Department of Veterans Affairs
Health Services Research & Development Service | Evidence-based Synthesis Program

QUERI

Early Warning System Scores: A Systematic Review

January 2014

Prepared for:
Department of Veterans Affairs
Veterans Health Administration
Health Services Research & Development Service
Washington, DC 20420

Prepared by:
Evidence-based Synthesis Program (ESP) Center
Portland VA Medical Center
Portland, OR
Devan Kansagara, MD, MCR, Director

Investigators:
Principal Investigator:
M. E. Beth Smith, DO

Co-Investigators
Joseph C Chiovaro, MD
Maya O'Neil, PhD
Devan Kansagara, MD, MCR
Ana Quinones, PhD
Michele Freeman, MPH
Makalapua Motu'apuaka, BS
Christopher G Slatore, MD, MS

VA HEALTH CARE Defining EXCELLENCE in the 21st Century

PREFACE

Quality Enhancement Research Initiative's (QUERI) Evidence-based Synthesis Program (ESP) was established to provide timely and accurate syntheses of targeted healthcare topics of particular importance to Veterans Affairs (VA) managers and policymakers, as they work to improve the health and healthcare of Veterans. The ESP disseminates these reports throughout the VA.

QUERI provides funding for 4 ESP Centers and each Center has an active VA affiliation. The ESP Centers generate evidence syntheses on important clinical practice topics, and these reports help:

- develop clinical policies informed by evidence,
- guide the implementation of effective services to improve patient outcomes and to support VA clinical practice guidelines and performance measures, and
- set the direction for future research to address gaps in clinical knowledge.

In 2009, the ESP Coordinating Center was created to expand the capacity of QUERI Central Office and the 4 ESP sites by developing and maintaining program processes. In addition, the Center established a Steering Committee comprised of QUERI field-based investigators, VA Patient Care Services, Office of Quality and Performance, and Veterans Integrated Service Networks (VISN) Clinical Management Officers. The Steering Committee provides program oversight, guides strategic planning, coordinates dissemination activities, and develops collaborations with VA leadership to identify new ESP topics of importance to Veterans and the VA healthcare system.

Comments on this evidence report are welcome and can be sent to Nicole Floyd, ESP Coordinating Center Program Manager, at nicole.floyd@va.gov.

Recommended citation: Smith MEB, Chiovaro J, O'Neil M, Kansagara D, Quinones A, Freeman M, Motu'apuaka M, Slatore CG. Early Warning Scoring Systems: A Systematic Review. VA-ESP Project #05-225; 2013

This report is based on research conducted by the Evidence-based Synthesis Program (ESP) Center located at the Portland VA Medical Center, Portland OR funded by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Quality Enhancement Research Initiative. The findings and conclusions in this document are those of the author(s) who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the Department of Veterans Affairs or the United States government. Therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs. No investigators have any affiliations or financial involvement (e.g., employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in the report.

EXECUTIVE SUMMARY

Background: Recognizing early signs of clinical deterioration of hospitalized patients is thought to improve patient outcomes by activating more attentive care in a timely fashion. Early warning system scoress are tools used by care teams to potentially predict a patient's risk of deterioration and facilitate changes in management.

Objective: To systematically review the evidence on the predictive ability of Early Warning System (EWS) scores and the impact of EWS interventions on health and resource related outcomes.

Data Sources: MEDLINE, CINAHL, and Cochrane Library through April 2013 and hand search of reference lists.

Study Selection: Independent dual review to identify English language studies of early warning systems tested with adult patients admitted to medical or surgical wards. Non-systematic reviews, opinions, and case series were excluded.

Data Extraction: Data were extracted by 2 reviewers on the population, setting, sample size, duration, model discrimination and calibration, health outcomes, and resource utilization. Quality was assessed as applicable using a modified version of the Quality in Prognosis Studies (QUIPS) assessment tool for observational studies. For predictive ability, primary outcomes were model discrimination on 48-hour mortality, cardiac arrest, or pulmonary arrest. Outcomes for impact of EWS implementation included 30-day mortality, cardiovascular events, use of vasopressors, number of ventilator days, respiratory failure, and resource utilization.

Results: Of 13,595 citations reviewed, 17 studies of 11 unique models met criteria. All were based on some combination of vital signs and clinical evaluation. Six observational studies tested in large urban hospitals in developed countries found a strong predictive value for death (AUROC 0.88-0.93) and cardiac arrest (AUROC 0.77-0.86) within 48 hours. Eleven observational cohort studies with historical controls provided evidence on the impact of EWS implementation but were insufficient to draw firm conclusions due to methodological limiitations.

Conclusions and Relevance: Current early warning system scores perform well for predicting cardiac arrest and death within 48 hours although the impact on health outcomes and utilization of resources remains uncertain. Efforts to more rigorously assess their performance and effectiveness are needed as use becomes more widespread.

TABLE OF CONTENTS

BACKGROUND

Early warning system (EWS) scores are tools used by hospital care teams to recognize the early signs of clinical deterioration in order to initiate early intervention and management, such as increasing nursing attention, informing the provider, or activating a rapid response or medical emergency team.[1] These tools involve assigning a numeric value to several physiologic parameters (e.g., systolic blood pressure, heart rate, oxygen saturation, respiratory rate, level of consciousness, and urine output) to derive a composite score that is used to identify a patient at risk of deterioration. Most are based on an aggregate weighted system in which the elements are assigned different points for the degree of physiological abnormality. Observational studies suggest that patients often show signs of clinical deterioration up to 24 hours prior to a serious clinical event requiring an intensive intervention.[2] Delays in treatment or inadequate care of patients on general hospital wards may result in increased admissions to the intensive care unit (ICU), increased length of hospital stay, cardiac arrest, or death.[2]

The purpose of the EWS scores is to ensure timely and appropriate management of deteriorating patients on general hospital wards. This is potentially a significant topic for the VA, as the Portland, Oregon VA Medical Center has implemented a Modified Early Warning System (MEWS) and there are plans to implement this nationally. This evidence review will be used by the Office of Nursing Services Clinical Practice Programs ICU Workgroup to develop guidelines for the development and implementation of EWS scores at facilities within the VA system and will be used to identify gaps in evidence that warrant further research.

METHODS

TOPIC DEVELOPMENT

We followed a standard protocol for systematic reviews and developed an analytic framework (Figure 1) with input from key informants (clinicians, nurses, hospital administration and patient advocates) to answer the following key questions:

Key Question 1:

In adult patients admitted to the general medicine or surgical wards, what is the predictive value of EWS scores for patient health outcomes within 48 hours of data collection, including short-term mortality (all cause or disease specific), cardiac arrest and pulmonary arrest? Which factors contribute to the predictive ability of EWS scores and does predictive ability vary with specific subgroups of patients?

Key Question 2A:

What is the impact of using Early Warning Systems on patient health outcomes including 30-day mortality, cardiovascular events (cardiac arrest, acute coronary syndrome and cardiogenic shock), use of vasopressors, number of ventilator days, respiratory failure and length of hospital stay?

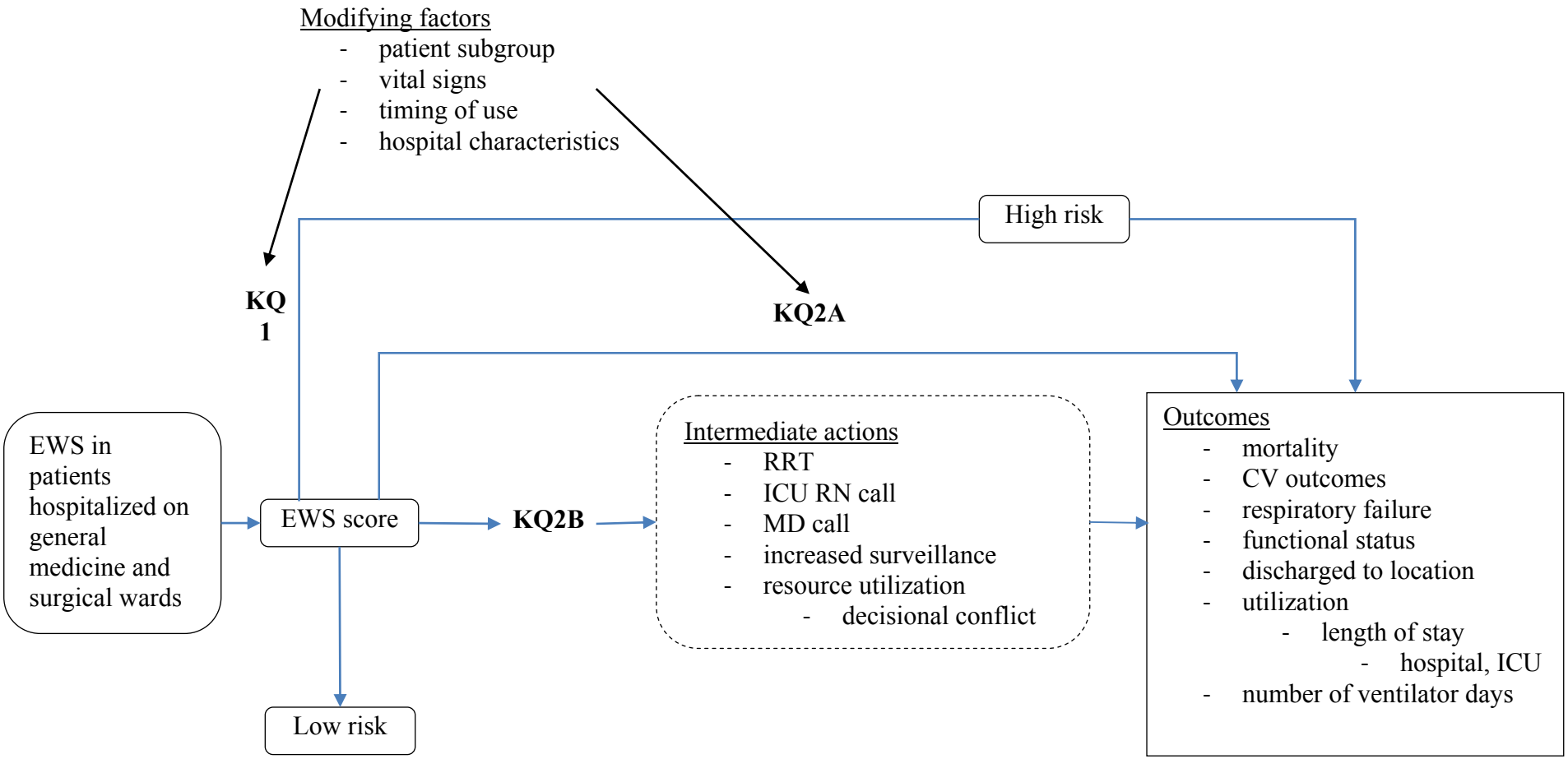
Key Question 2B:

What is the impact of EWS on resource utilization including but not limited to admissions to the intensive care unit (ICU), length of hospital stay, and use of Rapid Response Teams (RRT)?

We met regularly throughout the review with members of a technical expert panel, some of whom served as key informants during the development phase, to oversee the clinical applicability, content completeness, and methodological rigor of the review process.

The population comprises adults admitted to the general medicine or surgical wards. Interventions include any Early Warning System scoring or other established scoring system designed to identify deteriorating patients on hospital wards, including but not limited to Modified Early Warnings Systems (MEWS), Patient at Risk (PAR) score, Physiological Scoring Systems (PSS), Vital Sign Score (VSS), Manchester Triage System, BioSign, VitalPAC Early Warning Score (ViEWS) and Physiological Observation Track and Trigger System (POTTS).

Figure 1. Analytic Framework



SEARCH STRATEGY

We searched MEDLINE, CINAHL, and the Cochrane Central Register of Controlled Trials for literature published from database inception to April 2013. We searched for English-language studies of EWS in medical and surgical ward populations. Appendix A provides the search strategy in detail. We obtained additional articles from systematic reviews, reference lists of pertinent studies, reviews, editorials, and by consulting experts. All citations were imported into an electronic database (EndNote X4).

STUDY SELECTION

We included English language articles that provided primary data relevant to either key question. For studies on predictive value, we included hospital settings including the emergency department (ED). For studies on the impact of EWS intervention, we excluded studies conducted in settings other than the general medicine or surgical wards such as the emergency department (ED), and studies that did not include outcomes (opinion articles, descriptive/discussion articles, and consensus statements) or that examined outcomes outside of the proposed scope of work (e.g., impact on time interval of activating Rapid Response Team). A team of investigators individually reviewed citations and abstracts identified from electronic literature searches; if at least one reviewer indicated that a citation might be relevant, a second reviewer screened the citation for concordance. Full-text articles of potentially relevant references were retrieved for further review. Each article was independently assessed by 2 reviewers using the eligibility criteria shown in Appendix B. Disagreement was settled by consensus of the 2 investigators or by group discussion when agreement could not be reached. To assess the predictive ability of EWS, we examined observational studies reporting associations between EWS scores and 48-hour mortality, cardiac arrest, or pulmonary arrest. We chose outcomes with short time frames because the ability of models to predict longer-term clinical deterioration is not directly relevant to their utility in guiding acute care decisions.

To assess the impact of EWS on health outcomes, we included studies examining the effects of EWS interventions on 30-day mortality, cardiovascular events (cardiac arrest, acute coronary syndrome, and cardiogenic shock), use of vasopressors, number of ventilator days, and respiratory failure. We also included studies reporting the effects of EWS on the utilization of resources including admissions to the intensive care unit (ICU), use of rapid response teams (RRT), and length of hospital stay. We used a best evidence approach to guide inclusion of studies.[3] After an initial search failed to identify any controlled clinical trials, we opted to include any intervention study regardless of design except for case series.

DATA ABSTRACTION AND QUALITY ASSESSMENT

From each study, details were abstracted by one investigator and checked for accuracy by a second investigator including population characteristics, setting, number of subjects, number lost to full analysis, name and elements of scoring system, comparator, harms, implementation characteristics (e.g., staff training, pilot phase), and funding. For studies addressing the question of predictive ability, we abstracted data on model discrimination for outcomes of mortality,

cardiac arrest, and pulmonary arrest within 48 hours. For all other studies we abstracted data on patient health outcomes and resource utilization.

We report the c-statistic, with 95% confidence interval when available, to describe model discrimination. The c-statistic, which is equivalent to the area under the receiver operating characteristic curve (AUROC), is the proportion of times the model correctly discriminates a pair of high- and low-risk individuals.[4] A c-statistic of 0.5 indicates the model performs no better than chance; a c-statistic of 0.7 to 0.8 indicates modest or acceptable discriminative ability, and a threshold of greater than 0.8 indicates good discriminative ability.[5,6] If the c-statistic was not reported, we abstracted other operational statistics such as sensitivity, specificity, and predictive values for representative risk score cut-offs when available.

For key question one, we adapted criteria described in the Quality in Prognosis Studies (QUIPS) assessment tool.[7] We provide a discussion of the study strengths and limitations and describe the overall potential for bias as applicable. For key question 2, we did not find any tools designed to evaluate historically controlled study designs so provided a discussion of strengths and limitations.

DATA SYNTHESIS

Data were synthesized qualitatively focusing on model discrimination, the populations in which the model has been tested, impact of model implementation on health outcomes and resource utilization, practical aspects of model implementation, and the types of variables included in each model. The included studies did not allow meta-analysis due to heterogeneity and qualitative nature of the data.

ROLE OF THE FUNDING SOURCE

This research was funded by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Quality Enhancement Research Initiative. The draft report was reviewed by content experts and collaborative partners. Investigators worked with VA ESP staff to develop and refine the scope, analytic framework and key questions; resolve issues arising during the project; and review the final report to ensure methodological standards for systematic reviews were met. The investigators are solely responsible for the content and the decision to submit the manuscript for publication.

RESULTS

LITERATURE FLOW

From 13,595 titles and abstracts, 129 articles were selected for full-text review (Figure 2). Of these, we included 17, 6 providing primary data on predictive value of EWS scores and 11 pertaining to the impact of EWS interventions. We found 11 unique models ranging from 4 to 12 items with scores based on aggregate weighted systems (Table 1).

Figure 2. Literature flow diagram

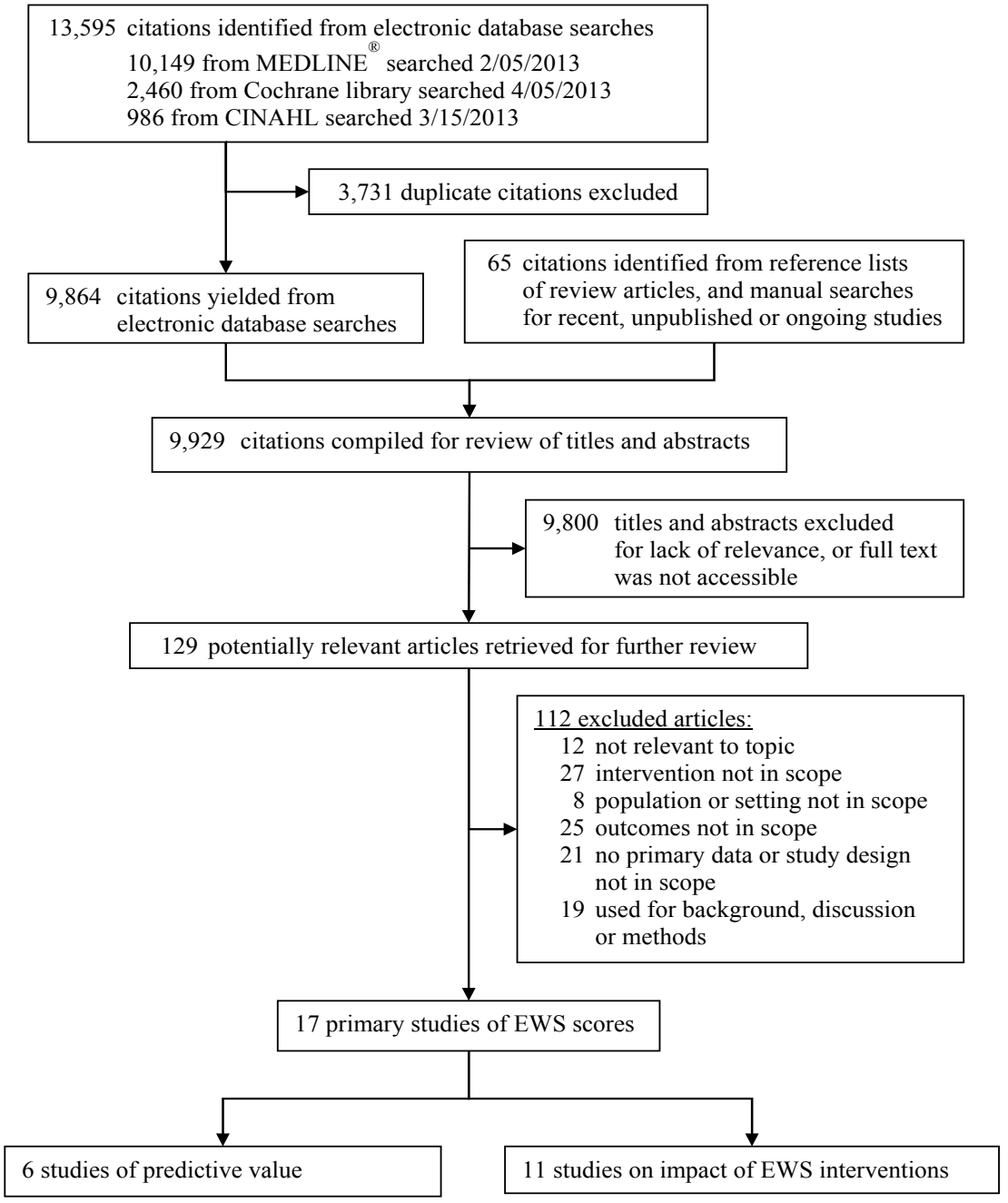


Table 1. Parameters used in studies of the predictive ability and effectiveness of early warning system scores for clinical deterioration in medical and surgical inpatients

Study Country	N parameters; name of scoring system	Parameters used in the system scores										
		Heart rate	Resp rate	SBP	Temp	Urine output	O2 Sat	Difficulty breathing	Supp O2	Mental Status (LOC)	Concern	Other, specify
Rothschild, 2010[8] USA	Single items, not combined	X	X	X	X	X	X	---	X	X	X	DBP, seizures, uncontrolled bleeding, color change
Churpek, 2012[9] USA	4-item CART	X	X	---	---	---	---	---	---	---	---	DBP, Age
Maupin, 2009[10] USA	5-item MEWS	X	X	X	X	---	---	---	---	X	---	---
Jones, 2011[11] UK	Patientrack EWS	X	X	X	X	---	---	---	---	X	---	---
Subbe, 2003[12] UK	5-item MEWS	X	X	X	X	---	---	---	---	X	---	---
Churpek, 2012[13] USA	5-item MEWS	X	X	X	X	---	---	---	---	X	---	---
O'Dell, 2002[14] UK	5-item MEWS	X	X	X	---	X	---	---	---	X	---	---
DeMeester, 2012[15] Belgium	6-item MEWS	X	X	X	X	---	X	---	---	X	---	---
Smith, 2006[16] UK	6-item EWS	X	X	X	X	X	---	---	---	X	---	---
Patel, 2011[17] UK	6-item MEWS	X	X	X	X	X	---	---	---	X	---	Catheterized
Kellett, 2012[18] Canada	6-item ViEWS	X	X	X	X	---	X	---	X	---	---	---
Mitchell, 2010[19]	7-item MEWS	X	X	X	X	X	X	---	---	X	---	---
Moon, 2011[20] UK	7-item MEWS	X	X	X	X	X	X	---	---	X	---	---
Green, 2006[21] Australia	7-item clinical marker tool	X	X	X	---	X	X	X	---	---	X	---
Smith, 2013[22] UK	7-item NEWS	X	X	X	X	---	X	---	X	X	---	---
Prytherch, 2010[23] UK	7-item ViEWS	X	X	X	X	---	X	---	X	X	---	---
Albert, 2011[24] USA	12-item MEWS	X	X	X	X	X	X	X	X	X	X	WBC; new focal weakness

PREDICTIVE VALUE OF EWS

Six observational studies (4 prospective cohort, 2 case-control) met our inclusion criteria (Table 2).[8,9,13,18,22,23] They reported the predictive values of 4 distinct models of early warning system scores for the outcomes of interest, death and cardiac arrest within 48 hours of measurement. No study reported on the predictive ability of EWS for respiratory arrest. The studies were conducted in several countries (3 USA, 2 UK, 1 Canada) though most were in academic, urban hospitals. Although one study considered single predictors, the 4 models ranged from 4 to 7 items, all of which included heart rate, respiratory rate, and blood pressure, and most included temperature and mental status. All models were based on an aggregated weighting system. In general, the early warning system scores appeared to have strong predictive ability as judged by the area under the receiver operator characteristic curve (AUROC). Patients with favorable EWS scores were unlikely to suffer cardiac arrest or death within 48 hours. Relatively few patients had less favorable scores and while higher scores corresponded to higher rates of adverse outcomes, the sensitivity was poor as a large majority still did not.

Though the included studies provide important information related to predictive ability and the studies were generally well-conducted for their study design, the body of evidence is limited by some of risk of bias (Table 3). The case-control design of 2 of the studies[8,13] have the potential for bias related to differential exposure assessment as it is unclear whether vital sign measurement was different for cases compared to controls. Additionally, although 2 studies conducted validation of models,[13,18] 4 were derivation studies.[8,9,22,23] Studies in which statistical models are derived but not validated in separate populations of patients are at risk for "over-fitting" data to the population under study. Such studies may exaggerate the predictive ability of models and, furthermore, may not be broadly applicable to populations of interest.

Kellett and colleagues provide the strongest evidence on the predictive ability of EWS scores.[18] They conducted a prospective cohort study of a consecutive sample of adult patients (n=75,419) using an abbreviated version of a previously derived EWS, ViEWS (temperature, systolic blood pressure, oxygen saturation, use of supplemental oxygen, heart rate, respiratory rate). Vital signs were recorded using Vitalpac software. Among medical patients, the abbreviated ViEWS score had an area under the receiver operator characteristic curve (AUROC) of 0.89 (95% CI, 0.85-0.92) as a predictor of death within 48 hours of the observed score (from admission). This group evaluated the AUROC among multiple subsets of patients in terms of sex, year of admission, age, indication for admission (medical vs surgical), and specific diagnoses. All of the AUROC values among patients not admitted to the intensive care unit were above 0.85. Low abbreviated ViEWS scores were associated with a very good prognosis as only 11 of 49,077 total patients with scores <3 (of 21 possible points) died within 48 hours of admission. Although 14% of the 519 patients with scores >11 died within 48 hours, the false positive rate would be 86% using this cut-off.

Two other studies reported on the AUROC values utilizing the same cohort of patients. Pyrthech and colleagues conducted a prospective cohort study to derive the EWS ViEWS (plus mental status) among patients on a general medical service with vital sign data recorded using Vitalpac software (n=35,585).[23] The ViEWS score had an AUROC of 0.888 (95% CI, 0.880-0.895) as a predictor of death with 24 hours of the observed score. The sensitivity of the ViEWS was

approximately 67% at a specificity of 90%. They also evaluated an "efficiency curve" for the ViEWS and death within 24 hours. As an example, approximately 20% of observations had a score of 5 or more (of 21 possible points) and approximately 83% of deaths within 24 hours had a score of 5 or more. Higher scores were associated with a higher risk of death. However, using a higher cut-off lowered the sensitivity and more patients classified as low risk died.

In a related study, Smith and colleagues used the same vital sign data to evaluate the NEWS: heart rate; diastolic and systolic blood pressure; respiratory rate; oxygen saturation; temperature; and mental status using the Alert-Verbal-Painful-Unresponsive (AVPU) scale or the Glasgow Coma Score (GCS).[22] The NEWS score had an AUROC of 0.894 (95% CI, 0.887-0.902) and 0.857 (95% CI, 0.847-0.868) as a predictor of death and cardiac arrest, respectively, within 24 hours of the observed score. This study did not report sensitivity at different specificities or show the ROC curve for death and cardiac arrest to estimate these data. However, as an example of the risk for a false positive, approximately 97% of patients with a NEWS value of 8 (of 20 possible points) did not die with 24 hours.

Churpek and colleagues conducted a retrospective cohort study of a novel EWS, the Cardiac Arrest Risk Triage (CART), among patients admitted to medical and surgical units, n=47,427.[9] The CART model included respiratory rate, heart rate, diastolic blood pressure, and age. They compared the CART model to the MEWS (includes respiratory rate, pulse, systolic blood pressure, temperature, and the AVPU mental status scale) and found that the CART model was superior in predicting cardiac arrest (AUROC 0.84 vs 0.78, p = 0.001). At a specificity of 90%, the CART score (cut-off 17 of 57 possible points) had a sensitivity of 53% compared to 48% for the MEWS (cut-off >4 of possible 14) for cardiac arrest.

Churpek and colleagues found similar results in a nested case-control study (88 cases of cardiac arrest, 352 controls) of MEWS using data from the same cohort among medical and surgical patients.[13] The MEWS score had an AUROC of 0.77 (95% CI, 0.71-0.82) as a predictor for cardiac arrest. Using a cut-off MEWS score of >3 (of possible 14), the specificity for cardiac arrest was 87% with a sensitivity of 51%.

Rothschild conducted an observational, case-control study (262 patients with 271 incidents of cardiac arrest and ICU transfer versus 318 controls) of at least one early warning criterion within 8 hours of a life-threatening event among medical and medicine subspecialty patients.[8] Vital signs were recorded during the time the patient was on the general ward. Criteria most associated with a life-threatening event included respiratory rate >35 (OR 31.1, 95% CI, 7.5-129.6), need for supplemental oxygen to 100% or use of a non-rebreathing mask (OR 13.7, 95% CI 5.4-35), and heart rate >140/minute (OR 8, 95%CI 2.4-27.5). Multiple positive criteria were more common in cases than controls (≥3 positive criteria, 23 vs 16, p=0.00027). However, among the 26 patients with a cardiac arrest, 20 (77%) did not have a preceding positive early warning sign within 8 hours of the event.

In general, the early warning system scores appeared to have strong predictive ability for death and cardiac arrest within 48 hours however most patients with high scores did not suffer an event and low scores did not preclude one. The evidence is insufficient to determine if one system is superior to another or to determine which factors contribute most to models' predictive ability.

Table 2. Observational studies of the predictive value of EWS scores

Author, Year; Model; Setting	Population; Age; Gender	Comparison	N	Predictive measures of mortality occurring within 72 hours of EWSS data collection/analysis	Predictive measures of cardiac arrest occurring within 72 hours of EWSS data collection/analysis
Churpek, 2012[9]; predictive model, retrospective cohort; US academic tertiary care hospital 500 beds	Medical/surgical pts, including telemetry pts; Mean age: Cardiac arrest pts 64 ICU Txf pts 60 Control pts 54; Male: Cardiac arrest pts 43% ICU Txf 52% Control pts 43%	Pts without an event (cardiac arrest or ICU transfer)	See previous Cardiac arrest: 88 pts ICU Txf: 2820 pts Control: 44,519 pts	NR At a specificity of 90%, the CART score (cut-off 17 of 57 possible points) sensitivity was 53%, compared to 48% for the MEWS (cut-off >4 of possible 14) for cardiac arrest.	AUROC: 0.84 (vs 0.78 for MEWS), 95% CI NR
Churpek, 2012[13]; case-control; US academic tertiary care hospital	73% medical, 27% surgical;l Mean age: Cases 64 Controls 58 (p=0.002); Male: Cases 43% Controls 49% (P=ns)	Control pts without cardiac arrest	88 cases, 352 controls	NR	Maximum MEWS AUC was the highest compared to other individual predictors at 0.77 (95% CI 0.71-0.82) for CA outcome. Other statistically significant AUCs were for max respiratory rate, heart rate, pulse pressure index, and minimum diastolic BP; no others were significant. Using a cut-off MEWS score of >3(of possible 14), the specificity for cardiac arrest was 87% with a sensitivity of 51%.
Kellett, 2012[18]; predictive model; Canada urban academic 375 beds	Med/Surg inpts; Mean age 63; 48.9% male	None	75,419 total Medical, non-ICU: 43693 Surg: 30485 (presumably includes some ICU pts)	48hr mortality AUROC for all pts: 0.93 (95% CI 0.91-0.95) AUROC for Surg: 0.89 (95% CI 0.78-1.0) AUROC for Medical, non-ICU: 0.89 (95% CI 0.85-0.92) 14% of the 519 pts with scores >11 died within 48 hours (86% false positives).	

Author, Year; Model; Setting	Population; Age; Gender	Comparison	N	Predictive measures of mortality occurring within 72 hours of EWSS data collection/analysis	Predictive measures of cardiac arrest occurring within 72 hours of EWSS data collection/analysis
Prytherch, 2010[23]; predictive model; prospective UK Hospital	General Medical Emergency patients consecutively admitted; Mean age 67.7 Median age 72.6; 47.5% male	Pts who were alive at 24 hours following observation	35,585 pt episodes	AUROC = 0.888 (95%CI 0.885-0.895) for death within 24 hrs Sensitivity was approximately 67% at a specificity of 90%.	NR
Rothschild, 2010[8] case-control, predictive model; US urban academic 745 beds	Medical and medicine subspeciality inpts; Mean age: 61; 49.6% male	Control patients matched on day of admit	262 pts 318 controls	NR	Among 26 patients with a cardiac arrest, 20 (76.9%, 95% CI 60.7-93.1%) did not have a preceding positive early warning sign.
Smith, 2013[22] UK predictive model; prospective Hospital	General Medical patients; Mean age 67.7 Median age 72.6; 94,376/198,755 observations from males	Pts without an event	See previous	Outcome: death within 24hrs AUROC: 0.894 (95% CI 0.887-0.902) 97% of patients with a NEWS value of 8 (of 20 possible points) did not die with 24 hours.	Cardiac arrest within 24hrs AUROC: 0.857 (95% CI 0.847-0.868)

Table 3. Assessment of study methods for potential sources of bias in studies of the predictive value of EWS scores

Study	**Study Population:** study sample adequately represents population of interest	**Study Attrition:** study data available adequately represent study sample	**Prognostic Factor Measurement:** measured in a similar way for all participants	**Outcomes Measurement:** outcome of interest measured in similar way for all participants	**Study Confounding:** Important potential confounding factors are appropriately accounted for	**Statistical Analysis:** statistical analysis is appropriate	**Reporting:** All primary outcomes reported
Churpek, 2012[13]	Yes; "consecutive"	Yes	Yes	Yes	Unclear: Cases might be significantly different from controls; not enough information provided	Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all (all AUCs include a 95% CI)	Yes
Kellett, 2012[18]	Yes; "every patient"	Yes	Yes	Yes	Yes	Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all (all AUCs include a 95% CI)	Yes
Prytherch, 2010[23]	Yes; "consecutive"	Yes	Yes	Yes	Yes	Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all (all AUCs include a 95% CI)	Yes
Rothschild, 2010[8]	Unclear: No statement of "all" etc.	Yes	Yes	Yes	Unclear: Cases might be significantly different from controls; not enough information provided	Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all	Yes
Smith, 2013[22]	Yes; "consecutive"	Yes	Yes	Yes	Yes	Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all (all AUCs include a 95% CI) though all p values < .001	Yes

IMPACT OF EWS INTERVENTIONS

We found 11 observational cohort studies with historical controls of adult patients admitted to general medical or surgical wards (n=89 to over 200,000 patients) that met our inclusion criteria for impact of EWS implementation on outcomes of 30-day mortality, cardiac arrest, and utilization of resources (Table 4).[10-12,14-17,19-21,24] We found no studies reporting on the other outcomes of interest. The early warning systems ranged from 5-item to 12-item models. All included heart rate, respiratory rate, and systolic blood pressure; all but one included level of consciousness or mental status; and most included temperature and urinary output. The studies were conducted in several countries (5 UK, 2 USA, 1 Australia, 1 Belgium) and most were in academic, urban hospitals.

This represents an insufficient body of evidence due to methodological limitations. Given that all of the studies used historical controls, they are at risk of unknown, unmeasured confounding factors affecting outcomes and are subject to the effects of time. None adjusted for pre-intervention trends in mortality rate, accounted for other secular changes in care that could simultaneously have impacted mortality, and none reported or compared the rate of change for key outcomes pre-intervention (i.e., the slope of the outcome) to the rate of change (slope) of the outcome for a period of time following implementation.

IMPACT OF EWS ON HEALTH OUTCOMES

Mortality

We found 6 studies that directly addressed the effects of Early Warning System (EWS) implementation on mortality. Four studies found a decrease in overall mortality after implementation of an EWS,[11,12,17,20] but only one study found this to be statistically significant.[20] Moon et al (2011) implemented an EWS at 2 hospitals and at both institutions deaths per hospital admission decreased significantly from 1.4% to 1.2% ($p<0.0001$). In addition, both the mortality per cardiac arrest call and the in-hospital mortality of patients admitted to the intensive care having undergone CPR fell significantly (26% to 21%, $p<0.0001$ and 70% to 40%, $p<0.0001$, respectively). Notably, the hospitals underwent an expansion of critical care outreach services during this period so the independent impact of EWS implementation is unknown.

Mortality rates after EWS implementation were lower in 3 other studies, but none reached statistical significance.[11,12,17] One additional study by DeMeester and colleagues used a combined end point of death in patients without a "Do Not Resuscitate" order or readmission to the ICU and demonstrated a non-significant decrease in this outcome.[15]

Conversely, in a study by Green and Williams, there was a non-significant increase in overall mortality after implementation of an EWS. However, for patients who were spontaneously breathing with a pulse at the time of "code blue" call, there was a significant improvement in survival (59% to 75%, $p=0.0003$).[21]

Cardiac arrest

Three studies evaluating the effects of EWS implementation on cardiac arrests met inclusion criteria and found mixed results.[11,12,20] No studies were found that specifically addressed

pulmonary arrests, though most studies did not differentiate between cardiac and pulmonary arrests.

Moon et al found that the proportion of cardiac arrest calls per adult admission decreased in both hospitals involved in the study (0.4% to 0.2%, p<0.0001 and 0.34% to 0.28%, p<0.0001).[20] Subbe et al stratified the incidence of cardiac arrest by EWS score on admission using the MEWS tool. Low (scores 0-2) and high (scores 5-15) risk groups did not have a significant difference in cardiac arrests after EWS implementation. However, the moderate risk group (scores 3-4) demonstrated a significant increase in cardiac arrests in the study population when compared to the control population (5% vs 0%, p<0.016).[12] No significant difference was found in incidence of cardiac arrest after EWS implementation in the study by Jones et al.[11]

Two studies looked at the impact of EWS implementation on "code blue" events. In Maupin et al, the authors reported a decrease in the number of "code blues" outside the ICU from 0.77 per 1000 patient days to 0.39 per 1000 patient days, but the statistical significance was not reported.[10] In the study by Green and Williams, the overall trend in cardiac arrest in relation to number of patients was not reported. However, among "code blue" calls, there was a significant decrease in the percentage of patients who had suffered cardiac arrest (52.1% to 35%, p=0.0024).[21]
In summary, studies on the impact of EWS scores on mortality and cardiac arrest provided limited information. In general, the data suggests a trend toward decreased mortality. However, only one study reported statistically significant improvements in overall mortality and none of the studies controlled for trends in mortality independent of the intervention. The data on cardiac arrest was even more limited and the results were mixed.

IMPACT OF EWS ON RESOURCES

Length of hospital stay

Three studies evaluated the impact on length of hospital stay before and after implementing an early warning system and found mixed results.[11,19,21] One study included ward patients with abnormal vital signs referred to the ICU liaison team or patients with an unplanned ICU admission or medical emergency comparing data from 12 months prior to EWS implementation to 24 months after implementation.[21] They found no difference in the total length of hospital stay: 19 (9-39) days vs 18 (8-33.7) days.[21] The second study was similar but included all patients admitted to the ward and had a shorter pre-post design (47 days pre and 38 days post). They found a significant decrease in length of stay: 9.7 vs 6.9 days, p<0.001.[11] The difference in study population (patients with an unplanned ICU admission or medical emergency vs all ward patients) likely accounted for the difference in outcomes. The third study compared 4 months before and 4 months after implementation and showed an increase in length of hospital stay from 4.0 (1.8-8.3) days to 4.8 (2.2-9.8) days.

Admissions to the intensive care unit

Four studies evaluated the impact of the EWS on the number of admissions to the intensive care unit (ICU).[11,12,20,21] Two studies found a significant increase in the number of ICU admissions after implementing EWS, accounting for differences in overall hospital admission rates.[11,21] One study involving 2 institutions however found a decrease in the number of unplanned admissions to the ICU after implementing EWS (3% vs 2% and 6.65% vs 2.63%).[20] The length of time an

unstable patient was on the ward before ICU admission was found to be less after implementing EWS (percentage of patients with clinical markers of instability for ≥6 hours, 41.2% vs 24.5%)[21] and 2 studies found no difference in the length of the ICU stay.[12,21]

Use of rapid response and code teams

Four studies evaluated the impact of EWS on rapid response and code teams.[10,20,21,24] All of the studies found at least a 50% increase in the number of rapid response or ICU liaison team calls. Three of the studies found a 6-33% decrease in the number of code blue calls.[20,21,24] Of interest, one study found that the number of code blue calls for a patient still breathing and with a pulse increased from 47.9% to 64.4%.[21]

Nursing

The impact on nursing was not well studied. Three studies did measure the accuracy and compliance of scoring with compliance being as low as 53% in one study[15] and accuracy being as high as 81-100% with the use of electronic calculations.[11] One study found that the most inconsistently recorded elements were urinary output and level of consciousness (45.6% missed values) and that respiratory rate had the highest errors (9.6%).[16] The number of observations and clinical attention by nursing increased with the use of the EWS with greater attention for EWS scores >5.[11] One study found that the frequency of patient observations per nursing shift increased during the daytime but not at night.[15]

In summary, the studies evaluating the impact of EWS on the use of resources are limited by their study designs and only provide a suggestion of the true effect. The results suggest that the use of staffing including nursing care and rapid response teams will increase; however, the effect on the length of hospital or ICU stay remains uncertain.

Table 4. Studies of the impact of EWS interventions on patient outcomes and resource utilization

Author Year; Setting	Study design	Observation period	Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)	Score that triggers activation of response; Description of Response	In-hospital outcomes: • Mortality (30-day) • Cardiovascular events	Resource utilization: • length of hospital stay; • ICU	Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing
Albert, 2011[24]; US 550-bed tertiary care academic hospital	Retrospective cohort; comparing MEWS score 4-8hr prior to event and at time of emergent event	6mo pilot study	N=140; age: RRT 64.7 v code 65.8; telemetry ward	≥3 (total 15); 1 or 2: nurse alert to reassess patient in ≤4hr; 3: discuss with charge nurse and decision made on RRT activation based on stability	NR	NR	N=78 (55.7%) RRT calls; mean MEWS for RRT 6.35; 6mo after implementation found a 33% reduction in code team calls and 50% increase in RRT calls
De Meester, 2012[15]; Belgium tertiary care hospital, 14 medical and surgical wards	Pre-retrospective and post-prospective	5dy period after ICU discharge	N=1039; age: 59; % male: 60; MEWS at ICU discharge: 2 5-days post ICU-discharge to medical and surgical wards	Score increase by 2 points; one vital sign scored 3 points; 4: nurse felt patient unsafe (total NR); observation frequency increased to every 30min and physician notified	Patients who died without a DNR or re-admission to ICU, %: pre 5.7 vs post 3.5 ARR of 2.2% (95% CI -0.4% to -4.67%)	NR	53% compliance for completion of MEW total population: patient observation frequency per nursing shift increased from 0.9993 (95% CI .9637-1.0035) to 1.07 (95% CI 1.10362-1.1101); p=.005; no difference during night but significant daytime increase from 1.1404(95% CI 1.1067-1.1742) to 1.2262 (95% CI 1.1899-1.2625)

Author Year; Setting	Study design	Observation period	Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)	Score that triggers activation of response; Description of Response	In-hospital outcomes: • Mortality (30-day) • Cardiovascular events	Resource utilization: • length of hospital stay; • ICU	Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing
Green, 2006[21]; Australia 323 bed tertiary care academic hospital surgical and medical ward	Pre-retrospective and post-prospective	Pre: February 2002-January 2003; post: February 2003-January 2005	N=415; Demographics NR; all ward patients with abnormal VS referred to the ICU liaison team and any patient who had an unplanned admission to the ICU from the wards or who suffered a cardiac arrest/medical emergency	NR; activate ICU Liaison Team	Mortality, %: pre 33.9 vs post 34.5	Hospital: pre 19dy vs post 16dy; ICU: pre n=153 including 111 admissions and 42 readmissions; LOS: 3.0 (1.3-6.9); post n=412 including 320 admissions and 92 readmits; LOS: median 2.6 (1.2-6.4)	Number of code blue calls, %: pre 52.1 to 35; number of code blue calls for patients still breathing and had a pulse, %: pre 47.9 to 64.4; ICU liaison Team visits: pre n=630 for 1958 visits, post n=1889 for 4586 visits; length of time on ward before admission to ICU: pre markers present for ≤6h n=90(58.8%), post n=311 (75.4%); markers present for ≥6h: pre n=63(41.2%), post 101 (24.5%), p<0.0002

Author Year; Setting	Study design	Observation period	Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)	Score that triggers activation of response; Description of Response	In-hospital outcomes: • Mortality (30-day) • Cardiovascular events	Resource utilization: • length of hospital stay; • ICU	Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing
Jones, 2011[11]; UK academic hospital medical assessment unit	Pre-post with 3 phases: baseline, PDA with downloading to ward-view, Patient Track; hospital and clinical outcomes and LOS collected retrospectively	Pre: 47dy November-December, 2007; alert 38dy August-September, 2008	Baseline n=705, (7820 observations); alert phase n=776 (5848 observations); median age: 70 vs 65, p=0.01; % male: 52 vs 53; all patients admitted to ward excluding those discharged within 24hr	3 (total 14); EWS 3-5: inform charge nurse and nurse intervention, recheck in 1hr, if still ≥3, call junior DR, if still ≥3 in 1hr, call senior DR and recheck 1hr, if still ≥call critical care medical team	Mortality, %, pre 9.5 vs post 7.6; cardiac arrest, %: pre 0.4 vs 0	Hospital: 9.7dy vs 6.9dy, p<0.001 ICU: pre 51dy vs 26dy, p=0.04	Clinical attendance with EWS 3-5, %: pre 29 vs post 78, p<.001; clinical attendance with EWS >5, %: pre 67 vs post 96.2; EW accuracy with electronic calculation 81%-100%; clinical attendance to patients with EWS 3, 4,or 5 increased from 29% to 78% with automated alerts, p<0.001; for EWS >5 clinical attendance increased from 67% to 95%, p<0.001; EW accuracy with electronic calculation 81%-100%; clinical attendance to patients with EWS 3,4,or 5 increased from 29% to 78% with automated alerts, p<0.001; for EWS >5 clinical attendance increased from 67% to 95%, p<0.001
Maupin, 2009[10]; US 200 bed general hospital	Pre-post; RRT calls per 1000 patient days increased from 7.8 to 16.4 after implementation (Figure 3)	Pre: 1yr (2007) pilot for 2mo post: 1yr	N=NR; general medical, surgical, tele floors	3: increased VS frequency; 4: notify physician; 5: call RRT (total 14); EWS 3: increase VS frequency to q2 hours x 3, calculate MEWS each time. Inform charge nurse	NR, the cardiac arrest was listed as either outside ICU or not	NR	RRT calls per 1000 patient days increased from 7.8 to 16.4 after implementation

Author Year; Setting	Study design	Observation period	Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)	Score that triggers activation of response; Description of Response	In-hospital outcomes: • Mortality (30-day) • Cardiovascular events	Resource utilization: • length of hospital stay; • ICU	Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing
Mitchell 2010[19]; Australia 2 teaching hospitals	Before-after;	Before: 4mo; implementation: 8mo; after: 4mo	Before: n=1157, age 58.6 +/-19.7, male 55.7%; after: n=985, age 57.4 +/-19.8, male 54.8%	≥4	NR	ICU admissions: 1.8% vs 0.5%, RRR 0.28 (95% CI 0.11-0.74): LOS: 4.0 (1.8-8.3) vs 4.8 (2.2-9.8), p=0.03	NR
Moon, 2011[20]; UK academic tertiary care hospital	Pre-post retrospective analysis	8yr: pre: 2002-2005 post: 2006-2009	Pre: 213,117 post: 235,516; demographics NR; all patients admitted to the hospital	NR (total 21); response NR	Mortality of patients undergoing in-hospital CPR, %: pre 42 vs post 52, p=0.05; hospital deaths decreased from 750/3001 to 697/2789; CV outcomes, %: pre 0.2 vs post 0.4, p<0.0001; cardiac arrests, %: 0.35 to 0.25	Hospital: NR ICU: admissions after undergoing in-hospital CPR, %: pre 2 vs post 3, p=0.004; ICU admission after CPR, %: pre 3 vs post 2; Increase in total admissions to ICU: 857/yr to 1135/yr	Decrease in proportion of code blue calls: pre 723/213,117 vs 668/235,516; p<0.0001
Odell, 2002[14]; UK surgical wards	Pre-retrospective vs post-prospective	7mo pre with 3mo pilot	N=NR; surgical patients	3 (total 15); referral to patient's medical team and the critical care outreach team	NR	NR	Calls to outreach increased from 432 in 7mo to 231 in 3mo; 70% found the outreach service "beneficial"

Author Year; Setting	Study design	Observation period	Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)	Score that triggers activation of response; Description of Response	In-hospital outcomes: • Mortality (30-day) • Cardiovascular events	Resource utilization: • length of hospital stay; • ICU	Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing
Patel, 2011[17]; UK trauma and orthopedic wards	Retrospective cohort - pre-post	7yr: pre: 2002-2005; post: 2006-12/2009	N=32149 admissions; % male: 55; trauma and orthopedic ward patients	>4 (total 21); score >4: nursing to seek senior medical advice, referral to RRT (critical care outreach team)	Proportion of percentage admissions to deaths, pre vs post: males 0.4% (95% CI 0.003-0.81) p=0.214; females 1.5% (95%CI 0.81-2.21), p=0.108; total 0.9% (95%CI 0.53-1.31%), p= 0.092	NR	NR
Smith, 2006[16]; UK general hospital	Retrospective cohort with matched control (Legionnaire negative patients)	August 3-22, 2002 during outbreak	N= 89 Legionnaires positive, 2045 record sets vs 100 negative, 1700 record sets; median age: 64.7 vs 61.0, p<0.03; ward patients with suspected Legionnaire's disease during an outbreak	3 (total 17); referral for critical care advice	NR	NR	Median 4.9 observation sets /patient day median 3.6 EWS/patient day; UO and LOC inconsistently recorded 2036/3739 (54.4%) observation sets contained a correct EWS; RR had highest scoring errors 264/2757 errors (9.6%) vs HR 5.4%, SBP 4.3% vs Temp 3.9%; 66/270 (24.4%) observation sets were underscored and should have triggered an intervention but did not proportion of incorrect EWS scores higher in the LP group (17%) vs LN group (12%) for difference of 5% (95% CI 0-10.7), p=0.02

Author Year; Setting	Study design	Observation period	Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)	Score that triggers activation of response; Description of Response	In-hospital outcomes: • Mortality (30-day) • Cardiovascular events	Resource utilization: • length of hospital stay; • ICU	Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing
Subbe, 2003[12]; UK general hospital 56-bed medical admissions unit	Pre-retrospective vs post-prospective; patients admitted to same unit in February of the previous year served as pre-intervention comparison	February-April, 2001	Study group n=1695 vs control n=659; age: 64 vs 63; % male: 4 vs 45	>4 (total 15); doctors examined patients within 1hr	Scores 0-2, %: 6 vs 6; score 3-4, %: 17 vs 13, p=0.29; score 5-15, %: 28 vs 20, p=0.25; ICU mortality, %: 33 vs 67, p=0.21; increased incidence of cardiopulmonary arrest in study group with score of 3 or 4, %: 5 vs 1, p<0.016	Hospital: NR ICU: LOS in ICU was 2dy in study group vs 4dy in control, p=0.3	NR

DISCUSSION

We found that early warning systems, mostly using vital sign abnormalities, seem to reasonably predict the occurrence of cardiac arrest and death within 48 hours of measurement. In general, patients with favorable EWS scores were unlikely to suffer imminent cardiac arrest or death. While patients with less favorable scores had higher rates of these adverse outcomes, a large majority did not, purporting their low sensitivity. Studies evaluating the impact of implementing EWS scores on patient outcomes and resource utilization are insufficient. The results suggest that the use of rapid response teams will increase as will the manpower to obtain, record, and react to triggering scores, but the impact on health outcomes and length of hospital or ICU stays remain uncertain.

The predictive ability of these systems for unfavorable outcomes is expected given that all of these scores are based on core signs of physiological function which have been recognized as vital to life. The term "vital sign" was first coined by Dr. Edward Seguin in 1886 and included temperature, heart rate, and respiratory rate.[25] Blood pressure became a part of vital sign recording by the 20th century and more recently level of consciousness has been added in response to needs of the more prevalent level one trauma hospitals.[25] Although it is understood that significant alterations in these signs of life are ominous, there has remained uncertainty about the utility in recognizing early changes or some combination of early changes and whether identifying these patients improves outcomes. Our results reveal that most scoring systems have good predictive ability for cardiac arrest and mortality but studies thus far have been inconclusive as to predicting other events given the inadequacy of the evidence. We do know that their use is at the expense of increasing hospital resources.

A priori, we elected to not evaluate studies that reported predictive values for EWS on outcomes such as ICU admission because vital sign abnormality is an indication for critical care monitoring. Additionally, we only included studies that reported on critical events within 48 hours since the scores are intended to identify patients with more immediate critical needs so that interventions can be implemented before the actual event occurs. Death occurring beyond the acute time frame may actually reflect a subsequent change rather than the score that triggered the initial intervention. Kellet and colleagues, in addition to reporting death within 48 hours, also reported on death at multiple time points up to 30 days. The AUROC for all patients, including 2% admitted to an ICU, was 0.93 for death at 48 hours (14% death rate), it retained good predictive ability over time as the AUROC was 0.87 for death at 5 days and 0.81 at 30 days. Thus, it is likely that some of the strong predictive ability of an EWS to predict both short and long-term adverse outcomes is because patients with profound vital sign abnormalities are at high risk of death, regardless of specific interventions or their timeliness.

While EWS scores do have the ability to identify patients at risk of clinical deterioration, there is limited data on the impact of their implementation on patient outcomes. Despite very good diagnostic accuracy, the use of pulmonary artery catheters proved not to improve mortality among adult patients admitted to intensive care units and its use among medical patients substantially decreased.[26,27] Similarly, EWS may be predictive of important outcomes but their impact on clinically important outcomes such as mortality has not been established. Although we found some evidence on the impact of implementation on mortality and cardiac arrest, we found no evidence on additional outcomes such as other cardiac events, acute coronary syndrome, use of vasopressors, number of ventilator days, respiratory failure, or quality of life at discharge. Two issues arise in

studies designed to address these questions. The first is that early warning systems are frequently implemented as part of a more comprehensive critical care outreach system. In these instances, it is difficult to distinguish between the impact of the scoring system and the impact of the outreach team. We have found that the use of EWS increases the use of critical care teams and we question whether this intervention has provided the benefit rather than the EWS itself. The second issue involves controlling appropriately for the effects of time during comparative before and after study periods. Mortality or cardiac arrest numbers may decrease during the intervention phase of an EWS trial, however it is difficult to know whether this was a trend occurring naturally or in relation to concomitant and unrelated system change. Even when EWS implementation is an isolated variable in a system, there is variability in both the threshold for triggering response as well as the response that is triggered. This heterogeneity raises the possibility that it is the response to the abnormal EWS score that ultimately impacts patient outcomes rather than the use of the EWS.

Given the paucity of studies with optimal methodology, we elected to take a best evidence approach to understanding the literature by including studies providing the strongest evidence for outcomes of interest.[3] However, the included studies did have significant methods and design limitations. When using studies with historical controls, there is the risk of unmeasured confounding variables and temporal bias. For example, it is very likely that vital signs are not recorded at random but are instead measured after or because of a clinical change in the patient. Thus, all these studies likely suffer from a confounding by indication bias. Furthermore, these studies likely are limited by immortal time bias since patients without adverse events have more time available for vital sign measurement.

Additionally, advances in medical care or other changes in practice cannot be ruled out as the cause of the outcomes. For example, adoption of a highly successful sepsis campaign might influence the outcome of mortality, independent of the study intervention. The strongest study design using historic controls will be: close in time compared to intervention; utilize accurate databases with little missing data; assess objective outcomes not likely to be influenced by lack of provider, participant or assessor blinding; and report a large magnitude of effect. Additionally, the best historic control group studies report both means and rate of change for key outcomes pre-intervention (i.e., the slope of the outcome) and compare this to the rate of change (slope) of the outcome for a period of time following intervention implementation. None of the included studies employed this method of slope comparison for the outcomes of interest.

Other concerns revolve around the resources needed to implement EWS. A prospective study in Amsterdam compared nurses trained to those untrained in MEWS for recognizing signs of a simulated deteriorating patient and responses to their assessments. They found that trained nurses were better able to identify and react but still missed multiple elements of the MEWS, failed to calculate the aggregate score and failed to take action informing the physician, which was part of their trained protocol.[28] Other studies also reveal that accuracy and compliance of scoring is low[15] particularly with respiratory rate, urinary output and more subjective elements such as level of consciousness.[11,16] Calculation errors are common and although improved with an electronic system, responses to the triggering score remain variable.[29-31]

The applicability of our findings to real-world clinical settings is supported by several features of the body of literature we reviewed. First, the populations studied included a broad representation of patients on general medicine or surgical wards in both general hospitals and academic,

tertiary care settings. The interventions or intermediate actions were similar including increase in nursing attention and activation of a critical care team or rapid response team. Other features of the studies we identified, however, limit the applicability of our findings. First, characteristics of training and implementation were varied, ranging from minimal to extensive inservice and ongoing training. Second, triggering scores to initiate increased attention or activation of a critical care outreach team were heterogeneous, as was the subsequent response. Recognizing deterioration of a patient's status often involves multiple steps, many of which occur simultaneously rather than sequentially, limiting the ability to detect the importance of each step independent of the other including the implementation of the EWS itself. Additionally, given the observational study designs, it is difficult to generalize findings as outcomes may be influenced by other features of the local environment and system changes independent of the intervention.

The most important potential limitation of our review is that due to the paucity of the evidence, we elected to follow a 'best evidence' approach and included studies of low methodological quality, prone to inaccuracies and leaving us with low confidence in the results. This is reflective of the early nature of this literature with the understanding that future research may substantially change our impression. It does help to inform the inadequacies of the current body of evidence in order to help shape the direction of future research. Secondly, we only searched English language studies and although we performed a comprehensive search strategy, important studies whose findings might influence clinical and policy decision making may not have been identified. Our review focused on what we identified as clinically important and temporally related outcomes of EWS systems. However, except to assess the impact on resources, we did not evaluate intermediate outcomes such as rapid response team calls and outcomes associated with their activation as this was beyond the scope of this review and it was felt that these intermediate outcomes do not necessarily reflect the important clinical outcomes. Finally, we did not evaluate or compare the different cut-offs that triggered an intervening response given that both the cut-offs and the responses were heterogeneous.

The major gaps in research identified by our review relate to the limitations of the evidence base as described above. Future studies that are randomized trials, with more rigorous adherence to methodological standards for observational studies including an active control arm with standardization of comparators and more standardized and clinically meaningful outcome measures (i.e., 48 hour cardiac or pulmonary arrest, 48 hour mortality, number of ventilator days, quality of life at discharge), are needed to inform clinical practice and policy. Standardization of cut-offs to trigger a response and standardization of responses would improve the applicability of study findings. Decisions about defining other aspects of patient populations, interventions, comparators, outcomes, study timing and duration, and study settings should be guided by clinical practice, expertise, and factors most relevant to decision makers, including patients, clinicians, and policymakers. Institutions implementing EWS systems should prospectively track use of resources while measuring these outcomes to gain better understanding of the true impact.

CONCLUSION

Current early warning scoring systems perform well for predicting death within 48 hours although the impact on health outcomes remains uncertain given the increased use of hospital resources. Efforts to better test and improve their performance and effectiveness are needed as use becomes more widespread.

REFERENCES

1. Whittington J, White R, Haig K, Slock M. Using an Automated Risk Assessment Report to Identify Patients at Risk for Clinical Deterioration. The Joint commission Journal on Quality and Patient Safety 33(9): 569-574.2007.

2. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. *Cochrane Database Syst Rev.* 2009(3):CD005529.

3. Treadwell J, Reston J, Singh S, Talati R, McPheeters M. A Framework for "Best Evidence" Approaches in Systematic Reviews. Methods Research Report. (Prepared by the ECRI Institute Evidence-based Practice Center under Contract No. HHSA 290-2007-10063-I.) AHRQ Publication No. 11-EHC046-EF. Rockville, MD: Agency for Healthcare Research and Quality. June 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

4. Iezzoni LI, Ed. *Risk adjustment for measuring health care outcomes. 3rd ed.* . Chicago, IL: Health Administration Press; 2003.

5. Ohman EM, Granger CB, Harrington RA, Lee KL. Risk stratification and therapeutic decision making in acute coronary syndromes. *JAMA.* Aug 16 2000;284(7):876-878.

6. Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *American Journal of Epidemiology.* Nov 1 2001;154(9):854-864.

7. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine.* Feb 19 2013;158(4):280-286.

8. Rothschild JM, Gandara E, Woolf S, Williams DH, Bates DW. Single-parameter early warning criteria to predict life-threatening adverse events. *J Patient Saf.* Jun 2010;6(2):97-101.

9. Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP. Derivation of a cardiac arrest prediction model using ward vital signs*. *Crit Care Med.* Jul 2012;40(7):2102-2108.

10. Maupin JM, Roth DJ, Krapes JM. Use of the Modified Early Warning Score decreases code blue events. *Jt Comm J Qual Patient Saf.* Dec 2009;35(12):598-603.

11. Jones S, Mullally M, Ingleby S, Buist M, Bailey M, Eddleston JM. Bedside electronic capture of clinical observations and automated clinical alerts to improve compliance with an Early Warning Score protocol. *Crit Care Resusc.* Jun 2011;13(2):83-88.

12. Subbe CP, Davies RG, Williams E, Rutherford P, Gemmell L. Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions. *Anaesthesia.* Aug 2003;58(8):797-802.

13. Churpek MM, Yuen TC, Huber MT, Park SY, Hall JB, Edelson DP. Predicting cardiac arrest on the wards: a nested case-control study. *Chest.* May 2012;141(5):1170-1176.

14. Odell M, Forster A, Rudman K, Bass F. The critical care outreach service and the early warning system on surgical wards. *Nurs Crit Care.* May-June 2002;7(3):132-135.

15. De Meester K, Das T, Hellemans K, et al. Impact of a standardized nurse observation protocol including MEWS after Intensive Care Unit discharge. *Resuscitation.* Jul 14 2012.

16. Smith AF, Oakey RJ. Incidence and significance of errors in a patient 'track and trigger' system during an epidemic of Legionnaires' disease: retrospective casenote analysis. *Anaesthesia.* 2006;61(3):222-228.

17. Patel MS, Jones MA, Jiggins M, Williams SC. Does the use of a "track and trigger" warning system reduce mortality in trauma patients? *Injury.* Dec 2011;42(12):1455-1459.

18. Kellett J, Kim A. Validation of an abbreviated Vitalpac Early Warning Score (ViEWS) in 75,419 consecutive admissions to a Canadian regional hospital. *Resuscitation.* Mar 2012;83(3):297-302.

19. Mitchell IA, McKay H, Van Leuvan C, et al. A prospective controlled trial of the effect of a multi-faceted intervention on early recognition and intervention in deteriorating hospital patients. *Resuscitation.* Jun 2010;81(6):658-666.

20. Moon A, Cosgrove JF, Lea D, Fairs A, Cressey DM. An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR. *Resuscitation.* Feb 2011;82(2):150-154.

21. Green AL, Williams A. An evaluation of an early warning clinical marker referral tool. *Intensive Crit Care Nurs.* Oct 2006;22(5):274-282.

22. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation.* Jan 4 2013.

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

24. Albert BL, Huesman L. Development of a modified early warning score using the electronic medical record. *Dimens Crit Care Nurs.* Sep-Oct 2011;30(5):283-292.

25. Stewart JV. *Vital Signs and Resuscitation*. Georgetown, Texas: Landes Bioscience; 2003.

26. Rajaram SS, Desai NK, Kalra A, et al. Pulmonary artery catheters for adult patients in intensive care. *Cochrane Database of Systematic Reviews.* 2013;2:CD003408.

27. Wiener RS, Welch HG. Trends in the use of the pulmonary artery catheter in the United States, 1993-2004. *JAMA.* Jul 25 2007;298(4):423-429.

28. Ludikhuize J, de Jonge E, Goossens A. Measuring adherence among nurses one year after training in applying the Modified Early Warning Score and Situation-Background-Assessment-Recommendation instruments. *Resuscitation.* Nov 2011;82(11):1428-1433.

29. Subbe CP, Gao H, Harrison DA. Reproducibility of physiological track-and-trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med.* Apr 2007;33(4):619-624.

30. Prytherch DR, Smith GB, Schmidt P, et al. Calculating early warning scores--a classroom comparison of pen and paper and hand-held computer methods. *Resuscitation.* Aug 2006;70(2):173-178.

31. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalized patients on the wards. Chest. Jun 1 2013;143(6):1758-1765.

APPENDIX A. SEARCH STRATEGIES

PubMed searched 1/23/2013 & 1/30/2013

Search	Key Question	Concept	Search Terms
A	All	Early Warning Scoring Systems (general) *EWSS – EWSS General*	((track and trigger{Title/Abstract})) OR (((((("Hospital Information Systems"{Mesh}) OR "Risk Assessment/methods"{Mesh}) OR "Point-of-Care Systems"{Mesh}) OR "Monitoring, Physiologic/methods"{Mesh})) OR ((clinical deterioration{Title/Abstract}) OR risk assessment report{Title/Abstract})) OR ((("early warning"{Title/Abstract}) OR "warning system*"{Title/Abstract}) OR "warning scoring"{Title/Abstract}))
B	All	Named Systems *EWSS – EWSS Named*	((((((((((bispebjerg{Title/Abstract})) OR (physiological scoring system{Title/Abstract})) OR ((vital sign{Title/Abstract}) AND score{Title/Abstract})) OR (worthing{Title/Abstract})) OR (vialpac{Title/Abstract})) OR (sbar{Title/Abstract})) OR (situation background assessment recommendation{Title/Abstract})) OR (mews{Title/Abstract})) OR (manchester triage system{Title/Abstract})) OR (biosign{Title/Abstract})
C	All	Either of the Above	A OR B
D	KQ1	Predictive Value *EWSS - Predictive Value*	((((((((((C statistic*{Title/Abstract})) OR (Likelihood ratio{Title/Abstract})) OR (expected to observed{Title/Abstract})) OR (calibration{Title/Abstract})) OR ("Calibration"{Mesh})) OR (area under curve{Title/Abstract})) OR ("Area Under Curve"{Mesh})) OR ((((PPV{Title/Abstract}) OR positive predictive value{Title/Abstract}) OR NPV{Title/Abstract}) OR negative predictive value{Title/Abstract})) OR ("Predictive Value of Tests"{Mesh})) OR (((Receiver Operating{Title/Abstract}) OR Receiver Operator{Title/Abstract}) OR ROC{Title/Abstract})) OR ("ROC Curve"{Mesh})) OR (discriminative function{Title/Abstract})
E	KQ2	Mortality *EWSS - Mortality*	((mortality{Title/Abstract})) OR ("Hospital Mortality"{Mesh})
F	KQ2	Length of Stay *EWSS - LOS*	(("length of stay"{Title/Abstract})) OR ("Length of Stay"{Mesh})
G	KQ2	Coronary Outcomes *EWSS - Coronary*	(((((cardiac arrest{Title/Abstract}) OR cardiogenic shock{Title/Abstract}) OR ACS{Title/Abstract}) OR acute coronary syndrome{Title/Abstract})) OR (("Shock, Cardiogenic"{Mesh}) OR "Acute Coronary Syndrome"{Mesh})

Search	Key Question	Concept	Search Terms
H	KQ2	Respiratory Failure *EWSS - Resp*	(("Respiratory Insufficiency"{Mesh})) OR (respiratory failure{Title/Abstract})
I	KQ2	Sepsis *EWSS - Sepsis*	(("Sepsis"{Mesh}) OR "Shock, Septic"{Mesh})) OR ((sepsis{Title/Abstract}) OR septic{Title/Abstract})
J	KQ2a	ICU admissions *EWSS - ICU*	(((((ICU{Title/Abstract}) OR intensive care unit{Title/Abstract})) AND ((admission{Title/Abstract}) OR admissions{Title/Abstract}))) OR (("Patient Transfer"{Mesh}) OR "Intensive Care Units/utilization"{Mesh})
K	KQ2a	Use of rapid response teams *EWSS - RRT*	((((((medical emergency team{Title/Abstract}) OR rapid response team{Title/Abstract}) OR RRT{Title/Abstract}) OR critical care outreach team{Title/Abstract}) OR patient at risk team{Title/Abstract})) OR (("Patient Care Team"{Mesh}) OR "Hospital Rapid Response Team"{Mesh})
X		Animals	Other Animals
Y		Children	Child: birth-18 years
Z		Labor and Fetal Monitoring *EWSS - Fetal*	("Fetal Monitoring"{Mesh}) OR "Labor, Obstetric"{Mesh}

Key Question	Search
KQ 1	Predictive Value (C AND D) Not (X,Y OR Z)
KQ2	Mortality (C AND E) Not (X,Y OR Z)
	Length of Stay (C AND F) Not (X,Y OR Z)
	Coronary outcomes (C AND G) Not (X,Y OR Z)
	Respiratory Failure (C AND H) Not (X,Y OR Z)
	Sepsis (C AND I)
KQ2a	ICU Admissions (C AND J) Not (X,Y OR Z)
	Rapid Response Teams (C AND K) Not (X,Y OR Z)

CINAHL searched 3/15/13

KQ	Concept	Search Terms
All	Early Warning Scoring Systems (general)	(MH "Hospital Information Systems") OR (MH "Risk Assessment/MT") OR "point of care systems" OR (M H "Monitoring, Physiologic/MT") OR "t rack and trigger" OR "clinical deterioration" OR "risk assessment report" OR "early warning" OR "warning system*" OR "warning scoring"
	Named Systems	(TI vital sign AND TI score) OR TI bispebjerg OR TI physiological scoring system OR TI worthing OR TI vialpac OR TI sbar OR TI situation background assessment recommendation OR TI mews OR TI manchester triage system OR TI biosign
Either or the above		
KQ1	Predictive value	(M H "Calibration") OR (MH "ROC Curve") OR (MH "Predictive Value of Tests") OR TI C statistic* OR TI likelihood ratio OR TI expected to observed OR TI calibration OR TI area under curve OR TI ppv OR TI positive predictive value OR TI NPV OR TI negative predictive value OR TI receiver operating OR TI receiver operator OR TI ROC OR TI discriminative function
KQ2	Mortality	(MH "Hospital Mortality") OR TI mortality
	Length of Stay	(MH "Length of Stay") OR "length of stay"
	Coronary outcomes	(M H "Shock, Cardiogenic") OR "cardiogenic shock" OR "ac cute coronary syndrome" OR (MH "Acute Coronary Syndrome") OR "cardiac arrest" OR TI ACS
	Respiratory failure	(M H "Respiratory Failure") OR "respiratory failure"
	Sepsis	"sepsis" OR (MH "Shock, Septic") OR (MH "Sepsis") OR "septic"
KQ2a	Icu admit	((TI ICU OR TI Intensive care unit) AND (TI Admission OR TI Admission)) OR "patient transfer" OR (M H "Intensive Care Units/UT")
	Use of rapid response teams	TI Medical emergency team OR "patient care team" OR "rapid response team" OR TI critical care outreach team OR TI RRT OR TI patient at risk team
All	Children	Limit: all child
	Labor and fetal monitoring	(MH "Fetal Monitoring") OR (MH "Labor")

KQ	All concepts below are ANDed with the search for early warning scoring systems and have fetal monitoring, labor and children NOTed out
KQ1	Predictive value
KQ2	Mortality
	Length of Stay
	Coronary outcomes
	Respitory failure
	Sepsis
KQ2a	ICU admissions
	Use of rapid response teams
All above searches de-duplicated against each other for a total of:	

Search Strategy for Cochrane Central Register of Controlled Trials (OVID) April 5, 2013

#	Searches
1	hospital information systems.mp. or exp Hospital Information Systems/
2	exp Risk Assessment/
3	exp Point-of-Care Systems/
4	exp Monitoring, Physiologic/
5	(track and trigger).mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
6	clinical deterioration.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
7	risk assessment report.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
8	early warning.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
9	warning system*.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
10	warning scoring.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
11	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12	vital sign.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
13	score.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
14	12 and 13
15	bispebjerg.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
16	physiological scoring system.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
17	worthing.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
18	vialpac.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
19	sbar.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
20	situation background assessment recommendation.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
21	mews.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
22	manchester triage system.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
23	biosign.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
24	14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23
25	11 or 24
26	calibration.mp. or exp Calibration/
27	area under curve.mp. or exp Area Under Curve/
28	predictive value of tests.mp. or exp "Predictive Value of Tests"/
29	roc curve.mp. or exp ROC Curve/
30	c statistic*.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
31	likelihood ratio.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
32	expected to observed.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
33	calibration.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
34	area under curve.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
35	ppv.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
36	positive predictive value.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
37	npv.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
38	negative predictive value.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
39	receiver operating.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}

40	reciever operator.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
41	roc.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
42	discriminative function.mp. {mp=title, original title, abstract, mesh headings, heading words, keyword}
43	26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42
44	25 and 43

KQ1 predictive value

#	Searches
45	mortality.mp. or exp Hospital Mortality/
46	length of stay.mp. or exp "Length of Stay"/
47	cardiogenic shock.mp. or exp Shock, Cardiogenic/
48	acute coronary syndrome.mp. or exp Acute Coronary Syndrome/
49	cardiac arrest.mp. or exp Heart Arrest/
50	acs.mp.
51	47 or 48 or 49 or 50
52	respiratory failure.mp. or exp Respiratory Insufficiency/
53	sepsis.mp. or exp Sepsis/
54	exp Shock, Septic/ or septic.mp.
55	45 or 46 or 51 or 52 or 54
56	25 and 55

KQ2 patient outcomes

Search Strategy:

#	Searches
57	icu.mp.
58	intensive care unit.mp.
59	57 or 58
60	admission.mp.
61	admissions.mp.
62	60 or 61
63	59 and 62
64	exp Patient Transfer/
65	exp Intensive Care Units/
66	63 or 64 or 65
67	exp Patient Care Team/
68	exp Hospital Rapid Response Team/
69	medical emergency team.mp.
70	rapid response team.mp.
71	rrt.m_titl.
72	critical care outreach team.m_titl.
73	patient at risk team.m_titl.
74	67 or 68 or 69 or 70 or 71 or 72 or 73
75	66 or 74
76	25 and 75

APPENDIX B. INCLUSION/EXCLUSION CRITERIA

Code	Definition	Exclusion criteria/notes	KQ1 – Predictive value	KQ2 –Patient health outcomes and resource utilization
X1	Non-English language			
X2	Not relevant to topic			
X3	Intervention not in scope	Exclude if the risk factors examined, methods of risk factor/exposure measurement, or early warning systems studied are not in scope.	**Included interventions:** any Early Warning Systems (EWS) or other established scoring system designed to identify deteriorating patients on general hospital wards. Potential interventions include but are not limited to the following: Modified Early Warnings Systems (MEWS), Patient at Risk (PAR) score, Physiological Scoring Systems (PSS), Vital Sign Score (VSS), Manchester Triage System, BioSign, ViaIPAC Earl Warning Score (ViEWS), and Physiological Observation Track and Trigger System (POTTS).	Same interventions as KQ1
X4	Study population or setting not in scope	Excluded populations: obstetric, fetal, or pediatric patients; Excluded settings: emergency department, ICU, outpatient.	Admitted patients on general medicine or surgical wards.	Same populations/settings as KQ1
X5	No primary data or study design not in scope	Excluded study designs: non-systematic or narrative reviews, opinions, case series	English language validity studies: include if the article is an observational study that reports measures of predictive ability (e.g., discriminative factor, concordance statistic, sensitivity, specificity, likelihood ratios, positive predictive value, and/or negative predictive value) and validates the predictive model.	Controlled studies: randomized controlled trials, controlled clinical trials, before and after studies, and interrupted-time-series designs. Also include: English language systematic reviews of controlled studies.

Code	Definition	Exclusion criteria/notes	KQ1 – Predictive value	KQ2 –Patient health outcomes and resource utilization
X7	Outcomes that are not in scope		Patient health outcomes occurring within 72 hours of EWSS data collection/analysis: • short-term mortality (all cause or disease specific) • cardiac arrest • pulmonary arrest	Outcomes for KQ2 generally apply to in-hospital stay: • mortality (in-hospital or mortality up to 30-days; all cause or disease-specific) • in-hospital cardiovascular events (cardiac arrest, acute coronary syndrome, cardiogenic shock) • use of pressors • number of ventilator (or ventilator-free) days • respiratory failure • length of hospital stay • resource utilization outcomes including ICU admission and use of rapid response teams
X8	Other reason: specify	Add comments or keywords as needed.		
X99	Full text not accessible			
B	Background	Add to any of the above X codes (e.g., X6–B) if the article contains information that may be useful for the introduction, discussion, limitations, future research, or other contextual purposes. Add comments or keywords as needed.		
I-1 **I-2**			Validity studies that address: KQ1a: What is the predictive value in using EWS scores for predicting patient health outcomes? KQ1b: Which factors contribute to the predictive ability of EWS scores? KQ1c: Does predictive ability of EWS scores vary with specific subgroups of patients?	Controlled studies that address: KQ2a: What are the effects of EWS on health outcomes including mortality, morbidity, cardiac arrest, and pulmonary arrest? KQ2b: What is the impact of EWS on resource utilization (e.g., ICU admissions, length of hospital stay, use of Rapid Response Teams) and provider/nurse satisfaction?
I-SR	Include – systematic review		Systematic review or meta-analysis that addresses either KQ1 or KQ2.	

APPENDIX C. PEER REVIEW COMMENTS AND RESPONSES

	Reviewer	Comment	Response
	Question 1: Are the objectives, scope, and methods for this review clearly described?		
1.	1	Yes	
2.	2	Yes	
3.	3	Yes	
4.	4	Yes	
	2. Is there any indication of bias in our synthesis of the evidence?		
5.	1	No	
6.	2	No	
7.	3	No	
8.	4	No	
	3. Are there any published or unpublished studies that we may have overlooked?		
9.	1	No	
10.	1	Very thorough.	
11.	2	Yes	
12.	2	Chan, P. et. al. (2010). Rapid Response Teams: A systematic review and analysis. Archives of internal med. 170(1), 18-26.	Although we did look at the impact of RRT's as it applied to the implementation of an EWS, we did not consider the use of rapid response teams in isolation of the EWS. The use of RRT's is not a direct indication of patient benefit.
13.	2	Winters, B. et. al. (2013). Rapid response systems as a patient safety strategy. A systematic review. Annals of Internal Medicine. 158(5) part 2, 417-42. (this one may have been published in May of 2013 and may have missed your date range)	As above, the direct use of rapid response teams is not the primary objective of this report.
14.	3	No	
15.	4	Yes	
16.	4	Resuscitation. 2013 Sep 20. pii: S0300-9572(13)00737-5. doi: 10.1016/j.resuscitation.2013.08.277. [Epub ahead of print] **CREWS: Improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia.** Eccles SR, Subbe C, Hancock D, Thomson N. **Source** Wrexham Maelor Hospital, Croesnewydd Road, Wrexham LL13 7TD, UK. Electronic address: sinaneccles@gmail.com.	Thank you for directing us to these additional resources. The CREW study was outside of our scope given that it was in a population of patients with chronic hyoxaemia rather than the general ward population.

	Reviewer	Comment	Response
	4 (cont.)	**Abstract** **BACKGROUND:** The National Early Warning Score (NEWS) is being introduced across the UK, but there are concerns about its specificity in patients with chronic hypoxaemia, such as some patients with COPD. This could lead to frequent clinically insignificant triggers and alarm fatigue. **AIMS OF STUDY:** To investigate whether patients with chronic hypoxaemia trigger excessively with NEWS, and to design a simple variant of NEWS for patients with chronic hypoxaemia: a Chronic Respiratory Early Warning Score (CREWS). **METHODS:** Data was collected from respiratory wards at 2 hospitals in North Wales. Components of NEWS and frequency of trigger thresholds being reached were recorded. CREWS was applied retrospectively to patients' observations. **RESULTS:** 196 admissions were analysed, including 78 for patients with chronic hypoxaemia. Patients with chronic hypoxaemia frequently exceeded trigger thresholds using NEWS during periods of stability/at discharge. Using CREWS, triggers during stability/at discharge were reduced from 32% of observations to 14% using a trigger threshold of a score greater than 6, and from 50% to 18% using a score greater than 5. All patients with chronic hypoxaemia who died within 30 days still reached CREWS trigger thresholds, and the area under receiver operated curves for NEWS and CREWS was comparable. **CONCLUSION:** CREWS is a simple variant of NEWS for patients with chronic hypoxaemia that could reduce clinically insignificant triggers and alarm fatigue, whilst still identifying the sickest patients. Copyright © 2013 Elsevier Ireland Ltd. All rights reserved. **KEYWORDS:** COPD, Death, Early Warning Score, Hypoxia, NEWS, Pulmonary fibrosis ********** Resuscitation. 2013 Aug 17. pii: S0300-9572(13)00421-8. doi: 10.1016/j.resuscitation.2013.08.006	

	Reviewer	Comment	Response
17.	4 (cont.)	PLoS One. 2013 Jul 26;8(7):e70068. doi: 10.1371/journal.pone.0070068. Print 2013. **Imperfect implementation of an early warning scoring system in a Danish teaching hospital: a cross-sectional study. (might not meet criteria for inclusion)** Niegsch M, Fabritius ML, Anhøj J. **Source** Anaesthesiology Department Z, Bispebjerg Hospital, Copenhagen, Denmark. mark@niegsch.dk **Abstract** **BACKGROUND:** In 2007, the initiation of a patient safety campaign led to the introduction of Ward Observational Charts (WOC) and Medical Early Warning Score (MEWS) at Naestved Regional Hospital. This included systematic measuring of vital signs of all patients in order to prevent patient deterioration and assure timely and correct initiation of treatment. The aim of this study was to assess to what degree WOC guidelines being followed by ward staff. **DESIGN AND SETTING:** A 7-day prospective, observational, randomised, cross-sectional, point prevalence study of WOC guideline compliance in hospitalised patients on twelve wards at Naestved Hospital. **RESULTS:** The study included 132 patients. Of these, 58% had been observed and managed correctly according to WOC guidelines. 77% had all MEWS elements recorded by staff. One patient had no MEWS elements recorded. Only 38% of patients with abnormal MEWS were correctly escalated by nursing staff. Staff was aware of the abnormal MEWS observed by investigator in 60% of the patients. Each element of WOC was on average recorded by staff in 90% of the patients. **CONCLUSION:** At the time of our study, the long-term implementation of WOC guidelines has not been completed satisfactorily. The lacking component in the implementation of MEWS and WOC is the documentation of action taken upon finding an abnormal value. Unsuccessful implementation could result in incorrect results from evaluation of an early warning system. We suggest a redesign of the training programme to educate staff in recognising and caring for critically ill patients at Naestved Hospital. PMID: 23922906 [PubMed - in process] PMCID: PMC3724921	This study does not meet our inclusion criteria given that the outcome measured was compliancy rather than utilization of nursing resources.

	Reviewer	Comment	Response
18.	4 (cont.)	Resuscitation. 2010 Jun;81(6):658-66. doi: 10.1016/j.resuscitation.2010.03.001. Epub 2010 Apr 7. **A prospective controlled trial of the effect of a multi-faceted intervention on early recognition and intervention in deteriorating hospital patients.** Mitchell IA, McKay H, Van Leuvan C, Berry R, McCutcheon C, Avard B, Slater N, Neeman T, Lamberth P. **Source** The Canberra Hospital, Department of Intensive Care, Yamba Drive, Canberra, Garran, Australian Capital Territory, Australia. Imogen.mitchell@act.gov.au **Abstract** **AIM:** To determine whether the introduction of a multi-faceted intervention (newly designed ward observation chart, a track and trigger system and an associated education program, COMPASS) to detect clinical deterioration in patients would decrease the rate of predefined adverse outcomes. **METHODS:** A prospective, controlled before-and-after intervention of trial was conducted in all consecutive adult patients admitted to 4 medical and surgical wards during a 4 month period, 1157 and 985, respectively. A sub-group of patients underwent vital sign and medical review analysis pre-intervention (427) and post-intervention (320). The outcome measures included: number of unplanned admissions to the intensive care unit (ICU), Medical Emergency Team (MET) reviews and unexpected hospital deaths, vital sign documentation frequency and incidence of a medical review following clinical deterioration. This study is registered, ACTRN12609000808246. **RESULTS:** Reductions were seen in unplanned admissions to ICU (21/1157 [1.8%] vs 5/985 [0.5%], p=0.006) and unexpected hospital deaths (11/1157 [1.0%] vs 2/985 [0.2%], p=0.03) during the intervention period. Medical reviews for patients with significant clinical instability (58/133 [43.6%] vs 55/79 [69.6%] p<0.001) and number of patients receiving a MET review increased (25/1157 [2.2%] vs 38/985 [3.9%] p=0.03) during the intervention period. Mean daily frequency of documentation of all vital signs increased during the intervention period (3.4 [SE 0.22] vs 4.5 [SE 0.17], p=0.001). **CONCLUSION:** The introduction of a multi-faceted intervention to detect clinical deterioration may benefit patients through increased monitoring of vital signs and the triggering of a medical review following an episode of clinical instability. 	Have reviewed this paper and will include the outcome of length of hospital stay.

	Reviewer	Comment	Response
19.	4 (cont.)	Lancet. 2005 Jun 18-24;365(9477):2091-7. **Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial.** Hillman K, Chen J, Cretikos M, Bellomo R, Brown D, Doig G, Finfer S, Flabouris A; MERIT study investigators. **Erratum in** • Lancet. 2005 Oct 1;366(9492):1164. **Abstract** **BACKGROUND:** Patients with cardiac arrests or who die in general wards have often received delayed or inadequate care. We investigated whether the medical emergency team (MET) system could reduce the incidence of cardiac arrests, unplanned admissions to intensive care units (ICU), and deaths. **METHODS:** We randomised 23 hospitals in Australia to continue functioning as usual (n=11) or to introduce a MET system (n=12). The primary outcome was the composite of cardiac arrest, unexpected death, or unplanned ICU admission during the 6-month study period after MET activation. Analysis was by intention to treat. **FINDINGS:** Introduction of the MET increased the overall calling incidence for an emergency team (3.1 vs 8.7 per 1000 admissions, p=0.0001). The MET was called to 30% of patients who fulfilled the calling criteria and who were subsequently admitted to the ICU. During the study, we recorded similar incidence of the composite primary outcome in the control and MET hospitals (5.86 vs 5.31 per 1000 admissions, p=0.640), as well as of the individual secondary outcomes (cardiac arrests, 1.64 vs 1.31, p=0.736; unplanned ICU admissions, 4.68 vs 4.19, p=0.599; and unexpected deaths, 1.18 vs 1.06, p=0.752). A reduction in the rate of cardiac arrests (p=0.003) and unexpected deaths (p=0.01) was seen from baseline to the study period for both groups combined. **INTERPRETATION:** The MET system greatly increases emergency team calling, but does not substantially affect the incidence of cardiac arrest, unplanned ICU admissions, or unexpected death.	This study was outside of our inclusion criteria as it applies to the use of medical emergency teams rather than an early warning scoring system.
	4. Please write additional suggestions or comments below. If applicable, please indicate the page and line numbers from the draft report.		
20.	1	I noted in table 1 that the parameters listed sometimes described the variation (for example, decreased urine output) but other times did not (Mental Status). Could this be made clearer across all the labels?	Thank you – changes made to the tables.
21.	4	I think this was a carefully written review avoiding any claims that EWS do not work and should not be used, although I anticipate that someone at least will use your review to claim that rapid response teams should not be continued across the country.	