overall in hospital mortality was lower on average and highest for the Goodacre and Worthing scores (0.810 and 0.800 respectively) but below 0.8 for the remaining scores. Results showed that NEWS was <0.02 AUC worse than tools with the highest AUC for 24-hour and in-hospital mortality. NEWS is much more widely studied as an EWS. This study was included in the systematic reviews above by Gerry et al. and Alhmoud et al.

**Campbell et al. (2020)**

This retrospective observational study compares four EWS head-to-head in the same population, including a single trigger system (BTF), two aggregate scoring systems (Q-ADDS and NEWS) and an advanced analytic tool (eCART). It provides external validation of the Q-ADDS, NEWS, and BTF in a Chicago hospital cohort. Outcomes assessed were ICU transfer or death. Additionally, four different variations of the Q-ADDS were tested, comparing two thresholds for low SBP and high respiratory rate. Unlike the study by Bedoya et al. (2019) above, in this case additional analysis examined NEWS ≥5 or any single parameter scoring 3 (triggering ward escalation), which is consistent with the NICE guidance.

Q-ADDS combines vital-sign based scoring like the NEWS tool with hard trigger limits, like the BTF tool used in NSW. The hard limits aim to add an extra level of safety. It also has an additional escalation tier and quantifies oxygen requirement as a contribution to the score. It includes pain and urine output, but pain doesn’t generate a score. Note, some other studies of Q-ADDS obtained for this review were specific to rural medical facilities and had been excluded as potentially not representative of national NZ hospital inpatient populations.

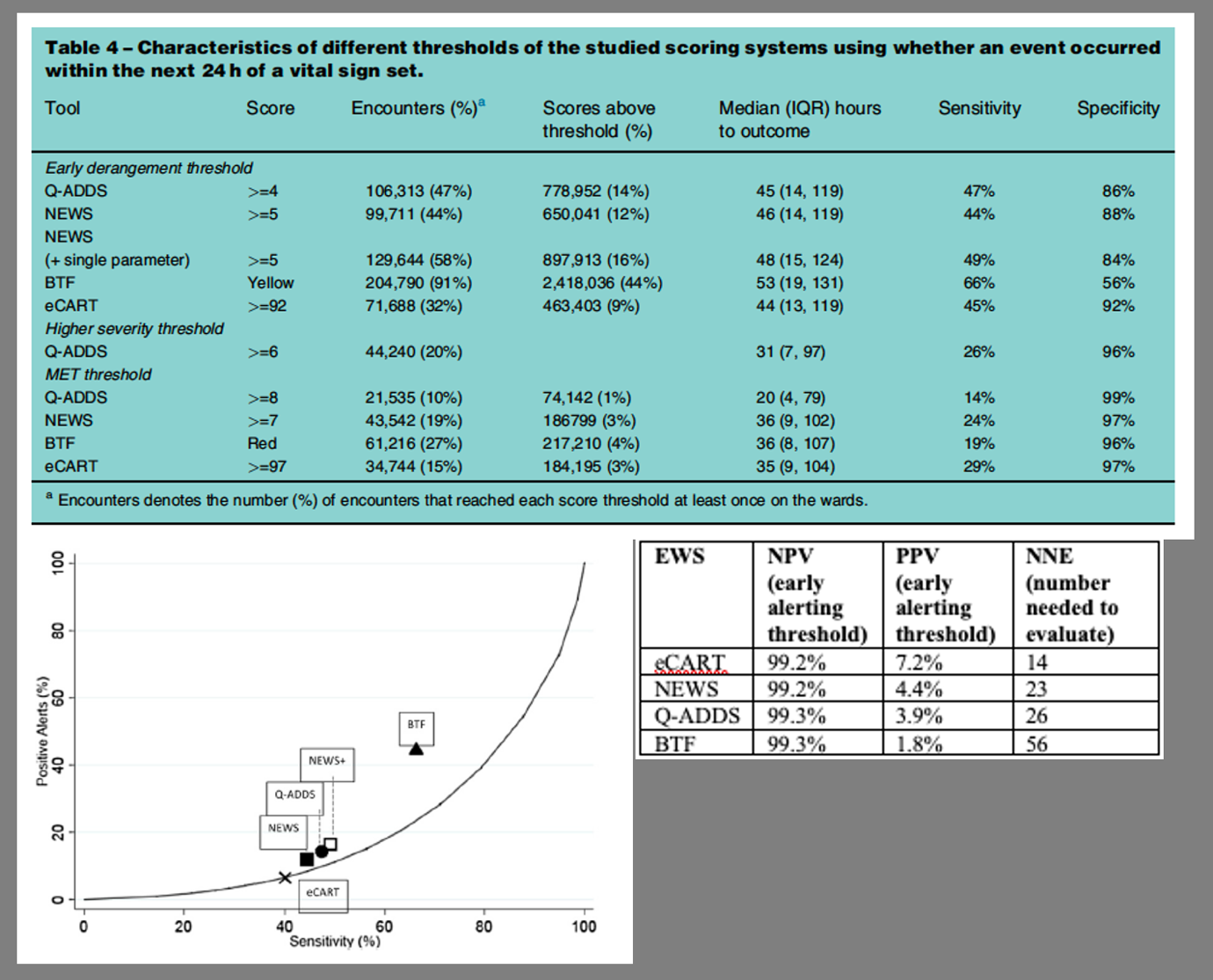
Results expressed as AUC for predicting the combined outcome showed Q-ADDS (AUC 0.71) and NEWS (AUC 0.72) had similar predictive accuracy, BTF (AUC 0.64) had the lowest, and eCART (AUC 0.76) the highest. Similar ordering of EWS was found for predicting ICU transfer (eCART 0.78, NEWS 0.72, Q-ADDS 0.71, BTF 0.63) and for predicting death (eCART 0.92, NEWS 0.88, Q-ADDS 0.87, BTF 0.78).

The four variations to Q-ADDS made no difference to its predictive ability as measured by AUC. Changing the high RR MET thresholds (≥36 vs ≥30) did not alter the NPV, PPV or sensitivity. However, lowering the SBP MET threshold from < 90 mmHg to <80 mmHg alone reduced MET alerts by 45% (18 to 10% of encounters) reducing sensitivity at the MET threshold from 17% to 14%, without impacting on NPV (99.0%/99.0%).

The authors conclude that it is likely that additional parameters improve accuracy rather than the manipulation of parameter thresholds. For example, the superior accuracy by adding patient demographics and laboratory results in eCART, versus the similarity between AUCs of the four variants of Q-ADDS.

The paper includes a useful table comparing the sensitivity and specificity of the tools. Q-ADDS, NEWS, and eCART all have similar sensitivity at the early derangement threshold (ie low cut-off score). BTF has higher sensitivity, but the price it pays is a much higher positive alert rate, as depicted in Figure 3.

**Figure 3**: Characteristics of eCART, NEWS, Q-ADDS, and BTF as reported by Campbell et al. (2020) (Adapted from Campbell et al. 2020, *Resuscitation*)



**Green et al. (2018)**

Like Campbell et al. (2020) above, this study compared the performance of multiple EWS (eCART vs NEWS vs MEWS vs BTF) in the same cohort as an external validation exercise. Data from five US hospitals (2008-13) on all admissions was collected, n = 107,868 (938 deaths, 4 million vital sign observations). The outcomes of interest were cardiac arrest, ICU transfer, or death within 24 hours of an observation.

Results showed that overall accuracy was highest for eCART, with an AUC of 0.801 (95% CI 0.799–0.802), followed by NEWS, MEWS and BTF respectively (0.718 [0.716–0.720]; 0.698 [0.696–0.700]; 0.663 [0.661–0.664]). The BTF ‘red zone’ had a specificity of 95.0% and a ‘yellow zone’ specificity of 27.5%, which corresponded to MEWS thresholds of ≥4 and ≥2, NEWS thresholds of ≥5 and ≥2, and eCART thresholds of ≥12 and ≥4, respectively. At those thresholds, eCART caught 22 more adverse events per 10,000 patients than BTF using the ‘yellow zone’ criteria and 13 more using ‘red zone’, while MEWS and NEWS identified the same or fewer. The sensitivity and specificity of the various EWS are displayed in Figure 4, along with the bar charts illustrating the number of additional false positives (top bar chart) and additional ‘saves’ (bottom bar chart) each tool generated compared to BTF.

**Figure 4**: Characteristics of eCART, NEWS, MEWS, and BTF as reported by Green et al. (2018). Note that MEWS generated the fewest false positives, while eCART facilitated the most saves (Figure credit Green et al. 2018, *Resuscitation*).



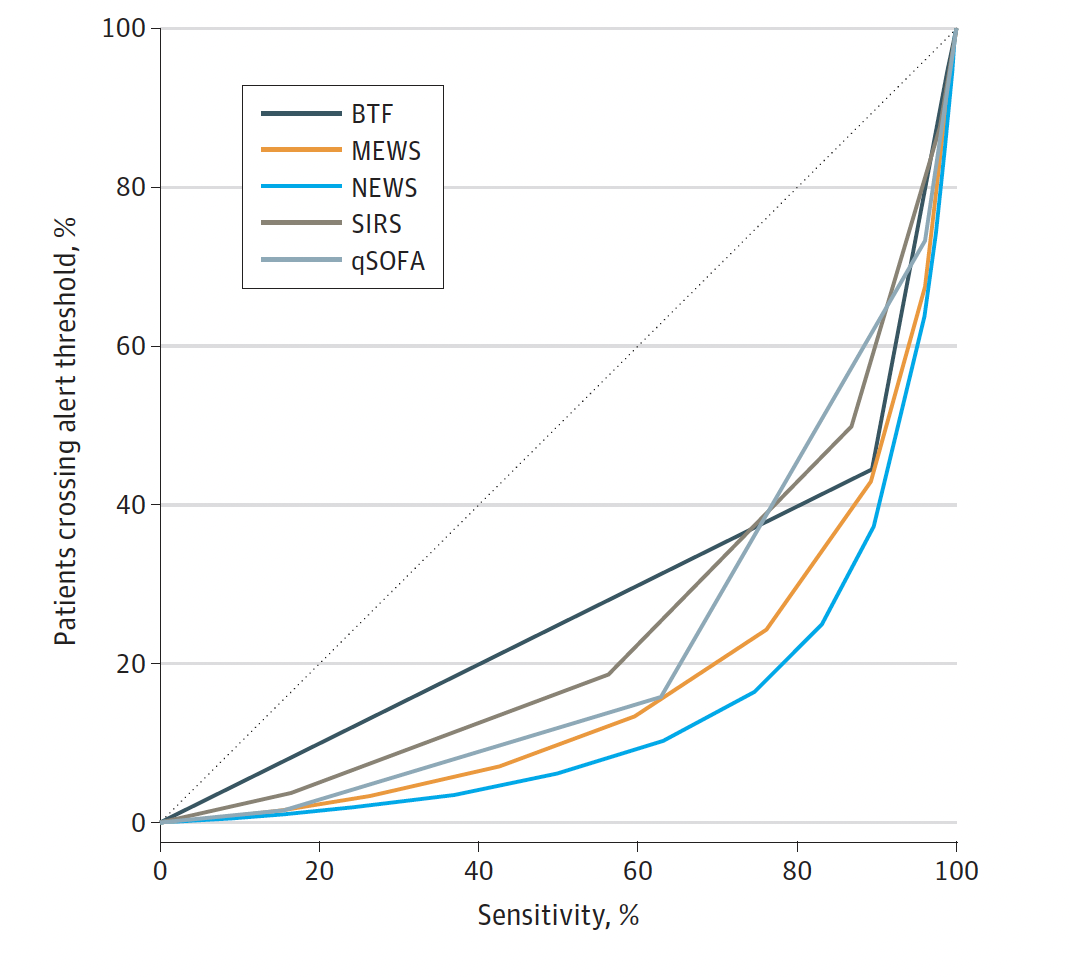
**Holland et al. (2023)**

In a study of adult patients (aged 16 years and older) who were emergency admissions to Salford Royal Hospital (UK) between 2014 and 2022, Holland et al. identify risk of in hospital mortality. Based on the NEWS score at admission, those with NEWS ≥3 experienced an 8-fold increased risk of in hospital death. However, 45% of all in-hospital deaths occurred in patients with an admission NEWS <3. Additionally, 60% of in-hospital deaths were in four ICD-10 chapters: infections, circulatory and respiratory diseases, or neoplasms. This additional information could be a useful aid to risk stratification. The study did not report metrics such as AUC, PPV, NPV. The authors conclude that mortality in emergency hospital admissions is associated with illness severity and four diagnostic chapters. NEWS should not be the only arbiter of hospital admission, as for certain diagnostic chapters the risk of death is high even if vital signs on presentation are normal.

**Liu et al (2020)**

This retrospective observational study of hospitalised patients at 21 California and 7 Illinois hospitals from 2006 to 2018 compared NEWS, MEWS, BTF, and sepsis specific tools qSOFA and SIRS in the same cohort. The outcome of interest was in-hospital mortality or ICU transfer. The NEWS exhibited the highest discrimination for mortality (AUC, 0.87 in California vs 0.86 in Illinois), followed by the MEWS (0.83, 0.84), qSOFA (0.78, 0.78), SIRS (0.76, 0.76), and BTF (0.73, 0.74). At specific decision thresholds, the NEWS outperformed the SIRS and qSOFA at all 28 hospitals either by reducing the percentage of at-risk patients who need to be screened by 5–20% or increasing the percentage of adverse outcomes identified by 3–25%. In this study NEWS outperformed sepsis specific tools, even in patients with suspected infection. This is not surprising given that NEWS includes all the variables in qSOFA and BTF.

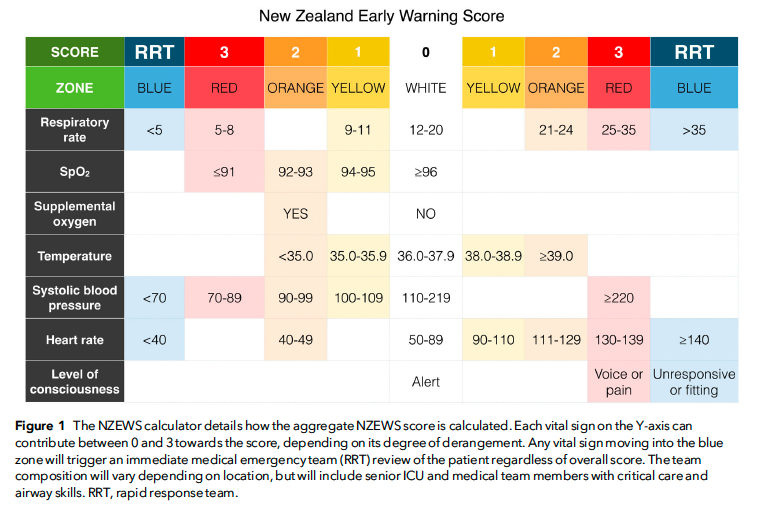
**Figure 5**: EWS efficiency curves for in-hospital mortality in all patients as reported by Liu et al. (2020). Note that NEWS was associated with the lowest proportion of patients crossing the alert threshold to require screening (Figure credit Liu et al. 2020, *JAMA Network Open*).



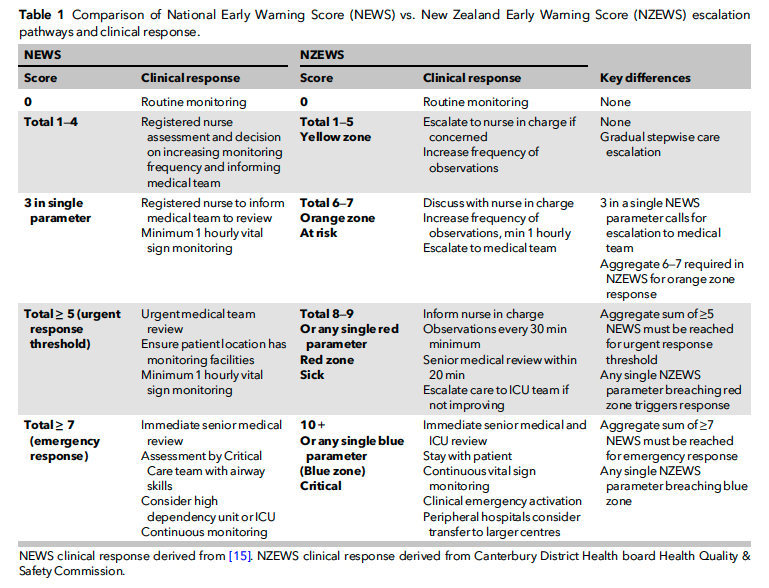
**Mohan et al. (2023)**

This retrospective observational study compared the performance of NZEWS vs NEWS at 6 hospitals in Canterbury District Health Board. NZEWS is presented in Figure 6 below. A total of 1,738,787 aggregate scores (13,910,296 individual vital signs) were obtained from 102,394 hospital admissions. The outcome of interest was cardiac arrest, death and/or unanticipated ICU admission. Results showed that the AUC for both EWSs for any adverse outcome in this population was 0.874. NZEWS performed better in surgical than medical patients (and the dataset included elective surgical patients). Over one-third of all aggregated vital sign sets taken scored zero. Most scores fall below the escalation trigger (NZEWS ≥ 6). For NZEWS scores of 10 (blue zone requiring immediate RRT review), 8% of patients died within 48 hours.

**Figure 6**: The NZEWS (reproduced from Figure 1 in Mohan et al. 2023, *Anaesthesia*)



**Table 2**: Comparison of escalation pathways NEWS vs NZEWS (Table credit: Mohan et al. 2023, *Anaesthesia*)



**Pimental et al. (2019)**

This multicentre, retrospective, observational study compared NEWS and NEWS2 head-to-head in the same cohort, with a focus on patients with type-II respiratory failure (T2RF) (one of the targets of the NEWS2 modifications). All completed adult admissions at five UK hospitals were included in the study population. The outcome of interest was death, cardiac arrest, or ICU within 24 hours of an observation set. A total of 251,266 adult admissions were examined, 48,898 were identified to be at risk of T2RF by diagnostic coding. Results indicated that NEWS2 modifications to NEWS do not improve discrimination of adverse outcomes in patients with documented T2RF (AUC 0.841 vs 0.862 for NEWS, not statistically significant) and decrease discrimination in patients at risk of T2RF (AUC 0.860 vs 0.881 for NEWS, significant). The PPV of NEWS2 in patients not at risk of T2RF was half that of NEWS (2.7 vs 5.0 at score ≥5, and 5.7 vs 11.2 at score ≥7).

**Tirkkonen et al. 2019**

Using different methods and different outcomes, Tirkkonen et al. (2019) found much the same regarding NEWS2 in a prospective observational study that included 886 patients triggering RRT in a Finnish hospital. This study included all adult RRT patients and calculated the NEWS at the time of RRT. Outcomes of interest were the prediction of (1) immediate need for ICU, (2) in hospital mortality or discharge with specified cerebral performance or limitations on treatment. Results showed that NEWS had fair discrimination for both the immediate (AUC 0.73; 0.69–0.77) and in hospital (0.68; 0.64–0.72) outcomes. However, use of the new SpO2 scale for the patients with confirmed T2RF did not improve the discrimination capability (AUC 0.73; 0.69–0.76 and 0.68; 0.64–0.71).

**Price et al. 2023**

Price et al. (2023) performed a retrospective observational analysis of 360,361 admissions (2010–2018) in a large NHS acute general hospital. The aim was to compare NEWS2 with 36 other kinds of EWS in 123 different diagnostic groups (note the similarity to the work of Alhmoud et al. 2021 above, with the Price et al. 2023 study published after that systematic review). Outcomes evaluated were death, ICU admission, or a combined outcome of either death or ICU admission within 24 hours of an observation set. The focus was on NEWS2, and NEWS does not appear to have been separately studied. Results showed that NEWS2 performed best or joint best in 120/123 patient groups. NEWS2 and Q-ADDS were the top two EWS for predicting primary outcome across the 123 disease groups (AUC 0.882 and 0.848 respectively). The only disease groups where NEWS2 was outperformed were peripheral arterial embolism orthrombosis, respiratory failure; insufficiency; arrest (adult), and lung disease due to external agents. The differences in AUC between the top performer and NEWS2 within these groups were 0.05, 0.02 and 0.03 respectively. The authors conclude that patients are not disadvantaged by use of NEWS2 versus other tools (again note NEWS was not included). AUC ranged from 0.733 to 0.917 for the combined outcome, 0.728 to 0.939 for death within 24 hours of an observation, and 0.646–0.907 for ICU admission within 24 hours of an observation. However, even when NEWS2 performed comparatively well, it had absolute poorer predictive performance in some groups. These groups were cardiac arrest and ventricular fibrillation, aortic & peripheral arterial embolism or thrombosis, and respiratory failure; insufficiency; arrest (adult), which all had AUC below 0.8.

### Additional Studies

Several other studies were obtained, which initial screening ranked as ‘4’ or ‘5’ (ie it appeared that they provided lower quality evidence based on smaller or single centre studies or provided additional validation of tools already assessed in systematic reviews). These studies are briefly tabulated below (Table 3), and their possible implications noted.

**Table 3:** Sixteen additional studies of adult EWS obtained through the literature searches but initially coded as lower-level evidence or relevance.

| **Study** | **Description** | **Implications** |
| --- | --- | --- |
| Bell et al (2021) | Examined an EHR-integrated, statistically derived, trend-based, deterioration index to predict unexpected death, unplanned ICU transfer, urgent surgery, and rapid-response alert (out of scope for this review) but compared it with MEWS/NEWS/BTF in two large Australian hospitals, finding the EHR system had higher discrimination.  The addition of vital sign and laboratory trend values to the logistic model increased the area under the curve from 0.84 to 0.89 and the sensitivity to predict an adverse event 1–48 hours prior from 0.35 to 0.41. | Supports other findings that adding variables increases discrimination. Adding demographics may be an easy way to increase sensitivity at high specificity.  Demonstrated that the BTF yellow flag threshold had very low PPV, which is support for using the red cut-offs as single value triggers. |
| Escobar et al. (2020) | Prospective study of staggered intervention of an EHR-integrated AAM statistical model (out of scope for this review) at 21 Kaiser Permanente hospitals.  Mortality within 30 days after an alert was lower in the intervention cohort than in the comparison cohort (adjusted relative risk, 0.84, 95% confidence interval, 0.78 to 0.90; P<0.001). | Like eCART the AAM electronic model appears effective. And the results demonstrate that EWS systems can impact hard endpoints like mortality. |
| Haegdorens et al. (2020) | Observational retrospective study in Belgium using cross-sectional sample (admitted patients, 1 day every 4 months) and a serious adverse event sample (all patients with unexpected death, cardiac arrest, and unplanned admission to the intensive care unit).  NEWS score ≥5 had a positive predictive value of 6.8% and a negative predictive value of 99.5% to predict unexpected death, cardiac arrest with cardiopulmonary resuscitation or unplanned admission to intensive care (AUC 0.841). | Authors argue that NEWS ≥5 is fit for purpose.  But, given the false positive load, hospitals need workable guidelines for response. |
| Hwang & Chin (2020 | Retrospective observational study in South Korea, examined NEWS2, MEWS, qSOFA, and clinical worry score, to predict patient deterioration.  At admission NEWS2 AUC was 0.77 (which was greater than MEWS and qSOFA). 24 hours before event both NEWS and NEWS2 AUC 0.91. Clinical Worry Score AUC 0.78 (weighted 0.85). | Clinical worry score probably doesn’t add much to NEWS/NEWS2 |
| Lee et al. (2018) | Observational retrospective study in South Korea, using NEWS and other scoring systems to predict in-hospital mortality. Combination model including age and diagnosis was tested.  Results showed the combination model that used other factors, such as age and diagnosis was more effective than NEWS alone in predicting hospital mortality (NEWS: 0.765; combination model: 0.861; p < 0.005). | Age and diagnosis can improve EWS prediction. However, younger patients who previously would have been salvaged by an early warning system might not be monitored effectively. |
| Loisa et al. (2022) | Large (n = 19,000 patients), 1-year, prospective, observational three-centre study at Finnish hospitals, general medical & general surgical wards. The first measured vital sign datasets on general wards were prospectively collected.  NEWS discriminated 1-day non-survivors with excellent accuracy (AUC 0.91, 95% CI 0.87 to 0.95) and 30-day mortality with acceptable accuracy (0.75, 95% CI 0.73 to 0.77).  The NEWS’s respiratory rate component discriminated 1-day non-survivors better (0.78, 95% CI 0.72 to 0.84) as compared with the oxygen saturation (0.66, 95% CI 0.59 to 0.73), systolic blood pressure (0.65, 95% CI 0.59 to 0.72) and heart rate (0.67, 95% CI 0.61 to 0.74) subcomponents (p<0.01 in all ROC comparisons). | External validation of NEWS in a larger, unselected, prospective population.  Prediction of 24-hour mortality/deterioration is probably the best use of NEWS.  Respiratory rate has best discrimination and repeated measures on general wards have value. |
| Luis et al. (2018) | Portuguese retrospective analysis of older population of patients in general medical, surgical and haematology wards. Compared NEWS to two candidate ‘short-NEWS’ EWS that excluded some parameters. Outcome was prediction of unanticipated ICU or death. Small study n = 300 patients.  NEWS presented an excellent discriminating capability (AUC 0.944). Temperature and systolic blood pressure (SBP) parameters did not contribute significantly to the model. Both short-NEWS models (without temperature, and without temperature and SBP) had an excellent discriminating capability (AUC 0.965 and 0.903, respectively) and a good predictive power. | Small study but could consider further research on dropping temperature or SBP from NEWS (in older populations). Note, low SBP causes most MET calls, and high temperature correlated with better outcomes in one study. |
| Malycha et al. (2019) | Large, multi-centre, cohort study in the UK comparing NEWS with NEWS + FiO2.  NEWS-FiO2 (AUC 0.823, 95% CI 0.819–0.824) outperformed NEWS (AUC 0.811, 95% CI 0.809–0.814) when predicting in-hospital death or unplanned ICU admission within 24 h of a complete set of vital sign observations. | Adding FiO2 gave small increase in AUC of +0.012 but unclear if this magnitude is clinically relevant. |
| Nestor et al. (2022) | Retrospective observational study at one hospital in Ireland, mixed patient general setting. Included only calls for NEWS ≥7, score of 3 in single variable, or concern about patient. These triggers identified a very high mortality population, late in the patient’s stay, and accounted for less than half transfers to critical care. Authors concluded that screening performance of EWS data is undermined by the more frequent use of other criteria for critical care transfer, suggesting that its value is casemix-dependent and might vary between – or even within – institutions. | Study has uncertain significance as it only examined cases of RRT activation, not lower NEWS scores where alterations to observation or management might avert deterioration. |
| Pankhurst et al. (2022) | Retrospective, single site, observational study in the UK, comparing different NEWS2 thresholds to predict outcome of death or ICU transfer, n = 100,362 consecutive admissions. Examined the index-NEWS2 (first one) and all-NEWS2. Analysed the not for CPR patient cohort separately.  Sensitivity of outcome prediction at Index-NEWS2  ≥5 was 0.46, and number needed to evaluate (NNE) was 52. Authors note that at this threshold, there would be a mean of 37.6 alerts/100 patient-days, occurring in 12.3% of patients on any day. Threshold changes to increase sensitivity by 0.1, would result in a twofold increase in alert rate and 1.5-fold increase in NNE. Overall, NEWS2 classification performance was significantly worse on Index-scores than all-scores (c-statistic= 0.78 vs 0.85; p<0.001). Note that ≥5 OR single=3, NNE=76, PPV 1.3%, NPV 99.9% versus ≥5 (only), NNE=52, PPV 1.9%, NPV 99.9%. The discussion notes that a health economic analysis was not possible but is needed.  Authors note that the EHR could be exploited to automate and limit the RRS. They cite Escobar et al. (2020) (see above, who studied their AAM system, which generated 2.8 alerts per 100 patients per day) | NEWS2 on admission is not as predictive as NEWS2 during admission.  NEWS2 threshold of ≥5 requires review of 12.3% of patients, and may not be efficient use of resources, a health economic assessment is needed.  Findings indicate there is a case for dropping the “or single value = 3” criteria (as this would reduce sensitivity by 3.5% and increase specificity by 3.1%, while reducing NNE from 76 to 52). See Table 2 in Pankhurst et al. (2022). |
| Pedersen et al. (2018) | Danish study of 2,835,331 NEWS records, which found 10% were incomplete and 0.2% had implausible values.  Anomalies include preponderance of RR divisible by 2 and 4, HR frequently just below 91 (the cut-off for generating points). Temperature recording the most frequently missing. | Manual RR recording could be improved.  Temperature is the value most frequently missing, but may not add much to model discrimination, and doesn’t have a ‘blue’ trigger.  However, note that NEWS was validated in a manual setting where these problems were likely present. |
| Sebat et al. (2020) | Prospective observational study of 6480 consecutive Rapid Response Team (RRT) patients from general medical, surgical, or intermediate care units at a 450-bed regional medical centre in California. Analysed capillary refill time vs 10 other EWS variables in the ’10-SOV’ tool.  20% of RRT-patients had prolonged-CRT (vs normal-CRT), were twice as likely to die (36% vs 17.8%, p < 0.001), more likely to experience the combined outcome (72.1% vs 54.2%, p < .001) and had longer hospital length of stays, 15.3 (SD 0.3) vs 13.5 days (SD 0.5) (p < .001). Multivariable logistic regression for mortality ranked CRT second to hypoxia among all 11 variables evaluated (p < 0.001).  Not all 10-SOV components were associated with increased mortality. Elevated temperature and pain were associated with an improved survival. | Capillary refill time warrants more research in EWS other than 10-SOV.  Temperature elevation was associated with improved survival (consistent with finding above by Luis et al. (2018) that temperature doesn’t add to model discrimination). |
| Shamout et al. (2019) | Retrospective observational study of 142,806 admissions in the UK, outcomes included prediction of ICU transfer, cardiac arrest, and death.  Study compared the performance of an age specific EWS (ASEWS) to the NEWS and a previous EWS (MCEWS). The ASEWS was based on statistical distributions of vital signs per age subset, rather than just adding age as a variable in the model.  ASEWS performed better in the 16–45 years age group than NEWS and MCEWS (Differences in performance were not consistent in the elder age group).  To achieve a sensitivity of 80% using ASEWS, the medical staff would respond to only half of the triggers generated by NEWS among younger patients. | ASEWS performed better than NEWS in <45 age group cohort.  Further research on including age in EWS is warranted, as it may help decrease RRT activations. |
| Rasmussen et al. (2018) | Retrospective observational study of 17,312 consecutively admitted acute medical patients to the Acute Medical Unit at Copenhagen University Hospital. Compared the ability of NEWS vs NEWS + age + sex vs NEWS + age + sex + suPAR to predict in mortality.  Results showed that for in-hospital mortality, NEWS had an AUC of 0.87 (with PPV 0.129), NEWS + Age + Sex AUC 0.89 (PPV 0.116), NEWS + Age + Sex + suPAR AUC 0.92 (PPV 0.130) | There may be small gains in AUC by adding age, sex, and suPAR, but when looking at the PPV we see that NEWS alone had equal best PPV for in-hospital mortality (0.129 vs 0.130 in the full model). So, there may be questionable gains from complicating the tool. |
| Elliot et al. (2016) | Prospective multisite survey (n = 477) of user acceptance of new chart design, when tested in parallel with existing charts. Some concerns were noted for chart size and style, use of ranges to graph vital signs and with specific human factors design features. | Explicit training may be needed on the principles and rationale of human factors chart design (results discussed under ‘Design of EWS Vital Sign Charts’ below). |
| Christofidis et al. (2016) | A 2x2x2x2 mixed factorial design trial of vital sign chart tested experimentally on 205 novice chart users.  Participants responded faster (scores present and absent) and made fewer errors (scores absent) using drawn-dot (vs. written-number) observations and an integrated colour-based (vs. non-integrated tabular) scoring-system. Participants responded faster using grouped (vs. separate) scoring-rows when scores were absent, but separate scoring-rows when scores were present. | Experimental observation should be considered when designing vital sign charts. |

## Studies of Maternity Inpatient EWS/Vital Sign Charts

### Background: Maternity EWS Tools

Several of the obstetric EWS in use are summarised in Table 4. Obstetric EWS systems may use a single vital sign trigger (eg MEWC, MERC) or may activate based on a combination of vital signs resulting in a score (eg MEOWS) or incorporate both single and aggregate parameters (NZMEWS), as well as requiring persistence of these abnormalities over time (eg MEWT). General EWS such as NEWS and MEWS have also been studied in obstetric populations. There is conflicting opinion about the benefit of the simplicity in the single parameter alerts versus the potential for improved detection in the aggregate systems (Blumenthal 2019).

**Table 4**: A range of obstetric-specific EWS evaluated in the studies obtained for this review. In some of the studies described below these tools were tested head-to-head with general EWS (including NEWS, MEWS), or electronic EWS (eg eCART).

|  | **NZMEWS**  NZ Maternity Early Warning Score | **MEWC**  Maternal Early Warning Criteria | **MERC**  Maternal Early Recognition Criteria | **MEOWS**  Modified Early Obstetric Warning Score | **MEWT**  Maternal Early Warning Trigger |
| --- | --- | --- | --- | --- | --- |
| Description | Aggregate-weighted scoring system, including single vital sign triggers. | Single parameter scoring system (one vital can trigger). | As per MEWC with addition of temperature >38.5C as a trigger (as per Blumenthal et al. 2019). | Aggregate scoring tool based on NEWS, with single triggers | Detailed algorithm, designed to identify 4 of the major causes of maternal morbidity.  *Triggers must be maintained for 20min.* |
| Triggers | Score ≥5  Any blue trigger requires RRT | 1 trigger | 1 trigger | Score ≥5 (varies by context)  1 red or 2 yellow triggers | 2 triggers >20min or 1 severe trigger |
| RR | B ≥31  R 26–30  Or 21–25  **10–20**  R 6–9  B ≤5 | >30  **10–30**  <10 | >30  **10–30**  <10 | R >30  Y 21–30  **10–20**  R <10 | Severe >30  >24  **12–24**  <12 |
| O2 L/min | **Room air**  Or On oxygen | NA | NA | NA | NA |
| Sat O2 | **≥95**  Or 92–94  ≤91 | **≥95**  <95 | **≥95**  <95 | **≥95**  <95 | **≥94**  ≤93  Severe <90 |
| HR | B ≥140  R 130–139  Or 120–129  Y 100–119  **60–99**  Y 50–59  R 40–49  B <40 | >120  **50–120**  <50 | >120  **50–120**  <50 | R >120  Y 100–120  **50–100**  Y 40–50  R <40 | >110  **50–120**  <50 |
| SBP | B ≥200  R 160–199  Or 140–159  **100–139**  Y 90–99  Or 80–89  R 70–79  B <70 | >160  **90–160**  <90 | >160  **90–160**  <90 | R >160  Y 150–160  **100–160**  Y 90–100  R <90 | >155  **80–155**  <80 |
| DBP | R ≥110  Or 90–109  **<90** | >100  **≤100** | >100  **≤100** | NA | >105  **45–105**  <45 |
| *Mean arterial BP* | NA | NA | NA | NA | ***≥55***  *Severe <55* |
| *Urine output* | NA | *<35 mL/h(for ≥2h)* | *<35 mL/h(for ≥2h)* | NA | NA |
| *Fetal HR* | NA | NA | NA | NA | *>160 if infection suspected* |
| Temp | R ≥39  Y ≥38–<39  **≥36–<38**  Y ≥35–<36  R <35 | NA | >38.5  **<38.5** | R >38  **36–38**  Y 35–36  R <35 | >38  **36–38**  <36 |
| LoC | **Normal**  Abnormal | **Normal**  Abnormal  OR  *Headache/SOB if preeclampsia* | **Normal**  Abnormal  OR  *Headache/SOB if preeclampsia* | Normal  Y voice  R pain | **Normal**  Altered |

**Bold** indicates normal parameters. *Italic* items are not part of current NZMEWS. B = blue, R = red, Or = orange, Y = yellow. In the NZMEWS scoring is attributed as follows: B: call RRT, R=3, Or=2, Y=1, with NZMEWS ≥5 suggesting possible acute illness or unstable chronic disease. Note that the thresholds used in some tools appear to vary in different contexts and studies of the same tool.

### Review Articles

**See Appendix Table A3 for summary details of studies described below.**

**Smith et al. (2021)**

There was one Cochrane review obtained, however this high-quality systematic review only included two studies, both were in low resource settings, and one study assessed an automatic device that only recorded BP and HR. The findings of this Cochrane review are out of scope for the present review, the study is not tabulated here.

**Umar et al. (2019)**

This systematic review aimed to define the predictive accuracy of obstetric EWS for morbidity and mortality, and their effectiveness in triggering corrective actions and improving health outcomes. It included women in labour, post-partum, or sick pregnant women. Outcomes of interest were death, ICU transfer, and morbidity. 17 studies were included. Of these 17, all studies published since 2016, that were not in low-resource settings, had already been obtained by the present review and are discussed below (though one multi-country study was excluded because it was specific to pre-eclampsia). Sensitivity, specificity, PPV, and NPV of the various obstetric EWS are reported (and compiled in Table 8 below).

Regarding variables included, 14 of the 16 EWS vital sign charts had heart rate, SBP, and RR, 13 charts had temperature and 12 charts had DBP and conscious level. On average four in five charts identified in Umar et al.’s review contained these 5 parameters.

Results indicate that in the general obstetric population, EWS was shown to be highly sensitive and specific in predicting morbidity and ICU admission, with comparatively low PPV. In critically ill obstetric patients, EWS are highly accurate in predicting maternal death, but there is limited evidence that they reduce maternal deaths. The authors conclude that obstetric EWS are highly sensitive and specific in predicting obstetric morbidity and ICU admission with relatively low, but acceptable PPV, and could be used in the general obstetric population.

**Bernstein et al. (2021)**

This was a scoping review focused on EWS use for *women in labour*. The aim was to evaluate EWS sensitivity and specificity to correctly identify women in need of care escalation. Eleven studies were included in the review. Four of the 10 studies described under 'Scoring Tools' on p.258 were published since 2016 and all 4 were already obtained in the present review (see below). Results showed a wide range of sensitivity and specificity, possibly due to the lack of consensus on normal vital sign ranges during labour, and the varying context of use as some hospitals used tools only for women deemed to be at greater risk to decrease the rate of false positives and address clinician concerns about alarm fatigue. This different context (versus the original validated use) may change the sensitivity and specificity of a tool.

Results extracted by Bernstein et al. from the individual studies included the following five specific scoring tools (note most of these were not head-to-head comparisons):

* **MEOWS**: showed 89% sensitivity and 79% specificity for morbidity (broadly defined) in 676 consecutive women in labour.
* **MEWT:** had 96.9% sensitivity and 99.9% specificity for ICU admission.
* **MEWC**: has 97% sensitivity and 39% specificity in predicting morbidity for women with a single trigger, and sensitivity 84% and specificity 62% with multiple or recurrent triggers.
* **MEWS vs MEOWS vs MERC vs MEWT** (head-to-head): By comparing 79 women with morbidity to 123 women without morbidity as controls, Blumenthal et al. (2019) found the MEWS and MEWT to be more specific (93.5% and 88.6% respectively) than the MEOWS (51.2%) or the MERC (60.2%; p < 0.001) for predicting morbidity. The MEOWS and MERC were more sensitive (67.1% and 67.1%, respectively) for predicting morbidity than the MEWS or MEWT (19.0% and 40.5%, respectively; p < 0.001), but had very high alert rates.

**Quinn et al. (2016)**

In a non-systematic review Quinn et al. describe the history and development (but not validation) of an obstetric EWS based on the NEWS. Recording of level of consciousness was replaced by DBP as this is an essential measurement in the detection of preeclampsia. The resulting ObsEWS measured respiratory rate, oxygen saturations, temperature, SBP, DBP, and heart rate. The article cites now dated literature and is more of an opinion piece.

### Head-to-head Studies of Maternity EWS

**See Appendix Table A4 for summary details of studies described below.**

**Arnolds et al. (2022)**

This study was an assessment of comparative performance of general and obstetric early warning scores for predicting deterioration and infection on maternal wards. The study was based on EHR data. It included 19,611 ante- and post-natal ward patients, so results may not apply to patients in labour or the immediate postpartum period. The study evaluated multiple obstetric-specific and general EWS head-to-head. Results showed that eCART had the highest discrimination for deterioration (p < 0.05 for all comparisons), with an AUC of 0.86, followed by MEOWS (0.74), NEWS (0.72), MEWC (0.71), MEWS (0.70), and MEWT (0.65). MEWC, MEWT, and MEOWS had higher accuracy than MEWS and NEWS but lower accuracy than eCART at specific cut-off thresholds (see Table 5 below).

For predicting infection, eCART (AUC 0.77) had the highest discrimination. Results support use of maternal EWS (eg MEOWS) rather than MEWS or NEWS on maternity wards. Additionally, RR was the best predictor of the individual variables with AUC 0.72 (95% CI 0.70–0.74), followed by creatinine (0.70 (95% CI 0.68–0.73), heart rate (0.68 (95% CI 0.65–0.71), and systolic blood pressure (0.67 (95% CI 0.65–0.70). Note that Bernstein et al. (above) found RR to be often omitted from recordings.

**Table 5**: Accuracy of early warning scores for predicting ICU transfer and/or death of ante and post-natal women at different score thresholds (Adapted from Arnolds et al. 2022, Supplementary Material, *BMC Pregnancy and Childbirth*)

| **EWS** | **Threshold** | **Sensitivity** | **Specificity** | **PPV** | **NPV** |
| --- | --- | --- | --- | --- | --- |
| MEWS | ≥2 | 59.5% | 69.2% | 0.3% | 99.9% |
|  | ≥3 | 44.1% | 93.2% | 1.0% | 99.9% |
|  | ≥4 | 27.5% | 98.6% | 3.0% | 99.9% |
|  | ≥5 | 14.8% | 99.7% | 6.7% | 99.9% |
| NEWS | ≥3 | 55.9% | 83.0% | 0.5% | 99.9% |
|  | ≥4 | 43.4% | 92.3% | 0.9% | 99.9% |
|  | ≥5 | 33.7% | 97.0% | 1.7% | 99.9% |
|  | ≥6 | 25.9% | 98.9% | 3.4% | 99.9% |
|  | ≥7 | 18.9% | 99.6% | 7.2% | 99.9% |
| MEOWS | Two yellow triggers | 20.2% | 91.2% | 0.4% | 99.9% |
|  | One red trigger | 51.5% | 94.9% | 1.5% | 99.9% |
|  | Two yellow triggers and/or one red trigger | 61.4% | 86.9% | 0.7% | 99.9% |
| MEWC | One abnormal parameter | 53.3% | 88.9% | 0.7% | 99.9% |
| MEWT | Two non-severe triggers | 4.9% | 99.3% | 1.0% | 99.9% |
|  | One severe trigger | 27.5% | 98.8% | 3.5% | 99.9% |
|  | Two non-severe triggers and/or one severe trigger | 31.4% | 98.1% | 2.5% | 99.9% |
| eCART | ≥0.0035 | 67.3% | 87.4% | 0.8% | 99.9% |
|  | ≥0.0045 | 55.5% | 93.0% | 1.2% | 99.9% |
|  | ≥0.0053 | 45.6% | 95.6% | 1.6% | 99.9% |
|  | ≥0.0060 | 40.9% | 97.0% | 2.1% | 99.9% |
|  | ≥0.0069 | 35.5% | 98.1% | 2.9% | 99.9% |
|  | ≥0.0075 | 32.9% | 98.6% | 3.5% | 99.9% |
|  | ≥0.0079 | 31.9% | 98.8% | 4.1% | 99.9% |
|  | ≥0.0090 | 27.5% | 99.3% | 5.6% | 99.9% |
|  | ≥0.0108 | 23.6% | 99.6% | 9.0% | 99.9% |

**Blumenthal et al. (2019)**

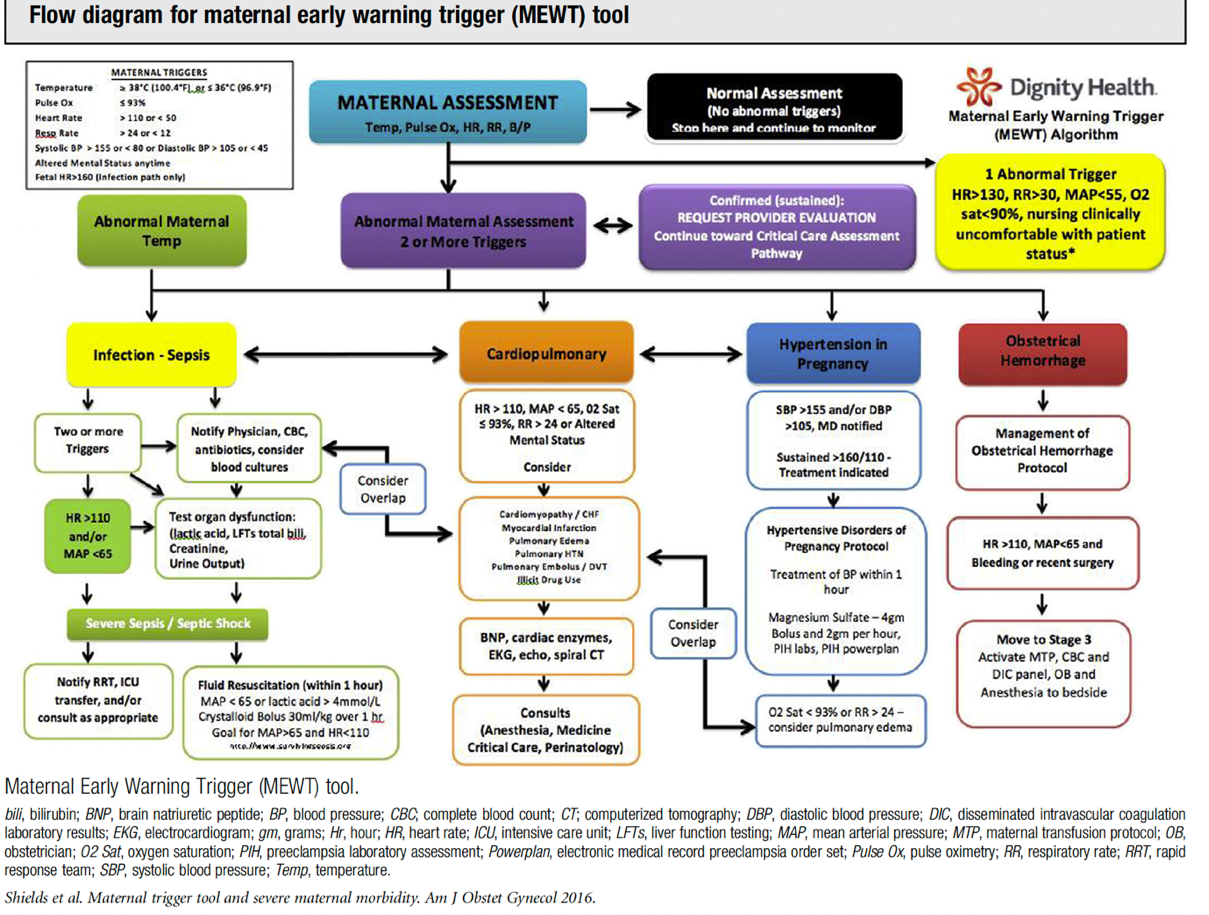
Already mentioned in the review by Bernstein et al. above, this retrospective case-control study compared 4 different EWS (MEWS, MEWT, MERC, MEOWS) head-to-head in the same population of 79 women experiencing severe maternal morbidity and 123 controls. MEOWS and MERC alerted in over 60% of the study population. None of the tools performed with 90% sensitivity and 95% specificity (see Table 8 below). The sensitivity of all systems improved if cases that were deemed 'emergencies without warning' were excluded from analysis. Frequent alerts included MEOWS ‘red’ (or single) alert with DBP > 90 mm Hg, also MERC alerts for oxygenation saturation less than 95% (note the findings of the 4P Study in Table 6 below suggesting both these thresholds may not be permissive enough), followed by SBP < 90 or > 160 mm Hg. The authors conclude that in this population at a high-volume delivery unit, the characteristics of MEWT best meet needs. MEWT's positive predictive value in the study population (70%) exceeded MEOWS (47%) and MERC (52%) and makes it more useful in the context despite lower sensitivity. A higher alert rate might be acceptable in a less busy environment, if desire to avoid false negatives outweighs the burden of a greater rate of false positives.

**Kern-Goldberger et al. (2022)**

This study was a retrospective cohort study testing the predictive ability of maternal EWS and their component vital sign variables for different kinds of maternal morbidity with a primary outcome defined as a composite of maternal morbidity including haemorrhage, infection, acute cardiovascular disease, and acute respiratory disease (all by ICD-10). The study cohort included 14,597 obstetric admissions, of which 2,451 patients experienced the composite morbidity outcome (16.8%) including 980 cases of haemorrhage (6.7%), 1,337 of infection (9.2%), 362 of acute cardiac disease (2.5%), and 275 of acute respiratory disease (1.9%). Results describe MEOWS, MERC and MEWT, and suggest that the clinical utility of vital signs to screen obstetric patients for developing morbidity is poor (see sensitivity and specificity in Table 9 below). MEWT had the lowest sensitivity (15.3%) and highest specificity (96.1%), PPV (44.5%), and LR+ (3.9) of the three scoring systems, and this was true for each subtype of morbidity. MERC consistently had the highest sensitivity (61.9–74.7%) and the lowest PPV (2.9–22.3%) and LR− (0.5–0.7). However, maternity EWS that use more stringent cut-off points may be most clinically relevant with high specificity able to discriminate patients who are well. The study tested different thresholds for HR, SBP, DBP and O2, and found that more stringent triggering thresholds increase the specificity of EWS. The trade-off between sensitivity and specificity are displayed in eg Table 6 and Table 8 in the study. For example, the following thresholds all have specificity >89%: DBP >110 or <40, SBP >160 or <80, O2 Sat <90%.

The authors conclude that all maternity EWS criteria demonstrated poor sensitivity for maternal morbidity and further refinement, or recalibration of parameters may be required for optimal clinical application. They noted that other studies found higher maternity EWS sensitivities, but this is likely related to non-standard definitions of maternal morbidity as an outcome (vs ICD-10 use here).

**Figure 7**: MEWT Algorithm (note the MEWT is quite different to most single trigger or aggregate weighted scoring tools with its inclusion of management pathways) (Figure credit: Shields et al. 2016, *American Journal of Obstetrics and Gynecology*)



### Studies of Physiology in Pregnancy

Two studies obtained for this review provided details of normal physiological values in pregnancy (labour may be another matter and Bernstein et al. (above) stated there is no consensus on normal values in labour)

**Loerup et al. (2019)** conducted a systematic review and meta-analysis of reports of vital signs in pregnancy, including analysis of SBP, DBP and heart rate by gestation. Results varied by gestational age. The authors note that there was no evidence of substantial decreases in blood pressure mid-pregnancy, also heart rate increases were lower than previously thought. They note that manual and automated blood pressure measurement cannot be used interchangeably. Also, DBP has increased over the last half-century. Key findings are tabulated below in comparison to the 4P Study and the NZMEWS ‘white’ vital sign ranges.

**Green et al. (2020)** reported the results of the 4P Study, a multicentre observational study of 1,401 pregnant UK women. Vital sign data from all participants were included in the primary analysis and smoothed centiles were reported for systolic and diastolic BP, heart rate, SpO2, temperature and respiratory rate by gestational age. The data are gestation specific so are curves, rather than fixed medians. This has implications for which stage of pregnancy an EWS is targeting. Examples of median values are presented in Table 6.

**Table 6**: Normal physiological values in pregnancy from Green et al. (2020) & Loerup et al. (2019) compared to the ‘white’ range used in NZMEWS.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **NZMEWS range (white band)** | **Green et al (2020) median or 50th centile (3rd–97th centile)** | **Loerup et al (2019)**  **mean (95% CI)** | **Comment** |
| RR | 10-20 | Median respiratory rate 15 (9–22), which did not change with gestation | - |  |
| O2 Sat | ≥95 | Median 97% (93–99%) at 40 weeks | - | 93 or 94% may be normal, authors suggest thresholds of <95% may be too high. |
| HR | 60-100 | Maximum median of  91 (68–115) bpm at 34.1 weeks | Mean 86.9 (82.2, 91.6) beats/min at 40 weeks | Note that 97th centile is substantially (15%) higher than the NZMEWS cut-off |
| SBP | 100-140 | Maximum median of 121(102–144) at 40 weeks | Mean 116.0 (113.6, 118.4) at 40 weeks | Authors hint that a threshold of <90 is too low, especially toward end of pregnancy, NZMEWS uses 100. |
| DBP | <90 | Maximum median 78 (62–95) mm Hg at 40 weeks | Mean 72.8 (71.0, 74.6) mmHg at 40 weeks | Refutes notion of a large BP drop in pregnancy, and authors urge users to abandon the notion that “relative hypotension from first BP in pregnancy may warrant investigation” |
| Temp | 36-38 | Minimum median of 36.5C (35.3–37.3) at 33.4 weeks | - | Green et al. (2020) state: “temperatures of >37.5°C were uncommon (fewer than 3% of observations), suggesting this may be an appropriate threshold to further investigate women.” |
| LoC | Normal | - | - |  |

### Additional Studies

Several other studies were obtained but not included in the tables above as the initial screening step ranked them ‘4’ or ‘5’ (ie it appeared that they provided lower quality evidence from smaller or single centre studies or provided additional validation of tools already assessed in systematic reviews). These studies are tabulated in Table 7 below and their possible implications noted.

**Table 7:** Seven additional studies of maternity EWS obtained through the literature searches but initially coded as lower quality or relevance – these included 6 studies of single EWS tools, plus a design audit.

| **Study** | **Description** | **Implications** |
| --- | --- | --- |
| Shields et al. (2016) | This was a prospective interventional study of maternity inpatients admitted to hospitals within a large US hospital system. Study included 36,832 deliveries at the pilot sites (24,221 pre and 12,611 post-MEWT testing) and 146,359 at the nonpilot sites (95,718 pre- and 50,641 post-MEWT testing) during the 2 study time periods. To be considered positive, MEWT triggers needed to be sustained for >20 minutes and were defined as severe (single abnormal value): maternal heart rate (HR) >130 beats/min (bpm), respiratory rate >30/min, mean arterial pressure <55 mm Hg, oxygen saturation <90%, or nurse concern; or nonsevere (required 2 abnormal values): temperature >38 or <36 C, blood pressure >160/110 or <85/45 mm Hg, HR >110 or <50 bpm, respiratory rate >24 or <10/min, oxygen saturation <93%, fetal HR >160 bpm, altered mental status, or disproportionate pain.  Results showed that use of MEWT resulted in significant reductions in CDC severe maternal morbidity (P < 0.01) and composite morbidity (P < 0.01). ICU admissions were unchanged. At nonpilot sites CDC severe maternal morbidity, composite morbidity, and ICU admissions were unchanged between baseline post-MEWT testing. Note that Umar et al. rated this a ‘low quality multicentre controlled trial’ in their systematic review (above). Note also that the MEWT tool includes guidance on management of different conditions. It is possible that improvements in management rather than detection improved outcomes. | Some evidence that use of sustained triggers may be associated with improved outcomes. |
| Arnolds et al. (2018) | This retrospective Chicago study examined a slightly modified MEWC, using a temperature threshold of 38.5C and did not include the subjective criteria. Original MEWC has good negative predictive value but low specificity. The aim was to increase its specificity by considering only patients with recurrent or multiple triggers. Thresholds used were Blood oxygen saturation (SpO2) on room air <95%; Oliguria <35mL/h, 2 hrs; Heart rate <50 or >120; SBP <90 or >160 mm Hg, DBP >100 mm Hg; RR <10 or >30.  In the intrapartum cohort of 400 women, 70% of women would have triggered MEWC ≥1 and 50% would have multiple triggers. Adding the requirement for recurrent or multiple triggers decreased sensitivity (from 0.97 to 0.84) and improved specificity from 0.39 to 0.62. | Although there was no comparison to other EWS, these findings suggest that performance gains, in terms of specificity, might be achieved by requiring sustained or multiple triggers, or relaxing triggering thresholds. |
| Ryan et al. (2017) | This small case-control study in Canada was the first non-UK validation of MEOWS. Population included Pregnant or recently delivered (<6 weeks) women, admitted for more than 24 hours, with 3 controls (n = 138) per ICU admission case (n = 46). Results showed a high-sensitivity and low specificity for two yellow triggers, but specificity markedly increased when using one ‘red’ rather than two ‘yellow’ (see Table below).  Four variables were significantly associated with ICU admission >24 hours: maximum temperature (b co-efficient 1.14; P = 0.005); heart rate (b 0.06; P < 0.001); SBP (b 0.05; P < 0.001); and respiratory rate (b 0.22; P = 0.001). The AUC for the four-variable equation was 0.91 (95% CI 0.83 to 0.95), with activation of 1 red or 2 amber triggers providing high sensitivity, specificity, and NPV. | The current threshold mandating a response may have been set too low and a modified ‘four variable’ MEOWS may better identify and focus early interventions (ie RR, HR, SBP, Temp). |
| Hedriana et al. (2016) | Small retrospective case-control study in California. Compared 50 obstetric patients requiring ICU admission with 50 controls. Six MEWTs were assessed. Two or more triggers were persistent for 30 minutes or more in 36 (72%) ICU patients versus 2 (4%) controls (OR 61.7, 95% CI 13.2–288.0).  Significant associations were recorded between ICU admission and tachycardia (OR 5.0, 95% CI 2.1–11.7), mean arterial pressure less than 65 mm Hg (OR 4.5, 95% CI 1.9–10.8), temperature of at least 38 °C (OR 44.1, 95% CI 13.0–839.1, p = 0.099, and altered mental state (OR 44.1, 95% CI 13.1–839.0).  The mean HR was 130.83± 16.34 beats per minute in the ICU group and 107.65±22.28 beats per minute in the control group.  Although the frequency of RR greater than 24 breaths per minute was higher in the ICU group, the difference was not significant (P=0.078) (Table 2). The frequency of SpO2 less than 94% did not differ significantly between the groups (P=0.076). | Two or more MEWTs often represents clinical deterioration and should mandate a more rigorous assessment of potential life-threatening complications.  HR had a lower OR (than temp or mental state) but note normal range (97th centile) goes up to 115 (see above), and a threshold of tachycardia >110 was used in the study.  Consider Temp 37.5C as the threshold since OR so high for 38C (though not statistically significant), and temp >37.5C is uncommon (see Green et al. 2020 above).  Evidence from this small study does suggest retaining LoC as a warning sign as OR was 44.1. |
| Hannola et al. (2021) | Finnish study of 828 high-risk women, in the first 24 hours on a post-natal ward. The sensitivity of the Obstetric early warning system at its best was 72% for pre-eclampsia, 52% for infection and 25% for postpartum haemorrhage. Single site, single tool study, limited by specific context and population. | Unclear implications for practice, as no comparison to other EWS, and study group limited to high-risk women 24-hours post-partum only. |
| Drake et al. (2021) | Before and after study of implementation of NZMEWS. Single site, single tool study. Following implementation of NZMEWS there was significant and sustained reduction in emergency response calls to women who were very unwell (emergency response team call), and a non-significant reduction in cardiorespiratory arrest team calls. | Unclear implications for practice as no comparison to other EWS, and potentially other explanations for reduction in response calls. |
| Isaacs et al. (2019) | Audit of 120 maternal EWS charts comparing them to the practice advice of Preece and Horswill. Many design flaws were noted commonly including inadequate patient identification, poor use of colour, illogical weighting, poor alignment and labelling of axes, and the opportunity for staff to ‘game’ the escalation. | Maternity EWS should adhere to established design principles. |

### Maternity EWS Test Performance

**Table 8**: Performance of maternity EWS in studies (published since 2016) identified in the review. Data displayed in terms of outcome studied, and test performance (ie sensitivity, specificity, AUC, PPV and NPV).

| **EWS Tool** | **Variables included** | **Study/outcome of interest** | **Sensitivity/specificity [AUC]** | **PPV\*/NPV** |
| --- | --- | --- | --- | --- |
| General EWS evaluated in obstetric settings | | | | |
| eCART | HR, RR, BP, Temp, mental status, O2 sat, supplemental oxygen, albumin, ALP, bilirubin, BUN, calcium, chloride, CO2, creatinine, glucose, Hb, platelet count, potassium, AST, sodium, total protein, WBC, age, ICU stays, ward hours | Arnolds (2022) – ICU/death | 45.6 / 95.6 (threshold ≥0.0053)  27.5 / 99.3 (threshold ≥0.0090)  [AUC: 0.86] | 1.6 / 99.9  5.6 / 99.9 |
| NEWS | RR, O2 sat, supplemental O2, HR, SBP, Temp, LoC | Arnolds (2022) – ICU/death | 43.4 / 92.3 (threshold ≥4)  [AUC: 0.72] | 0.9 / 99.9 |
| MEWS | RR, O2 sat, HR, SBP, Temp, LoC | Blumenthal (2019) – Morbidity (ICD-10)  Arnolds (2022) – ICU/death | 19.0 / 93.5  44.1 / 93.2 (threshold ≥3) | 1.0 / 99.9 |
| Maternity-specific EWS | | | | |
| NZMEWS | See Table 4 above | Drake (2021) – Response calls | NR | NR |
| MEWT | See Table 4 above | Shields (2016) – morbidity & ICU admission  Blumenthal (2019) – morbidity (ICD-10)  Hedriana (2016) - ICU admission  Arnolds (2022) - Death/ICU  Kern-Goldberger (2022) - morbidity | 96.9 / 99.9  40.5 / 88.6  72 / 96  31.4 / 98.1  [AUC: 0.65]  15.3 / 96.1 | 12.0 / 99.99  70 (PPV)  95 / 77  2.5 / 99.9  44.5 / 84.9 |
| MEWC | See Table 4 above | Arnolds (2018) – Morbidity  Arnolds (2022) – Death/ICU | Single trigger: 97 / 39  Multiple: 84 / 62  Single trigger: 53.3 / 88.9  [AUC: 0.71] | 0.34 / 0.97  0.42 / 0.92  0.7 / 99.9 |
| MERC | See Table 4 above | Blumenthal (2019) – Morbidity (ICD-10)  Kern-Goldberger (2022) - morbidity | 67.1 / 60.2  64.8 / 56.8 | 52 (PPV)  22.3 / 88.9 |
| MEOWS | See Table 4 above | Blumenthal (2019) – Morbidity (ICD-10)  Ryan (2017) - ICU > 24 hrs  Arnolds (2022) – Death/ICU  Kern-Goldberger (2022) – Morbidity  Hannola (2021) – Morbidity [modified MEOWS] | 67.1 / 51.2  96 / 54 (two yellow)  91 / 72 (one red)  61.4 / 86.9  [AUC: 0.74]  48.8 / 77.3  72 (best) / 83.2 | 47 (PPV)  41 / 97  52 / 96  0.7 / 99.9  30.4 / 88.2  32.2 (best) / 96.8 (best) |

Note that study methodology varies so the most useful comparisons may be within individual studies, eg Kern-Goldberger (2022) vs Kern-Goldberger (2022), or Arnolds (2022) vs Arnolds (2022).

\*Some high PPV values are reported based on case-control data, rather than general population sampling, these methods are not directly comparable.

## Design of EWS Vital Sign Charts

Sources obtained cited the vital sign chart design research of Preece & Horswill as underpinning good human factors recommendations. Much of this work, being pre-2016, was beyond the scope of the present review, however, the NZEWS is already informed by the work of Preece et al. (see Psirides & Pedersen’s proposal for a national EWS in 2015). A dearth of more recent publications on paper chart design may be due to most new user interfaces being electronic.

A survey of user experience of new vital sign charts by Elliot et al. (2016) found that:

* Users had preference for red or blue for highest alert level (but note this may not be optimal for those with red/green colour blindness).
* Nurses expressed concern that placement of a dot or ‘x’ on a chart meant that no exact values for vital signs may be noted.
* Participants’ overwhelming preference for a numerical value due to concern that parameter ranges were too wide to illustrate changes in a patient’s condition (however the ‘dot’ allows tracking of trend and reduces errors, see below)*.*
* The A3 chart size folded as a booklet was difficult to fit in bedside (A4 size) folders or to write on. Staff tended to fold the chart inside out or remove it from the folder, causing confusion over front and back. Front and back pages may need to look different.
* A bold line after ever 3 recordings confused users, even though it was intended to prevent ‘column shift’ errors, it resembled date boundaries.
* Users had difficulty identifying or targeting a patient’s usual blood pressure.
* Users need to be able to see the trend in a patient’s vital signs.
* One participant thought that an oxygen saturation of 95% coded as yellow was not warranted and mentioned recording it in the >95% to avoid having to report.
* Recording urine output was difficult to do consistently.

An experimental study by Chistofidis et al. (2016) investigated issues of chart design and concluded that participants responded faster (scores present and absent) and made fewer errors (scores absent) using drawn-dot (rather than written-number) observations and an integrated colour-based (rather than non-integrated tabular) scoring-system. Participants also responded faster using grouped (rather than separate) scoring-rows when scores were absent, but separate scoring-rows when scores were present. Considering these results, the authors argued that paper-based observation charts should employ drawn-dots and colour-based, rather than tabular, scoring-systems, with colour cues embedded in graphs to help identify criterion breaches rapidly.

In the UK, the NEWS2 chart was updated to overcome problems for those with red/green colour blindness (Williams 2022). Notably, the NZEWS vital sign chart does not use red or green but should be checked for colour issues. The order of vital signs was also updated in NEWS2 to reflect the ‘ABCDE’ Resuscitation Council priorities (airway, breathing, circulation, disability, exposure). The order of items is now: RR, SpO2, air/oxygen, SBP, HR, Consciousness, Temperature. Treacy et al. (2019) cites Preece et al. (2012) who note that ordering of vital signs on the chart influences how likely they are to be assessed. This ordering is also consistent with RR having the most predictive ability (as has often been reported), and temperature not contributing much to model discrimination (as shown in one small study described in the present review).

A design study specific to maternity EWS was conducted by Isaacs et al. (2019) who audited 120 maternal EWS charts comparing them to the practice advice of Preece and Horswill.[[2]](#footnote-3) Many design flaws were noted. Common issues included inadequate patient identification, poor use of colour, illogical weighting, poor alignment and labelling of axes, and the opportunity for staff to ‘game’ the escalation. The authors noted that maternity EWS should have a clear layout and style, appropriate colour scheme, correct language and labelling, and the ability for vital signs to be documented accurately and quickly.

Finally, the author of the present report noted (without supporting evidence) that the use of thresholds in the form ‘≤XXs’ for example ‘≤34s’ as used for temperature in the NZMEWS, could be misinterpreted as meaning ‘≤XX’. For example, ‘≤34s’ could be interpreted as ‘≤34’ rather than ‘between 34 and 35, OR less than or equal to 34’, which is what it means in context. The risk is that the action associated with the ‘red’ attribution might register consciously only if the score is <34. The issue is flagged for consideration.

## Discussion

A clinical early warning tool that is fit for purpose will predict what is wanted, be superior to other comparable options, be optimised (ie difficult to improve further), and be cost-effective or efficient to implement. However, it is important to specify what is wanted, because, for examples, *prediction of 24-hour ICU-admission* is not the same as *reduction of 30-day mortality*. Validation studies are highly heterogeneous (Fu et al. 2020), and this must be borne in mind when assessing the evidence. EWS policy needs to make a choice between ensuring that all cases at risk of deterioration are identified (ie few false negatives) and ensuring that resources are deployed only when there is something valuable for them to do (ie few false positives). The sector needs to decide on the goals of using EWS and titrate tool choice, triggers and score thresholds, and response plans to these goals. Only then can it be fully determined whether the tools used are fit for purpose.

### Adult

**Summary**

At a high level the evidence obtained for this review suggests that EWS can perform well in the general adult inpatient population and across most diagnostic groups, particularly for predicting mortality in the shorter term (<24 hours). Aggregate weighted EWS appear to perform better than activation criteria tools. The NEWS tool is widely studied and has been found by review studies to consistently outperform other manual or paper based EWS, although electronic EWS and those employing advanced analytics such as AI and ML appear more promising. Adding trend-based data did not greatly improve the performance of tools in the few studies identified. Compliance with EWS measurement and protocols has been observed to be poor in some settings. There may be a dose-response relationship between EWS protocol compliance and improved outcomes, with benefit observed to be highest in populations with comorbidities. Outcome benefits reported include reduced in-hospital cardiac arrests and reduced mortality. However, there is a lack of high-quality evidence and little data on the net benefit of using EWS.

**Why use an EWS?**

Studies of EWS use have found associations with reduced mortality and in-hospital cardiac arrests. This includes use of the electronic AAM tool (Escobar et al. 2020), the single trigger ‘between the flags’ tool (Bhonagiri et al. 2021), and the aggregate weighted score NEWS tool (Hogan et al. 2020, Haegdorens et al. 2019). Impact on patient outcome of a wider range of electronic tools was unclear (Mann et al. 2020), and cost-effectiveness or net-benefit analysis is needed (Gerry et al. 2020).

**EHR-integrated vs paper-based EWS**

The evidence collected indicates that EHR-integrated systems perform better than paper-based EWS tools (Bell et al. 2021, Hogan et al. 2020). Computational and aggregate weighted charts were better than unweighted activation trigger tools (Fu et al. 2020). Data driven (statistical/ML) complex computational scores perform best (Fu et al. 2020, Mann et al. 2021, Campbell et al. 2020, Bell et al. 2021). In particular, eCART outperformed manual vital sign charts (Campbell et al. 2020, Green et al. 2018) and eCART was superior to NEWS in head-to-head comparison (Kramer et al. 2019, Campbell 2020). The evaluation of electronic EWS was out-of-scope for this report, so no claims are made as to which electronic tools are optimal. For example, no research was obtained evaluating AAM vs eCART head-to-head.

Although the evaluation of electronic and advanced analytic tools was beyond the scope of this report, it appears that such tools are superior to paper-based and manual tools. A move to electronic vital sign recording and automatic EWS score generation should be considered, and this ought to be done in such a way that ensures a vital sign dataset is compiled for future analysis and tool refinement and adaptation to local clinical contexts, such as demographics and condition prevalence. Design of such electronic systems should take account of evidence about human factors and effective electronic interfaces.

**Which paper vital sign chart?**

If a manual, paper-based, vital sign EWS is to be used, then NEWS appears best overall in head-to-head comparisons (Liu et al. 2020, Campbell et al. 2020, Green et al. 2017, Holland & Kellett 2022), and has been widely studied (Gerry et al. 2020, Holland & Kellett 2022). NEWS was associated with improved outcomes compared to hospitals using other track and trigger systems (Hogan et a. 2020). NEWS had higher discrimination than MEWS and BTF for predicting cardiac arrest, ICU transfer, or death within 24 hours (Campbell 2020, Green 2018), and similar performance to Q-ADDS, which is not as widely studied (Campbell 2020). Additionally, the modifications in NEWS2 do not appear to have improved on NEWS (Kramer et al. 2019, Hwang et al. 2020, Holland & Kellett 2022, Tirkkonen et al. 2019). Indeed, NEWS2 may perform worse than NEWS (Pimental et al. 2019). NEWS2 did perform better than Q-ADDS in one study, though this study did not assess NEWS (Price 2023). Furthermore, NEWS has been prospectively and externally validated (Loisa et al. 2022).

The system used in New Zealand (NZEWS) is a manual, paper-based, hybrid of aggregate weighted and single trigger systems and is based on NEWS. Both approaches have evidence of association with reduced mortality from large registry studies, such as BTF in NSW (Bhonagiri et al. 2021) and NEWS in UK (Hogan et al. 2020). NZEWS has similar discrimination to NEWS for in-hospital cardiac arrest, unplanned ICU admission or death (Mohan et al. 2023).The similarity in performance between NZEWS and NEWS begs the question of whether anything is gained by including the single trigger elements (or indeed if costs are paid through increased sensitivity), rather than simply using NEWS.

If we assume transitivity (ie if a relation between three elements holds between the first and second and it holds between the second and third, then it holds between the first and third), then the research described above appears to show that: eCART > (NEWS = Q-ADDS) > BTF, and that NZEWS = NEWS, and therefore NZEWS > BTF.

**Variables to include in EWS charts**

NEWS, the best performing manual, paper based EWS tool, includes the variables heart rate, respiratory rate, blood pressure, temperature, mental status, oxygen saturation, supplemental oxygen.

eCART, the best performing (automated, electronic) tool identified in the review, includes the variables heart rate, respiratory rate, blood pressure, temperature, mental status, oxygen saturation, supplemental oxygen, albumin, alkaline phosphatase, bilirubin, blood urea nitrogen, calcium, chloride, CO2, creatinine, serum glucose, hemoglobin, platelet count, potassium, AST, sodium, total protein, WBC, age, ICU stays, and ward hours.

Respiratory rate was the vital sign that provides the best discrimination for deterioration (Loisa et al. 2022, Churpek et al. 2016). However, recording of RR could be improved, as it is poorly done (Holland & Kellett 2023).

Reducing the low SBP trigger threshold for RRT activation from 90 to 80mmHg was found to significantly reduce RRT activation with little loss in sensitivity (Campbell et al. 2020). Currently NZEWS ‘red zone’ is <90mmHg and RRT activation occurs at <70.

It is suggested that undertaking a neurological evaluation using the AVPU system may miss cases of confusion/delirium (Holland & Kellett 2023, Williams et al. 2022). However, NEWS2 was updated to include a confusion screen and the revised EWS probably doesn’t perform as well overall as NEWS.

Temperature was found to contribute little to model discrimination (Luis et al. 2018). In fact, high temperature was associated with improved survival (Sebat et al. 2020). However, the NZEWS currently down-weights temperature as this variable does not have any ‘red’ or ‘blue’ value thresholds.

Capillary refill time was second best predictor of mortality, cardiac arrest and higher-level of care transfer (behind hypoxia) among the set of 10-SOV variables (Sebat et al. 2020) and may be worth further research in the context of vital sign EWS.

Inclusion of demographic variables appeared to increase performance of EWS (Bell et al. 2021, Lee et al. 2018, Shamout et al. 2019, Rasmussen et al 2018).

Considering diagnosis group (Lee et al. 2018, Holland et al. 2023), or comorbidity index (Fu et al. 2020) can also aid risk stratification.

Inclusion of vital sign trends provided minor improvement in test performance (Churpek et al. 2016, Brekke et al. 2019, Bell et al. 2021), and adding FiO2 slightly improved discrimination (Malycha et al. 2019), but adding ‘Clinical Worry Score’ didn’t help much (Hwang et al. 2020).

Any changes to existing vital sign EWS chart inclusions and thresholds would necessitate revalidation of the tool, preferably in head-to-head comparison with the old version and other highly considered tools.

Changing to the electronic eCART system would probably provide greater performance gains than tweaking the current NZEWS vital sign chart with any of the above modifications.

**EWS score thresholds**

The single trigger BTF ‘yellow’ cut-off may be overly sensitive and result in poor specificity (Bell et al. 2021). On the other hand, the aggregate weighted score of NEWS is a specific tool but not so sensitive, with 45% of deaths having admission NEWS score < 3 (Holland et al. 2023). Given the proportion of patients who activate NEWS, it was suggested that NEWS cut-off ≥5 is appropriate (Haegdorens et al. 2020). However, the appropriate activation thresholds for various kinds of responses must be titrated to the available resources and should take into consideration PPV and NPV not just discrimination ability (AUC) (Alhmoud et al 2021). The accuracy of an early warning chart to predict morbidity is defined by the PPV and NPV and these depend on the prevalence of the condition of interest. It is generally desirable that a screening test should have a high sensitivity and specificity, but when there is low prevalence, high sensitivity and specificity will result in low PPV and a high NPV, and when prevalence is higher a high PPV and a low NPV will result (Umar et al. 2019). These will determine the number needed to evaluate, which will in turn determine the resource demands. Alert fatigue was also noted as a risk. For example, in one study 86% of automatic alerts NEWS alerts were ignored (Bedoya et al. 2019).

The NNE should be manageable. Evidence collected in this review provides guidance for these decisions (Holland & Kellett 2022, Green et al. 2017). The number of calls can be reduced according to graphs mapping score to proportion of vital sign sets triggering response. For example, Holland & Kellett (2022), described above, report that 24-hour mortality increased from 1.8% for a NEWS ≥3 to 7.8% for NEWS ≥7. Although 24-hour mortality for NEWS *<*3 was only 0.07% these deaths accounted for 9% of all deaths within 24-hours; for NEWS *<*7 24-hour mortality was 0.23%, which accounted for 44% of all 24-hour deaths. Within 30-days of a NEWS recording 22% of all deaths occurred in patients with a NEWS *<*3, 52% in patients with a NEWS *<*5, and 75% in patient with a NEWS *<*7.

The present report can’t recommend a threshold as such choices are dependent on priorities and capacity of sites. Some initial work has been completed in New Zealand on this issue, but is beyond the scope of this report (see eg Ullah et al. 2022).

**How often to measure vital signs?**

EWS vital sign recordings taken 12-hourly were associated with reduced mortality in post-hoc analysis of an RCT deploying NEWS (Haegdorens et al. 2019). Another study found no difference between 8 and 12-hour recording (Petersen et al. 2016). Authors of a systematic review to identify NEWS thresholds and levels of risk note that patients with NEWS <3 have only a 0.07% chance of dying within 24-hours and note that some have suggested 24-hourly observation might be reasonable in such cases (Holland & Kellet 2022). Regardless of how often EWS are measured, it should be noted that there is often poor compliance with EWS recording and staff should be facilitated and encouraged to correctly record the scores (Credland et al. 2018).

**Where/when to use EWS?**

The evidence collected indicates that NEWS performs best during hospital admission and for predicting 24-hour outcomes (Pankhurst et al. 2022, Holland & Kellett 2022, Brabrand et al. 2017), with predictive performance diminishing for longer timeframes. NEWS performs well across all wards and for undifferentiated patients (Alhmoud et al. 2021, Price et al. 2023, Holland & Kellett 2023). Performance appears best in general medical and surgical populations (Alhmoud et al. 2021, Holland & Kellett 2022). NEWS performance is moderate in the ED, but scores when first measured are not as predictive as those recorded subsequently during a patient’s stay (Alhmoud et al. 2021, Holland & Kellett 2022).

NEWS was found to perform less well in haematology, oncology, or gastrointestinal conditions (Alhmoud et al. 2021), or respiratory illness or sepsis (Holland & Kellett 2022) but nevertheless performed better than SIRS & qSOFA for sepsis (Holland et al. 2023, Liu 2020), see below. The very similar NEWS2 did not perform as well for cases of cardiac arrest and ventricular fibrillation, aortic & peripheral arterial embolism or thrombosis, and respiratory failure; insufficiency; or arrest (Price et al. 2023). NEWS was found to provide most benefit on wards with highest comorbidities (Haegdorens et al. 2019).

Regarding sepsis specifically, although NEWS didn’t predict adverse outcomes as well as in many other patient groups (Holland & Kellett 2022), it performed better than the sepsis-specific tools qSOFA and SIRS (Holland et al. 2023, Liu 2020).

* Alhmoud et al. (2021) included 33 studies in their systematic review that specified ‘sepsis’ as a disease subgroup and found that EWS prediction of sepsis had reasonable predictive performance in all subgroups (AUC: 0.71–0.79) and in infectious diseases specifically (AUC: 0.79).
* Liu et al. 2020 compared sepsis-specific tools qSOFA and SIRS to NEWS in head-to-head comparison and found NEWS superior to SIRS and qSOFA for predicting in-hospital mortality or ICU transfer at all of 28 hospitals, at specific decision thresholds, either by reducing the proportion of patients that need to be screened or increasing the proportion of adverse outcomes identified (outcomes of interest were in-hospital mortality and the combined outcome of ICU transfer or mortality).
* It appeared to matter when the EWS assessment was done, as discrimination of NEWS has been repeatedly found to be lower on arrival to hospital than when repeated later (Holland et al. 2022). Some conditions have very effective treatments, so general EWS may not accurately predict mortality (see eg Hamilton et al. 2018), when there is effective management that is yet to be implemented (eg antibiotics for sepsis), this does not mean that that EWS cannot identify patients who need to be treated urgently.
* NEWS ≥5 has been recommended as a possible flag for sepsis by Holland et al. (2022), who cite a pre-hospital study (beyond the scope of the present review) where NEWS was compared to qSOFA.

With condition-specific prediction tools, there is a bit of a horse or cart situation. If a user is thinking of using a sepsis-specific tool, for example, then they are already considering sepsis (and might just manage accordingly). Whereas when a general EWS tool is used, it may provide a flag that serious conditions such as sepsis might need to be considered.

**Response to EWS triggers**

The response (efferent) component of EWS is necessarily context dependent. Appropriate response to EWS triggers must consider the prevalence and severity of adverse conditions expected in the relevant population. Response guidance must also consider the resources and expertise available.

When looking to the literature for guidance on response protocols, the relevant evidence are studies that provide outcome data (not mere prediction validation studies). There are fewer such studies. However, large real-world audits of the use of BTF (Bhonagiri et al. 2021), NEWS (Hogan et al. 2020, Haegdorens et al. 2019), and AAM (Escobar et al. 2020) have demonstrated associations with improved outcomes. Given that NZEWS is largely similar to NEWS, the response approaches used in studies of NEWS are likely relevant.

In the case of Hogan et al. (2020), the intervention was use of NEWS at 106 different NHS hospitals, so particulars of response (efferent pathways) are likely to vary. That said, many UK hospitals will be using the RCP escalation guidance recommended by NICE. This guidance has been critiqued as ‘arbitrary’ and not grounded in an evidence base (Holland & Kellett 2023). As such it is possible that response protocols might be improved, and this is an area for further research.

The RCT by Haegdorens et al. (2019), which found improved outcomes only in post-hoc analysis, implemented a response protocol that included changes to observation frequency, lists of phone numbers, a maximum waiting time for medical assistance, and a backup procedure if the regular medical response was late or unavailable. This protocol did not include an RRT/MET (due to resource constraints), but the ordinary resuscitation team was to be contacted as usual. The SBAR communication protocol was an additional component that linked the afferent and efferent processes. The intervention is described in a book chapter Haegdorens et al. (2018a).

The intervention found to be associated with reduced IHCA in NSW (using the BTF system activation thresholds) included standardised calling criteria and minimum standards for escalation, including clinical review and thresholds for RRT activation, as described in Bhonagiri et al. (2021).

In Chicago, the AAM system was associated with reduced mortality. The intervention used remote monitoring nurses, who would see data of flagged patients, and conduct a chart review and communicate concerning results to the ward staff, and to rapid response teams in hospitals (Escobar et al. 2020).

Additionally, the cost and cost-effectiveness (and net benefit) of various response protocols (at various EWS trigger thresholds) should be evaluated. This has not yet been done.

### Maternity

**Summary**

The picture in maternity is less clear than for general adult inpatients, due to far fewer head-to-head comparisons of EWS tools, fewer external validation studies, and fewer consistently reported settings, populations, and outcomes of interest. It is also noteworthy that physiology changes during gestation, there are not agreed normal vital sign parameters during labour, and a proportion of pregnancies are in non-adults, who also may have different normal physiological parameters to adult populations. There is also a philosophical trade-off between casting the net wide with a low threshold to detect all possible adverse events in maternity care (thereby potentially minimising ‘never’ events but evaluating many false positives) versus reluctance to over investigate and medicalise pregnancy (and thereby being at risk of failing to identify serious adverse events in a healthy population that are experiencing a normal process). All these factors make inferring the optimal EWS strategy for maternity/obstetric populations more challenging than for the general adult inpatient population.

**Why use a maternity EWS?**

Maternity EWS are highly accurate at predicting maternal death (Umar et al. 2019). Use of maternity EWS improved the quality of monitoring, which may improve reaction time (Umar et al. 2019). Use of maternity EWS (MEWT) was associated with reduced maternal morbidity (Shields et al. 2016).

**EHR-integrated vs paper-based**

In head-to-head comparison an electronic (general) EWS system (eCART) outperformed manual maternity EWS as well as other general EWS when used in a maternity population to predict deterioration and infection (Arnolds et al. 2022). Reports identified in this review concluded that machine learning for maternity EWS should be further investigated (Arnolds 2022).

**Which paper vital sign chart?**

When considering only non-electronic manual systems, maternity specific EWS (MEOWS) outperformed general EWS (NEWS) for predicting deterioration and infection in a head-to-head study (Arnolds et al. 2022). In this study the discrimination performance (AUC) of the various tools was as follows: MEOWS > NEWS > MEWC > MEWS > MEWT (Arnolds et al. 2022). However, MEWC, MEWT, and MEOWS had higher accuracy than MEWS and NEWS (Arnolds et al. 2022). The most sensitive EWS was MEOWS at the ‘one red or two yellow’ threshold (61.4%), however with a PPV of only 0.7%. The most specific EWS was the general (non-maternity) MEWS at the ≥5 threshold (99.7%) with PPV of 6.7%. The electronic eCART outperformed all these, with a higher sensitivity and/or PPV at various thresholds. The literature noted that the above published sensitivities and specificities may be biased by non-standard definitions of morbidity (Kern-Goldberger et al. 2022).

On average the most specific test was MEWT. MEWT is the only maternity EWS that requires vital sign derangement to be sustained for 20 minutes and is also the maternity EWS found to be associated with reduced maternal morbidity as noted above. Given the performance of the various tools, a high-volume delivery unit determined that MEWT, with its higher specificity, might best meet needs while also limiting the number of alerts (Blumenthal et al. 2019). It is notable that the NZMEWS is not currently recommended for use in the birthing suites.

**Variables to include**

Most maternity EWS include at least HR, SBP, RR, temperature, DBP, and consciousness level (Umar et al. 2019). Maternity EWS that perform well include components, which are not included in the NZMEWS (see Table 4). For example, MEWT requires sustained vital sign abnormalities across 20 minutes, as well as including mean arterial BP, and fetal HR (where infection is suspected). It also has a more permissive low O2 sat threshold than NZMEWS (≤93%) and includes a minimum DBP (≤45). MEOWS does not include scoring of supplemental oxygen, but is otherwise similar to NZMEWS. The tools including oliguria and headache/SOB if preeclampsia (MEWC, MERC) were not the most specific, or discriminatory tests. High performing eCART includes a range of laboratory values and demographics, as discussed above, and listed in Table 8.

As with general adult EWS, RR was the best predictor of deterioration, this was followed by creatinine (Arnolds et al. 2022). A more limited set of variables (RR, HR, SBP, Temp) had better AUC for discriminating pregnant or recently delivered women who will require ICU admission than the full MEOWS (Ryan et al. 2017). However, level of consciousness probably ought to be retained given its high odds ratio of 44 for ICU admission when abnormal (Hedriana et al. 2016).

Some of the additional variables just described could be considered for inclusion in a NZ maternity EWS, and authors have suggested recurrent, sustained, or multiple disturbances of vital signs might be preferred to single triggers (Arnolds et al. 2018, Shields et al. 2016). However, any modifications to the tool would require validation of the new tool’s predictive ability, preferably head-to-head against other high-performing EWS.

**Maternity EWS score thresholds**

DBP >110mmHg or <40, SBP >160 or <80, and O2 sat <90% all had specificity of >89% for maternal morbidity (Kern-Goldberger et al. 2022). Furthermore, results of the 4P Study (of maternal physiology through stages of gestation) (Green et al. 2020) suggest that some maternity EWS variable thresholds could be too strict (eg O2 sat, HR) or permissive (eg temperature). For example:

* O2 sat threshold of <95% leads to frequent alerts (Blumenthal et al. 2019). A lower O2 sat threshold could be reasonable, given that 93-94% (40 weeks) may be considered normal (Green et al. 2020, Kern-Goldberger et al. 2022).
* HR threshold of 115 bpm (34 weeks) would equate with the 97th centile of normal pregnancy (Green et al. 2020) and such HR would score points in the NZMEWS. Additionally, HR had a lower odds ratio for ICU admission than some other vital signs (Hedriana et al. 2016).
* Temperatures over 37.5C were found in the 4P Study to be rare in pregnancy (Green et al. 2020), and temperature of >38C had an odds ratio of 44 for ICU admission (Hedriana et al. 2016), so temperature thresholds could be titrated to these observations.

Bernstein et al. reported that there is a lack of consensus on normal vital sign ranges in labour (Bernstein et al. 2021), this could be further investigated.

**Where/when to use maternity EWS?**

The unclear normal parameters for physiology of labour, and the changing normal ranges for vital signs during stages of pregnancy, mean that the scope for use of specific maternity EWS are not yet clear. Evidence-based implementation can only be informed by the studies available. Comparisons of EWS are particularly important because many tools have been validated in some context (pre- or ante-natal, in labour, etc) for some outcome (ICU, morbidity, death, other outcomes), but unless we test many tools in the same population at the same time, and agree on the outcome of interest, we don’t know how EWS performance compares. That said, tools like eCART outperform other EWS in general settings, and outperform maternity specific EWS in maternity settings. Looking to the future such automated and electronic EWS could possibly be considered as a universal general screening tool.

The choice of vital sign charts/maternity EWS systems should be based on the sensitivity–specificity trade-off, site-specific logistics considerations, and how many false positives can be managed given available resources (Arnold et al. 2022, Bernstein et al. 2021). There is probably no set of EWS criteria that performs with high sensitivity and high specificity and obstetric EWS need to choose between the two (Kern-Goldberger et al. 2022). Rather than a screening tool for morbidity, obstetric EWS might best leverage high specificity to identify those not at risk.

## Recommendations

### Answers to the 5 Research Questions

Full justification for the following short responses, with many additional details and options, can be read in the ‘Discussion’ above.

Q1. **What vital sign parameters are validated to predict acute deterioration (death, cardiac arrest or admission to a higher level of care) in adult hospital inpatients?**

A number of vital sign based EWS tools have been validated to predict a range of measures of acute deterioration, including 24 hour mortality, unplanned ICU admission, and in-hospital cardiac arrest. The NEWS tool is one of the most studied and generally best performing of manual, paper-based EWS, and its subsequent iteration (NEWS2) does not appear to have improved performance. Performance is best in general patient settings and following admission, for outcomes within 24 hours. The NEWS is, however, outperformed by electronic tools such as eCART, which include variables such as demographics and laboratory values. Within the parameters measured by vital sign EWS, respiratory rate has the highest discrimination for adverse events.

Q2. **What vital sign parameters are validated to predict acute deterioration (death, cardiac arrest or admission to a higher level of care) in hospital inpatients who are currently pregnant or post-partum?**

Vital sign EWS are less studied in pregnant and post-partum populations, though a systematic review concluded that various tools have good sensitivity and specificity in predicting obstetric morbidity and ICU admission with relatively low, but acceptable PPV. The parameters commonly included in such tools are HR, SBP, RR, temperature, DBP, and consciousness level. The highest specificity was obtained by the MEWT tool, which requires that vital sign derangements persist across 20 minutes. However, the general population electronic eCART tool had higher discrimination for maternal morbidity than maternity specific tools in head-to-head evaluation. Performance of the various tools in various studies is presented in Table 8 of this report, and comparative performance in the same study population in Table 5.

Q3. **What other clinical parameters should be documented on vital sign charts?**

* In addition to the variables included in the NZEWS the following variables may have relevance for predicting patient deterioration:
  + Assessment of confusion/delirium (not part of the AVPU measure).
  + Capillary refill time (second best predictor of deterioration in one study).
  + Demographic variables (advocated to increase performance in several studies).
  + Diagnostic group (reported to enhance risk stratification in multiple studies).
  + Comorbidity index (found to aid risk stratification in one study).
  + Additionally, the following non-vital sign measures inform the high performing eCART score: albumin, alkaline phosphatase, bilirubin, blood urea nitrogen, calcium, chloride, CO2, creatinine, serum glucose, hemoglobin, platelet count, potassium, AST, sodium, total protein, WBC, age, ICU stays, and ward hours.
* In addition to the variables included in the NZMEWS the following variables may have relevance for predicting maternal deterioration:
  + Mean arterial blood pressure (included in MEWT)
  + Fetal heart rate (where infection is suspected, included in MEWT)
  + Oliguria (included in multiple tools, though not the best performing)
  + Headache/SOB (included in multiple tools, though not the best performing)

Q4. **What are the features of good vital sign chart design?**

* Vital sign chart design should follow the principles laid out by Preece & Horswill.
* Experimental evidence supports drawn-dots and colour-based, rather than tabular, scoring-systems, with colour cues embedded in graphs.
* Charts should be clear to those with colour blindness, especially red/green colour blindness.
* The ordering of items may influence which vital signs are recorded.
* Developers of NEWS2 settled on ordering vital signs according to the UK Resuscitation Council ‘ABCDE’ approach.
* Additional design principles will likely apply to electronic implementation of vital sign early warning charts.

Q5. **What are the key elements of effective escalation pathways and response systems?**

*This item was agreed at an initial meeting to be a low priority for focus as it is hospital resource dependent – this is borne out in the literature where several reviews state that response pathways must necessarily be context dependent*.

* Where improvement in adult inpatient outcomes has been found, implementation of the EWS ‘efferent’ arm tended to adhere to a graded response system where mild derangements trigger increased monitoring, moderate derangements trigger medical review, and severe derangements trigger review by those with critical care expertise (as per RCP/NICE advice).
* Communication protocols such as SBAR, relevant phone numbers, and alternative courses of action if there is no immediate response have also been included in a successful programme, as has remote monitoring by nurses.

### Additional Recommendations

Based on the findings discussed above, the following could also be considered.

#### Adult

* Continue to use an aggregate weighted scoring tool (eg NZEWS) in preference to single trigger only tools.
* NEWS (and hence NZEWS) has been demonstrated appropriate for use across most patient populations, including sepsis.
* Clearly determine the goals of EWS implementation, as this will guide which thresholds of NEWS (and hence NZEWS) to focus on and which graded response protocol is appropriate.
* Consider further investigating advanced analytic tools such as eCART, AAM, and others, as these appear to outperform manual, paper-based tools.
* Consider EHR integration of existing EWS because electronic implementation was associated with better outcomes than paper-based implementation.
* Consider further evaluating the additional variables listed under research Q3 above, though any modifications to existing EWS will require re-validation, ideally in head-to-head fashion.
* Conduct a net-benefit and/or cost-effectiveness analysis of EWS and RRT at various thresholds of response activation.

#### Maternity

* Clearly articulate the goal of maternity EWS implementation, as the choice of tool and triggering thresholds depends on what outcome is desired.
* No specific maternity EWS can be strongly recommended, as it depends on the goals, and therefore test characteristics desired.
* Consider electronic EWS as eCART outperformed maternity-specific EWS.
* Consider further investigating the requirement for sustained or multiple triggers to limit number of false positives (MEWT tool takes this approach, has high specificity and has been associated with improved maternal outcomes).
* Consider evaluating more permissive thresholds for HR, and O2 saturation before scoring points or triggering RRT, given the normal ranges found in the 4P Study.
* Consider evaluating a more stringent upper bound on temperature as >37.5C was found to be rare in pregnancy and 38C had a very high (though not statistically significant) odds ratio for ICU admission in one study.

## Limitations of this Study

This study has some limitations. Firstly, it was not a systematic review, but rather a time-limited scoping review, which aimed to find the most relevant evidence in the time available, not necessarily all the evidence. It is expected that there will be evidence that exists but is not described in this report. Additionally, grey literature was not searched as it was deemed out-of-scope from the outset. Finally, the overall quality of literature and evidence sourced tended to be low, which means that these findings should be seen as provisional, pending the publication of results from larger, or more methodologically robust research and evaluation.

## References

Alhmoud, B., Bonnici, T., Patel, R., Melley, D., Williams, B., & Banerjee, A. (2021). Performance of universal early warning scores in different patient subgroups and clinical settings: a systematic review. *BMJ Open, 11*(4), e045849. doi:10.1136/bmjopen-2020-045849

Arnolds, D. E., Carey, K. A., Braginsky, L., Holt, R., Edelson, D. P., Scavone, B. M., & Churpek, M. (2022). Comparison of early warning scores for predicting clinical deterioration and infection in obstetric patients. *BMC Pregnancy Childbirth, 22*(1), 295. doi:10.1186/s12884-022-04631-0

Arnolds, D. E., Smith, A., Banayan, J. M., Holt, R., & Scavone, B. M. (2019). National Partnership for Maternal Safety Recommended Maternal Early Warning Criteria Are Associated With Maternal Morbidity. *Anesth Analg, 129*(6), 1621-1626. doi:10.1213/ane.0000000000003889

Bedoya, A. D., Clement, M. E., Phelan, M., Steorts, R. C., O'Brien, C., & Goldstein, B. A. (2019). Minimal Impact of Implemented Early Warning Score and Best Practice Alert for Patient Deterioration. *Crit Care Med, 47*(1), 49-55. doi:10.1097/ccm.0000000000003439

Bell, D., Baker, J., Williams, C., & Bassin, L. (2021). A Trend-Based Early Warning Score Can Be Implemented in a Hospital Electronic Medical Record to Effectively Predict Inpatient Deterioration. *Crit Care Med, 49*(10), e961-e967. doi:10.1097/ccm.0000000000005064

Bernstein, S. (2021). Scoping Review on the Use of Early Warning Trigger Tools for Women in Labor. *J Obstet Gynecol Neonatal Nurs, 50*(3), 256-265. doi:10.1016/j.jogn.2021.01.003

Bhonagiri, D., Lander, H., Green, M., Straney, L., Jones, D., & Pilcher, D. (2021). Reduction of in-hospital cardiac arrest rates in intensive care-equipped New South Wales hospitals in association with implementation of Between the Flags rapid response system. *Intern Med J, 51*(3), 375-384. doi:10.1111/imj.14812

Blumenthal, E. A., Hooshvar, N., McQuade, M., & McNulty, J. (2019). A Validation Study of Maternal Early Warning Systems: A Retrospective Cohort Study. *Am J Perinatol, 36*(11), 1106-1114. doi:10.1055/s-0039-1681097

Brabrand, M., Hallas, P., Hansen, S. N., Jensen, K. M., Madsen, J. L. B., & Posth, S. (2017). Using scores to identify patients at risk of short term mortality at arrival to the acute medical unit: A validation study of six existing scores. *Eur J Intern Med, 45*, 32-36. doi:10.1016/j.ejim.2017.09.042

Brekke, I. J., Puntervoll, L. H., Pedersen, P. B., Kellett, J., & Brabrand, M. (2019). The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. *PLoS One, 14*(1), e0210875. doi:10.1371/journal.pone.0210875

Campbell, V., Conway, R., Carey, K., Tran, K., Visser, A., Gifford, S., . . . Churpek, M. (2020). Predicting clinical deterioration with Q-ADDS compared to NEWS, Between the Flags, and eCART track and trigger tools. *Resuscitation, 153*, 28-34. doi:10.1016/j.resuscitation.2020.05.027

Christofidis, M. J., Hill, A., Horswill, M. S., & Watson, M. O. (2016). Observation chart design features affect the detection of patient deterioration: a systematic experimental evaluation. *J Adv Nurs, 72*(1), 158-172. doi:10.1111/jan.12824

Credland, N., Dyson, J., & Johnson, M. J. (2018). What are the patterns of compliance with Early Warning Track and Trigger Tools: A narrative review. *Appl Nurs Res, 44*, 39-47. doi:10.1016/j.apnr.2018.09.002

Drake, M., Austin, G., Dann, L., Li, Y., Shuker, C., & Psirides, A. (2021). Introduction of a standardised maternity early warning system: indicative data from a before-and-after study at a large pilot site before national rollout in Aotearoa New Zealand. *Anaesthesia, 76*(12), 1600-1606. doi:10.1111/anae.15557

Elliott, D., Allen, E., McKinley, S., Perry, L., Duffield, C., Fry, M., . . . Roche, M. (2016). User acceptance of observation and response charts with a track and trigger system: a multisite staff survey. *J Clin Nurs, 25*(15-16), 2211-2222. doi:10.1111/jocn.13303

Escobar, G. J., Liu, V. X., Schuler, A., Lawson, B., Greene, J. D., & Kipnis, P. (2020). Automated Identification of Adults at Risk for In-Hospital Clinical Deterioration. *N Engl J Med, 383*(20), 1951-1960. doi:10.1056/NEJMsa2001090

Fang, A. H. S., Lim, W. T., & Balakrishnan, T. (2020). Early warning score validation methodologies and performance metrics: a systematic review. *BMC Med Inform Decis Mak, 20*(1), 111. doi:10.1186/s12911-020-01144-8

Fu, L. H., Schwartz, J., Moy, A., Knaplund, C., Kang, M. J., Schnock, K. O., . . . Rossetti, S. C. (2020). Development and validation of early warning score system: A systematic literature review. *J Biomed Inform, 105*, 103410. doi:10.1016/j.jbi.2020.103410

Gerry, S., Bonnici, T., Birks, J., Kirtley, S., Virdee, P. S., Watkinson, P. J., & Collins, G. S. (2020). Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. *Bmj, 369*, m1501. doi:10.1136/bmj.m1501

Green, L. J., Mackillop, L. H., Salvi, D., Pullon, R., Loerup, L., Tarassenko, L., . . . Watkinson, P. J. (2020). Gestation-Specific Vital Sign Reference Ranges in Pregnancy. *Obstet Gynecol, 135*(3), 653-664. doi:10.1097/aog.0000000000003721

Green, M., Lander, H., Snyder, A., Hudson, P., Churpek, M., & Edelson, D. (2018). Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. *Resuscitation, 123*, 86-91. doi:10.1016/j.resuscitation.2017.10.028

Haegdorens, F., Monsieurs, K. G., De Meester, K., & Van Bogaert, P. (2020). The optimal threshold for prompt clinical review: An external validation study of the national early warning score. *J Clin Nurs, 29*(23-24), 4594-4603. doi:10.1111/jocn.15493

Haegdorens, F., Monsieurs, K. G., De Meester, K., & Van Bogaert, P. (2019). An intervention including the national early warning score improves patient monitoring practice and reduces mortality: A cluster randomized controlled trial. *J Adv Nurs, 75*(9), 1996-2005. doi:10.1111/jan.14034

Haegdorens, F., Van Bogaert, P., Roelant, E., De Meester, K., Misselyn, M., Wouters, K., & Monsieurs, K. G. (2018). The introduction of a rapid response system in acute hospitals: A pragmatic stepped wedge cluster randomised controlled trial. *Resuscitation, 129*, 127-134. doi:10.1016/j.resuscitation.2018.04.018

Haegdorens, F., Monsieurs, K., & Van Bogaert, P. (2018a). Standardizing Care Processes Using Evidence-Based Strategies: Implementation of a Rapid Response System in Belgian Hospitals. In P. Van Bogaert & S. Clarke (Eds.), The Organizational Context of Nursing Practice: Concepts, Evidence, and Interventions for Improvement (pp. 239-257). Cham: Springer International Publishing.

Hamilton, F., Arnold, D., Baird, A., Albur, M., & Whiting, P. (2018). Early Warning Scores do not accurately predict mortality in sepsis: A meta-analysis and systematic review of the literature. *J Infect*, 76(3), 241-248. doi:10.1016/j.jinf.2018.01.002

Hannola, K., Hoppu, S., Mennander, S., Huhtala, H., Laivuori, H., & Tihtonen, K. (2021). Obstetric early warning system to predict maternal morbidity of pre-eclampsia, postpartum hemorrhage and infection after birth in high-risk women: a prospective cohort study. *Midwifery, 99*, 103015. doi:10.1016/j.midw.2021.103015

Hedriana, H. L., Wiesner, S., Downs, B. G., Pelletreau, B., & Shields, L. E. (2016). Baseline assessment of a hospital-specific early warning trigger system for reducing maternal morbidity. *Int J Gynaecol Obstet, 132*(3), 337-341. doi:10.1016/j.ijgo.2015.07.036

Hogan, H., Hutchings, A., Wulff, J., Carver, C., Holdsworth, E., Nolan, J., . . . Black, N. (2020). Type of Track and Trigger system and incidence of in-hospital cardiac arrest: an observational registry-based study. *BMC Health Serv Res, 20*(1), 885. doi:10.1186/s12913-020-05721-5

Holland, M., Dannatt, A., Kellett, J., & Green, D. (2023). Emergency admissions' diagnoses and risk of in-hospital death according to the primary ICD-10 chapter assigned at discharge and the National Early Warning Score on admission. *Acute Med, 22*(3), 113-119.

Holland, M., & Kellett, J. (2022). A systematic review of the discrimination and absolute mortality predicted by the National Early Warning Scores according to different cut-off values and prediction windows. *Eur J Intern Med, 98*, 15-26. doi:10.1016/j.ejim.2021.12.024

Holland, M., & Kellett, J. (2023). The United Kingdom's National Early Warning Score: should everyone use it? A narrative review. *Intern Emerg Med, 18*(2), 573-583. doi:10.1007/s11739-022-03189-1

Hwang, J. I., & Chin, H. J. (2020). Relationships between the National Early Warning Score 2, clinical worry and patient outcome at discharge: Retrospective observational study. *J Clin Nurs, 29*(19-20), 3774-3789. doi:10.1111/jocn.15408

Isaacs, R., Smith, G., Gale-Andrews, L., Wee, M., van Teijlingen, E., Bick, D., & Hundley, V. (2019). Design errors in vital sign charts used in consultant-led maternity units in the United Kingdom. *Int J Obstet Anesth, 39*, 60-67. doi:10.1016/j.ijoa.2019.01.001

Kern-Goldberger, A. R., Ewing, J., Polin, M., D'Alton, M., Friedman, A. M., & Goffman, D. (2023). The Predictive Value of Vital Signs for Morbidity in Pregnancy: Evaluating and Optimizing Maternal Early Warning Systems. *Am J Perinatol, 40*(14), 1590-1601. doi:10.1055/s-0041-1739432

Lee, Y. S., Choi, J. W., Park, Y. H., Chung, C., Park, D. I., Lee, J. E., . . . Moon, J. Y. (2018). Evaluation of the efficacy of the National Early Warning Score in predicting in-hospital mortality via the risk stratification. *J Crit Care, 47*, 222-226. doi:10.1016/j.jcrc.2018.07.011

Liu, V. X., Lu, Y., Carey, K. A., Gilbert, E. R., Afshar, M., Akel, M., . . . Churpek, M. M. (2020). Comparison of Early Warning Scoring Systems for Hospitalized Patients With and Without Infection at Risk for In-Hospital Mortality and Transfer to the Intensive Care Unit. *JAMA Netw Open, 3*(5), e205191. doi:10.1001/jamanetworkopen.2020.5191

Loerup, L., Pullon, R. M., Birks, J., Fleming, S., Mackillop, L. H., Gerry, S., & Watkinson, P. J. (2019). Trends of blood pressure and heart rate in normal pregnancies: a systematic review and meta-analysis. *BMC Med, 17*(1), 167. doi:10.1186/s12916-019-1399-1

Loisa, E., Kallonen, A., Hoppu, S., & Tirkkonen, J. (2022). Ability of the National Early Warning Score and its respiratory and haemodynamic subcomponents to predict short-term mortality on general wards: a prospective three-centre observational study in Finland. *BMJ Open, 12*(4), e055752. doi:10.1136/bmjopen-2021-055752

Luís, L., & Nunes, C. (2018). Short National Early Warning Score - Developing a Modified Early Warning Score. *Aust Crit Care, 31*(6), 376-381. doi:10.1016/j.aucc.2017.11.004

Malycha, J., Farajidavar, N., Pimentel, M. A. F., Redfern, O., Clifton, D. A., Tarassenko, L., . . . Watkinson, P. J. (2019). The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis. *Resuscitation, 139*, 192-199. doi:10.1016/j.resuscitation.2019.04.002

Mann, K. D., Good, N. M., Fatehi, F., Khanna, S., Campbell, V., Conway, R., . . . Cook, D. (2021). Predicting Patient Deterioration: A Review of Tools in the Digital Hospital Setting. *J Med Internet Res, 23*(9), e28209. doi:10.2196/28209

McGaughey, J., Fergusson, D. A., Van Bogaert, P., & Rose, L. (2021). Early warning systems and rapid response systems for the prevention of patient deterioration on acute adult hospital wards. *Cochrane Database Syst Rev, 11*(11), Cd005529. doi:10.1002/14651858.CD005529.pub3

Mohan, C., Entezami, P., John, S., Hewitt, J., Sylevych, V., & Psirides, A. (2023). Comparison of the Aotearoa New Zealand Early Warning Score and National Early Warning Score to predict adverse inpatient events in a vital sign dataset. *Anaesthesia, 78*(7), 830-839. doi:10.1111/anae.16007

Nestor, C. C., Donnelly, M., Connors, S., Morrison, P., & Boylan, J. (2022). Early warning scores and critical care transfer - patient heterogeneity, low sensitivity, high mortality. *Ir J Med Sci, 191*(1), 119-126. doi:10.1007/s11845-021-02558-7

Pankhurst, T., Sapey, E., Gyves, H., Evison, F., Gallier, S., Gkoutos, G., & Ball, S. (2022). Evaluation of NEWS2 response thresholds in a retrospective observational study from a UK acute hospital. *BMJ Open, 12*(2), e054027. doi:10.1136/bmjopen-2021-054027

Pedersen, N. E., Rasmussen, L. S., Petersen, J. A., Gerds, T. A., Østergaard, D., & Lippert, A. (2018). A critical assessment of early warning score records in 168,000 patients. *J Clin Monit Comput, 32*(1), 109-116. doi:10.1007/s10877-017-0003-5

Petersen, J. A., Antonsen, K., & Rasmussen, L. S. (2016). Frequency of early warning score assessment and clinical deterioration in hospitalized patients: A randomized trial. *Resuscitation, 101*, 91-96. doi:10.1016/j.resuscitation.2016.02.003

Pimentel, M. A. F., Redfern, O. C., Gerry, S., Collins, G. S., Malycha, J., Prytherch, D., . . . Watkinson, P. J. (2019). 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, 134*, 147-156. doi:10.1016/j.resuscitation.2018.09.026

Preece, M. H. W., Hill, A., Horswill, M. S., & Watson, M. O. (2012). Supporting the detection of patient deterioration: Observation chart design affects the recognition of abnormal vital signs. *Resuscitation*, 83(9), 1111–1118. https ://doi.org/10.1016/j.resuscitation.2012.02.009

Price, C., Prytherch, D., Kostakis, I., & Briggs, J. (2023). Evaluating the performance of the National Early Warning Score in different diagnostic groups. *Resuscitation, 193*, 110032. doi:10.1016/j.resuscitation.2023.110032

Quinn, A. C., Meek, T., & Waldmann, C. (2016). Obstetric early warning systems to prevent bad outcome. *Curr Opin Anaesthesiol*, 29(3), 268-272. doi:10.1097/aco.0000000000000338

Rasmussen, L. J. H., Ladelund, S., Haupt, T. H., Ellekilde, G. E., Eugen-Olsen, J., & Andersen, O. (2018). Combining National Early Warning Score With Soluble Urokinase Plasminogen Activator Receptor (suPAR) Improves Risk Prediction in Acute Medical Patients: A Registry-Based Cohort Study. *Crit Care Med, 46*(12), 1961-1968. doi:10.1097/ccm.0000000000003441

Ryan, H. M., Jones, M. A., Payne, B. A., Sharma, S., Hutfield, A. M., Lee, T., . . . von Dadelszen, P. (2017). Validating the Performance of the Modified Early Obstetric Warning System Multivariable Model to Predict Maternal Intensive Care Unit Admission. *J Obstet Gynaecol Can, 39*(9), 728-733.e723. doi:10.1016/j.jogc.2017.01.028

Sebat, C., Vandegrift, M. A., Oldroyd, S., Kramer, A., & Sebat, F. (2020). Capillary refill time as part of an early warning score for rapid response team activation is an independent predictor of outcomes. *Resuscitation, 153*, 105-110. doi:10.1016/j.resuscitation.2020.05.044

Shamout, F., Zhu, T., Clifton, L., Briggs, J., Prytherch, D., Meredith, P., . . . Clifton, D. A. (2019). Early warning score adjusted for age to predict the composite outcome of mortality, cardiac arrest or unplanned intensive care unit admission using observational vital-sign data: a multicentre development and validation. *BMJ Open, 9*(11), e033301. doi:10.1136/bmjopen-2019-033301

Shields, L. E., Wiesner, S., Klein, C., Pelletreau, B., & Hedriana, H. L. (2016). Use of Maternal Early Warning Trigger tool reduces maternal morbidity. *Am J Obstet Gynecol, 214*(4), 527.e521-527.e526. doi:10.1016/j.ajog.2016.01.154

Singh, S., McGlennan, A., England, A., & Simons, R. (2012). A validation study of the CEMACH recommended modified early obstetric warning system (MEOWS). *Anaesthesia*, *67*(1), 12-18. doi:10.1111/j.1365-2044.2011.06896.x

Smith, V., Kenny, L. C., Sandall, J., Devane, D., & Noonan, M. (2021). Physiological track-and-trigger/early warning systems for use in maternity care. *Cochrane Database Syst Rev*, 9(9), Cd013276. doi:10.1002/14651858.CD013276.pub2

Tirkkonen, J., Karlsson, S., & Skrifvars, M. B. (2019). National early warning score (NEWS) and the new alternative SpO(2) scale during rapid response team reviews: a prospective observational study. *Scand J Trauma Resusc Emerg Med, 27*(1), 111. doi:10.1186/s13049-019-0691-6

Treacy, M., & Caroline Stayt, L. (2019). To identify the factors that influence the recognizing and responding to adult patient deterioration in acute hospitals. *J Adv Nurs*, 75(12), 3272-3285. doi:10.1111/jan.14138

Umar, A., Ameh, C. A., Muriithi, F., & Mathai, M. (2019). Early warning systems in obstetrics: A systematic literature review. *PLoS One, 14*(5), e0217864. doi:10.1371/journal.pone.0217864

Ullah, E., Albrett, J., Khan, O., Matthews, C., Perry, I., GholamHosseini, H., & Lu, J. (2022). Workload involved in vital signs-based monitoring & responding to deteriorating patients: A single site experience from a regional New Zealand hospital. *Heliyon*, 8(10), e10955. doi:10.1016/j.heliyon.2022.e10955

Williams, B. (2022). The National Early Warning Score: from concept to NHS implementation. *Clin Med (Lond), 22*(6), 499-505. doi:10.7861/clinmed.2022-news-concept

## Appendix: Core Papers of the Review

(studies rated 1, 2, or 3 in the screening process, see ‘Methods’)

### Table A1: Adult (Reviews)

| **First author surname** | **Link** | **Year** | **Study type** | **Population/ Setting** | **Intervention/EWS studied** | **Comparator** | **Outcome** | **Main Finding** | **Implications** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| McGaughey | https://pubmed.ncbi.nlm.nih.gov/34808700/ | 2021 | Review, Cochrane systematic - up to 2020 (updates a 2007 review) | Acute hospital wards | EWS & RRS | Implementation of EWS and RRS in acute hospital wards compared to ward settings without EWS and RRS | Mortality, unplanned ICU admissions, length of hospital stay or adverse events | 4 RCTs (only 1 meets our inclusion criteria, included below)  7 non-randomised studies (all with critical to serious risk of bias)  Conclusions were drawn from the randomised evidence. Low-certainty evidence that EWS and RRS may lead to little or no difference in hospital mortality, unplanned ICU admissions, length of hospital stay or adverse events; and moderate-certainty evidence of little to no difference on composite outcome | High quality systematic review. Strict inclusion criteria. But did not analyse predictive validity of EWS. |
| Alhmoud | https://pubmed.ncbi.nlm.nih.gov/36044371/ | 2021 | Review, systematic | Hospital inpatients from different clinical settings, including ED, medical, ICU, and surgical settings, with various disease subgroups such as sepsis. | EWS to predict deterioration | Predictive validation among various EWS | Mortality, intensive care unit (ICU) transfer and cardiac arrest | 103 studies included.  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. | EWS may be most reliable in ED/medical and other non-specialised settings |
| Holland | https://pubmed.ncbi.nlm.nih.gov/34980504/ | 2022 | Review, systematic | Patients in various healthcare settings such as hospital admissions, EDs, emergency admissions, pre-hospital, acute medical units, and ICU, many observations from UK patients and ED patients in the USA. | EWS to predict deterioration | ViEWS vs NEWS vs NEWS2 | Death at different prediction time windows (eg 24 hours, 30 days) | 121 papers identified, the average AUC for mortality declined from 0.90 at 24-hours to 0.76 at 30-days. All studies with more than 50,000 observations had an AUC ≥0.880 at 24 hours (NEWS, ViEWS, and NEWS2) | Best validity for predicting mortality by NEWS/NEWS2 is 24 hours. Probably doesn't matter which of these EWS is used. Triggering threshold score must necessarily be arbitrary, but decision relevant guidance is given. |
| Fu | https://pubmed.ncbi.nlm.nih.gov/32278089/ | 2020 | Review, systematic | Acute care settings, general ward admissions and medical admissions, excluding ED. | Various EWS including single threshold, aggregate score, and electronic EWS | Validation against other EWSs, external validation. EWS compared using AUC or PPV metrics | Mortality, intensive care unit (ICU) transfer and cardiac arrest were most common | 29 studies included. The characteristics of the cohort, predictors, and outcome selection, as well as the metrics for model validation, vary greatly across EWS studies. Aggregate scores are better than single thresholds. Poor PPV reduces model usefulness and is seldom reported. | Keep using aggregate score EWS, analyse the PPV to set thresholds. |
| Gerry | https://pubmed.ncbi.nlm.nih.gov/32434791/ | 2020 | Review, systematic | General adult patients admitted to hospitals | Various EWS | Methodological assessment of studies describing the development of EWS | Death, cardiac arrest, ICU, or composite were most common, various prediction horizons | 11 studies were development only, 23 were development and external validation, and 61 were external validation only. Many EWS have methodological weaknesses. EWS might not perform as well as expected | EWS without further validation could have a detrimental effect on patient care (though note subsequent validation study of Loisa et al. reported above) |
| Brekke | https://pubmed.ncbi.nlm.nih.gov/30645637/ | 2019 | Review, systematic | Patients hospitalized with acute illness. Only 2 studies included: Hospitals in Illinois, USA, and a Canadian regional hospital | Trends in intermittently monitored vital sign variables; ViEWS | Evaluating different vital sign trends to predict clinical deterioration in hospitalized patients. | Death, cardiac arrest, ICU transfer | A lack of research in intermittently monitored vital sign trends. The included studies, although heterogeneous and imprecise, indicated some added value of trend analysis. | Using trend data for intermittently recorded vital signs may have minor improvement over current value |
| Credland | https://pubmed.ncbi.nlm.nih.gov/30389058/ | 2018 | Review, narrative | Adult patients admitted to acute hospital general wards in the United Kingdom, Denmark, and Amsterdam. | Aggregate weighted track and trigger systems used for monitoring patient deterioration and linked to a referral protocol | Not applicable | Studies which presented quantitative date measuring compliance with early warning scoring systems | Three key themes were identified, early warning score calculation accuracy, monitoring frequency and clinical response. This review identifies poor compliance with the Early Warning Score (EWS) protocol in all three themes. | Narrative review that offers insight into why EWS might not always help |
| Mann | https://pubmed.ncbi.nlm.nih.gov/34591017/ | 2021 | Review, systematic | Patients in general wards in hospitals, excluding those in ICU, ED, or single diagnosis patient groups. | EHR and digital EWS based on physiologic data | Predictive validation and outcomes of implementation among various digital EWS | AUC, ICU admission, cardiac arrest, death | Despite relative progress in the development of algorithms to predict patient deterioration, the literature has not shown that the deployment or implementation of such algorithms is reproducibly associated with improvements in patient outcomes. | Digital implementation of EWS and use of advanced analytics tools has potential but remains understudied |
| Fang | https://pubmed.ncbi.nlm.nih.gov/32552702/ | 2019 | Review, systematic | Adults admitted to hospitals, multiple countries including the UK, USA, and several other countries globally | Various EWS, the most reviewed EWS were the MEWS and NEWS. | Different methodologies and performance metrics used in the validation of EWS in various studies. | Mortality, ICU transfers, or cardiac arrest | Methodologies and performance metrics used in studies performing validation on EWS were heterogeneous making it difficult to interpret and compare EWS performance.  Standardizing EWS validation methodology and reporting can potentially address this issue. | It can be difficult to directly compare validation data across studies of EWS |
| Holland | https://pubmed.ncbi.nlm.nih.gov/36602553/ | 2023 | Review, narrative | Patients predominantly in acute UK hospitals, including ED, with a focus on various care settings and conditions. | National Early Warning Score (NEWS) | Different NEWS thresholds, NEWS2 | Discrimination of NEWS at various thresholds | How and who should respond to  different NEWS levels is uncertain and may vary according to the clinical setting and resources available. In the ED, simple triage scores which are quicker and easier to use may be more appropriate determinants of acuity. However, any alternative to NEWS should be easier and cheaper to use and provide evidence of outcome improvement. | Efferent arm needs to be tailored to likely conditions and resources in context-specific fashion |
| Williams | https://pubmed.ncbi.nlm.nih.gov/36427887/ | 2022 | Opinion, narrative | Development and implementation of NEWS/NEWS2 across the UK | National Early Warning Score (NEWS) |  |  | The NHS is the first healthcare system in the world to have a system-wide common language for acute clinical illness severity and clinical deterioration. |  |

### Table A2: Adult (Other Studies)

| **First author surname** | **Link** | **Year** | **Study type** | **Population/Setting** | **Intervention/EWS studied** | **Comparator** | **Outcome** | **Main Finding** | **Implications** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Haegdorens | https://pubmed.ncbi.nlm.nih.gov/29679694/ | 2018 | RCT, stepped wedge cluster | Hospital inpatients admitted to medical and surgical wards in Belgian acute care hospitals from October 2013 to May 2015, excluding patients below 17 years of age and pregnant patients. Control group 34,267 patient admissions, intervention group 35,389 | NEWS used as part of a standard observation and comms protocol | Control wards where EWS not yet deployed | Death, cardiac arrest or unplanned ICU admission | NEWS was consistently recorded in the intervention group. The intervention had no significant effect on the incidence of unexpected death rates (1.5 vs 0.7/1000, OR 0.82, 95%CI 0.34–1.95), cardiac arrest rates (1.3 vs 1.0/1000, OR 0.71, 95%CI 0.33–1.52) or unplanned ICU admissions (6.5 vs 10.3/1000, OR 1.23, 95%CI 0.91–1.65), when adjusted for clustering and study time | Unclear implications as NEWS not the only component of the intervention |
| Haegdorens | https://pubmed.ncbi.nlm.nih.gov/31012124/ | 2019 | RCT, stepped wedge cluster | Post-hoc analysis of the above RCT | NEWS used as part of a standardised observation and communication protocol | Control wards where EWS not yet deployed | Compliance with the NEWS protocol and unexpected death | Protocol compliance was negatively associated with patient mortality adjusted for comorbidity and age, combined mortality rate (r = −0.364, p = 0.080), unexpected death rate (r = −0.451, p = 0.027) | There may be a dose-response relationship between compliance with EWS and outcomes. |
| Petersen | https://pubmed.ncbi.nlm.nih.gov/26891861/ | 2016 | RCT | Acutely admitted surgical and medical patients (n = 1346) with an initial EWS of 0 or 1 at surgical and medical acute care wards at Bispebjerg University Hospital in Copenhagen, Denmark | EWS based on NEWS | The comparator in the study was the monitoring frequency of EWS measurements every 12 hours (vs 8 hours). | Clinical deterioration 24 h post-admission, estimated by the proportion of patients with an EWS ≥2 at 24 h after the initial EWS on admission. | No significant interactions for the primary outcome and the predefined variables: age, gender, ward type, and inclusion period, with an adjusted OR 1.20 (0.79–1.82). There were no significant differences regarding the secondary outcomes: cardiac arrests, ICU admissions, review by medical emergency team (MET), length of hospital stay, or elevated EWS at 48 hr. | Probably nothing to take from this study with high attrition and soft endpoint. |
| Hogan | https://pubmed.ncbi.nlm.nih.gov/32948171/ | 2020 | Observational study, retrospective, registry | Hospital level data from 106 acute NHS hospitals, National Cardiac Arrest Audit, inpatients who experienced in-hospital cardiac arrest (IHCA) | NEWS | NEWS vs non-NEWS TTS; Paper TTS vs electronic TTS. | In hospital cardiac arrest | 2009-2015 incidence of IHCA fell. Introduction or use of NEWS was associated with reduction of 9.4% in the rate of ward based IHCA compared to non-NEWS systems (incidence rate ratio 0.906, p < 0.001). Electronic TTS was also associated with a reduction of 9.8% in the rate of IHCA compared with paper-based TTS (incidence rate ratio 0.902, p = 0.009). There was no change in hospital survival. | NEWS appears to be better than other TTS, electronic EWS appears to be better than paper based. |
| Bhonagiri | https://onlinelibrary.wiley.com/doi/10.1111/imj.14812 | 2021 | Observational study, prospective registry | 35 New South Wales public hospitals with an ICU, patients experiencing in-hospital cardiac arrest | Between the Flags' (BTF) program, a two-tier rapid response system implemented in New South Wales, Australia. | Before and after implementation rates of cardiac arrests | Incidence of in-hospital cardiac arrests and the incidence and outcome of patients admitted to an ICU following cardiac arrest | The cardiac arrest rate per 1000 hospital admissions declined from 0.91 in the implementation period to 0.70. Propensity score analysis showed significant declines in ICU and hospital mortality and length of stay for cardiac arrest patients admitted to the ICU (all P < 0.001). | Single trigger systems may have promise, but no other EWS comparator |
| Bedoya | https://pubmed.ncbi.nlm.nih.gov/30247239/ | 2019 | Observational study, retrospective | The study population consisted of adult inpatient admissions (age greater than or equal to 18 years) on surgical and medical wards from a large academic hospital and a community hospital. | NEWS implemented in the EHR to automatically trigger a Best Practice Advisory (BPA) at NEWS ≥7 | Patient outcomes before and after implementation of NEWS. | Changes in mortality and ICU transfer rates | The primary outcome of rate of ICU transfer or death did not change after NEWS implementation, with adjusted HRs of 0.94 (0.84, 1.05) and 0.90 (0.77, 1.05) at the academic and community hospital respectively. | Not likely to inform practice, only threshold of 7 tested and 86% of warnings ignored |
| Brabrand | https://pubmed.ncbi.nlm.nih.gov/28986155/ | 2017 | Observational study, retrospective | Adult patients (15+ years) arriving at an acute medical unit at a 450-bed regional teaching hospital in Denmark, with a median age of 67 years and 50.4% male participants. | RAPS, REMS, NEWS, Goodacre physiological score, Groarke physiological score, Worthing physiological score | Different Early Warning Scores (EWS) based on vital signs for predicting mortality | 24-hour mortality and in-hospital mortality | Discriminatory power for 24 h mortality was above 0.8 for all scores (except the Groarke score (0.587)) and highest for the Worthing score (0.847). Discriminatory power for predicting overall in hospital mortality was highest for the Goodacre and Worthing scores (0.810 and 0.800 respectively) but below 0.8 for the remaining scores. | Hard to generalise as only the vital sign set on admission was studied. |
| Campbell | https://pubmed.ncbi.nlm.nih.gov/32504769/ | 2020 | Observational study, retrospective | The study population consists of adult patients (age 18 and older) hospitalized on the medical-surgical wards at the University of Chicago from November 2008 to November 2018. The median age of the study population was 57 years, with a majority being female and African American. | Q-ADDS, NEWS, Between-the-Flags (BTF), and eCART. | Comparison among the various EWS | Clinical deterioration outcomes such as ICU transfer or death | Q-ADDS (AUC 0.71) and NEWS (AUC 0.72) had similar predictive accuracy, BTF (AUC 0.64) had the lowest, and eCART (AUC 0.76) the highest. | Aggregate scoring systems appear to outperform BTF, eCART outperforms manual paper-based EWS. |
| Green | https://pubmed.ncbi.nlm.nih.gov/29169912/ | 2018 | Observational study, retrospective | All admissions to 5 hospitals in Illinois, all patients, 2008-13 | eCART, NEWS, MEWS, BTF | Comparison among the various EWS | Cardiac arrest, ICU transfer, or death within 24 hours of an observation | Overall accuracy was highest for eCART with an AUC of 0.801 (95% CI 0.799–0.802), followed by NEWS, MEWS and BTF respectively (0.718 [0.716–0.720]; 0.698 [0.696–0.700]; 0.663 [0.661–0.664]) | BTF generates a lot of false positives, may be better to use NEWS (without the single trigger thresholds? Need a health economic analysis) |
| Holland | https://pubmed.ncbi.nlm.nih.gov/37746679/ | 2023 | Observational study, retrospective | Adult patients (aged 16 years and older) who were emergency admissions to Salford Royal Hospital between 2014 and 2022. Excluding patients in the ED, covers all wards within the hospital. | National Early Warning Score (NEWS) | NEWS <3 on admission compared to NEWS ≥3, also diagnostic ICD-10 chapter. | In hospital mortality | 60% of in-hospital deaths were in four ICD-10 chapters, infections, circulatory and respiratory diseases, or neoplasms. NEWS ≥3 was associated with 8-fold increase in mortality. However, 45% of all in-hospital deaths occurred in patients with an admission NEWS <3 | Single site study finds ICD-10 chapter relevant to risk stratification |
| Liu | https://pubmed.ncbi.nlm.nih.gov/32427324/ | 2020 | Observational study, retrospective | Hospitalized inpatients in 21 California and 7 Illinois hospitals, with a mean age of 61.3-65.1 years and 53.9% women. | NEWS, MEWS, BTF, qSOFA, and SIRS | Comparison among the various EWS | In-hospital mortality and the combined outcome of ICU transfer or mortality using area under the receiver operating characteristic curves (AUCs). | The NEWS exhibited the highest discrimination for mortality (AUC, 0.87; 95%CI, 0.87-0.87 in California vs AUC, 0.86; 95%CI, 0.85-0.86 in Illinois), | NEWS appears to be better than other EWS, including single trigger EWS and sepsis specific EWS |
| Mohan | https://pubmed.ncbi.nlm.nih.gov/36991498/ | 2023 | Observational study, retrospective | Hospital inpatients from six hospitals within the Canterbury District Health Board, including both elective and emergency admissions, categorized as 'medical' or 'surgical' based on the specialty of their admitting medical service. | New Zealand Early Warning Score (NZEWS) | The comparator is the UK National Early Warning Score (NEWS) | AUC analysis to discriminate patients at risk of serious adverse events | The study found that NZEWS was similar to NEWS in predicting adverse outcomes. | Need to investigate PPV, NPV NEWS vs NZEWS, does NZEWS gain anything from its 'BTF' approach? |
| Pimentel | https://pubmed.ncbi.nlm.nih.gov/30287355/ | 2019 | Observational study, retrospective | Adult inpatients admitted to five acute hospitals from two UK NHS Trusts, specifically those who were not fit enough to be discharged alive on the day of admission. | NEWS and NEWS2 | Comparison among the two EWS, performance was evaluated in different risk groups, including patients with documented T2RF and those at risk of T2RF. | AUC analysis, with primary and secondary outcomes related to patient outcomes like in-hospital death, cardiac arrest, and ICU admission. | NEWS2 modifications to NEWS do not improve discrimination of adverse outcomes in patients with documented T2RF and decrease discrimination in patients at risk of T2RF. Further evaluation of the relationship between SpO2 values, oxygen therapy and risk should be investigated further before wide-scale adoption of NEWS2 | Does not support switch from NEWS to NEWS2 modifications. |
| Tirkkonen | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6915867/ | 2019 | Observational study, prospective | Adult Rapid Response Team (RRT) patients without limitations of medical treatment (LOMT) in a large Finnish university-associated tertiary level hospital. | NEWS and NEWS2 | NEWS without the new SpO2 scale for T2RF patients compared to NEWS with the new SpO2 scale for T2RF patients | Accuracy in predicting immediate need for intensive care, in-hospital death, or discharge with poor neurological status among RRT patients | In patients attended by a RRT, the NEWS predicts patient’s hospital outcome with moderate accuracy. No improvement using the new SpO2 scale in T2RF patients. | Does not support switch from NEWS to NEWS2 modifications. |
| Price | https://pubmed.ncbi.nlm.nih.gov/37931891/ | 2023 | Observational study, retrospective | Hospital inpatients at Portsmouth Hospitals University NHS Trust, excluding specialty wards like burns, spinal injury, neurosurgery, cardiothoracic surgery, maternity patients and individuals under 16. | NEWS2 | NEWS2 performance compared to 36 other Early Warning Scores in 123 patient groups. | Death, ICU admission, or a combined outcome of either death or ICU admission within 24 hours of an observation set. | Consistently high performance indicates that NEWS2 is a suitable early warning score to use for all diagnostic groups considered by this analysis. Patients are not disadvantaged through use of NEWS2 in comparison to any of the other evaluated Early Warning Scores | NEWS2 can be used in all patient groups. |

### Table A3: Maternity (Reviews)

| **First author surname** | **Link** | **Year** | **Study type** | **Population/Setting** | **Intervention/EWS studied** | **Comparator** | **Outcome** | **Main Finding** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Umar | https://pubmed.ncbi.nlm.nih.gov/31150513/ | 2019 | Review, systematic (to 2018), with narrative synthesis | Pregnant women in labour, sick pregnant women of any gestational age, and women who had recently given birth (within 6 weeks of delivery) admitted to hospital units including intensive care and high dependency units, studies from multiple countries including 6 high-income countries. | Various versions of Early Warning Systems (EWS) modified for the obstetric population, typically including pulse rate, respiratory rate, temperature, blood pressure, and consciousness level. | Use of a non-obstetric EWS on an obstetric unit, usual care practice with no use of any EWS | Two thematic categories: studies that investigated the predictive accuracy of EWS for adverse obstetric outcomes (validation studies) and those that investigated the effectiveness of EWS in improving measured outcomes (clinical outcomes and measures of the effectiveness of the EWS trigger mechanism). | 17 studies met the inclusion criteria. Eleven of the included studies evaluated the predictive accuracy of EWS for obstetric morbidity and mortality, 5 studies assessed the effectiveness of EWS in improving clinical outcomes, while one study addressed both. The obstetric EWS identified had very high median (inter-quartile range) sensitivity—89% (72% to 97%) and specificity—85% (67% to 98%) but low median (inter-quartile range) positive predictive values—41% (25% to 74%) for predicting morbidity or ICU admission. Obstetric EWS had a very high accuracy in predicting death (AUC >0.80) among critically ill obstetric patients. Obstetric EWS improves the frequency of routine vital sign observation, reduces the interval between the recording of specifically defined abnormal clinical observations and corrective clinical actions, and can potentially reduce the severity of obstetric morbidity. Obstetric EWS are effective in predicting severe morbidity (in general obstetric population) and mortality (in critically ill obstetric patients). There is limited evidence of the effectiveness of EWS in reducing maternal death across all settings |
| Bernstein | https://pubmed.ncbi.nlm.nih.gov/33549533/ | 2021 | Review, scoping | Women in labour who were the focus of the primary quantitative and qualitative studies evaluating trigger tools for obstetric care escalation. Multiple countries including US, UK, NZ, Denmark, Norway, Sweden. | Five specific scoring tools: MEOWS, MEWT, MEWC, MEWS, MERC, also a non-specific early warning system, and “maternal early warning triggers” that were not part of a named tool. | Performance among the various obstetric EWS | Morbidity, care escalation eg ICU admission | The paper presents a mixed view on the use of early warning scores in obstetrics, highlighting variations in sensitivity and specificity among different trigger tools and the need for further research to improve their accuracy. |
| Quinn | https://pubmed.ncbi.nlm.nih.gov/27031791/ | 2016 | Review non-systematic | Pregnant women admitted to acute care services within 42 days of delivery, including obstetric units, midwifery units, and some community settings, in the UK | ObsEWS, NEWS | N/A | Ability to predict patient deterioration and outcomes in both general and obstetric populations. | The paper provides evidence supporting the use of early warning systems but also criticizes the current state of obstetric early warning systems and calls for improvements and standardization. |
| Loerup | https://pubmed.ncbi.nlm.nih.gov/31506067/ | 2019 | Review, systematic with meta-analysis | Pregnant women, all gestational ages | Reports of normal physiological parameters in pregnancy | Comparison across time in gestation | Median and range of normal values | Significant gestational blood pressure and heart rate changes occur that should be considered when assessing pregnant women. Commonly taught substantial decreases in blood pressure mid-pregnancy were not seen and heart rate increases were lower than previously thought. |

### Table A4: Maternity (Other Studies)

| **First author surname** | **Link** | **Year** | **Study type** | **Population/Setting** | **Intervention/EWS studied** | **Comparator** | **Outcome** | **Main Finding** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Arnolds | https://pubmed.ncbi.nlm.nih.gov/35387624/ | 2022 | Observational retrospective cohort study | Hospital inpatients on antepartum or postpartum wards, University of Chicago, US | MEOWS, MEWC, MEWT, MEWS, NEWS, eCART | Comparison of different scoring systems for predicting outcomes in obstetric patients. | Discrimination and accuracy in predicting outcomes such as ward to ICU transfer, death, and new infection. | Study findings are generally supportive of EWS, particularly eCART for predicting deterioration and infection in ante-and postpartum patient populations. Switching to eCART could improve accuracy compared to other general and obstetric EWS. |
| Blumenthal | https://pubmed.ncbi.nlm.nih.gov/30856674/ | 2019 | Observational retrospective case-control study | Severe maternal morbidity patients (n = 79) compared with 123 controls who did not experience severe maternal morbidity during their delivery hospitalization. Long Beach Memorial Medical Center, California. | MEWS, MEOWS, MERC, and MEWT | Performance among the various obstetric EWS in predicting maternal morbidity. | Assessed whether these systems would have alerted prior to the occurrence of morbidity events. | MEOWS and MERC were more sensitive than MEWS or MEWT (67.1% and 67.1% vs. 19% and 40.5%, p < 0.001); however, MEWT and MEWS were more specific (88.6% MEWT and 93.5% MEWS vs. 51.2% MEOWS and 60.2% MERC, p < 0.001). In the control population, 70% of MEWT alerts still appeared “relevant” to the clinical scenario in contrast to the MEOWS (32%) or MERC systems (31%). None of the systems performs with high sensitivity and specificity. |
| Kern-Goldberger | https://pubmed.ncbi.nlm.nih.gov/35623625/ | 2023 | Observational retrospective cohort study | Pregnant patients admitted for delivery at four campuses of New York-Presbyterian Hospital, including individuals of any gestational age (n = 14,597) | MEOWS, MERC, and MEWT. | Comparison of different Maternal Early Warning Systems (MEWS) criteria and vital sign thresholds to detect maternal morbidity. | Composite of maternal morbidity including haemorrhage, infection, acute cardiovascular disease, and acute respiratory disease | Though all MEWS criteria demonstrated poor sensitivity for maternal morbidity, permutations of the most abnormal vital signs have high specificity, suggesting that MEWS may be better implemented as a trigger tool for morbidity reduction strategies in the highest risk patients, rather than a general screen. |
| Green |  | 2020 | Observational prospective | Pregnant women, all gestational ages, UK | Prospective collection of vital sign recordings in pregnancy | Comparison across time in gestation | Centiles for normal values | Gestation specific reference ranges for detecting abnormal BP, heart rate, respiratory rate, oxygen saturation and temperature during pregnancy were compiled. No clinically significant BP drop from 12 weeks of gestation was observed. |

1. Note that ‘MEWS’ is also used in NZ to stand for ‘maternity early warning system(s)’, however this could cause confusion given the pre-existence of the general Modified Early Warning System, so the former is not used in this report. [↑](#footnote-ref-2)
2. Horswill M, Preece M, Hill A, Watson M. Paper-based patient chart design information sheet. Sydney, New South Wales, Australia: Australian Commission on Safety and Quality in Health Care 2010. Available at: <https://espace.library.uq.edu.au/view/UQ:220755>

   Preece M, Hill A, Horswill M, et al. Developer’s guide for observation and response charts. Report prepared for the Australian Commission on Safety and Quality in Health Care’s program for Recognising and Responding to Clinical Deterioration. Available at: [www.safetyandquality.gov.au/sites/default/files/migrated/ORC-Developers-Guide-4-Oct-2010.pdf](http://www.safetyandquality.gov.au/sites/default/files/migrated/ORC-Developers-Guide-4-Oct-2010.pdf) [↑](#footnote-ref-3)