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
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.


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.
TABLE OF CONTENTS

BACKGROUND ..................................................................................................................... 1

METHODS
   Topic development ............................................................................................................. 2
   Key question #1 .................................................................................................................. 2
   Key question #2 .................................................................................................................. 2
   Search strategy .................................................................................................................. 4
   Study selection ................................................................................................................... 4
   Data abstraction and quality assessment .......................................................................... 4
   Data synthesis .................................................................................................................... 5
   Role of the funding source .............................................................................................. 5

RESULTS
   Literature flow .................................................................................................................... 6
   Predictive value of EWS .................................................................................................... 8
   Impact of EWS interventions ............................................................................................ 13
   Impact of EWS on health outcomes .................................................................................. 13
      Mortality .......................................................................................................................... 13
      Cardiac arrest .................................................................................................................. 13
   Impact of EWS on resources ............................................................................................. 14
      Length of hospital stay .................................................................................................... 14
      Admissions to the intensive care unit .............................................................................. 14
      Use of rapid response and code teams .......................................................................... 15
      Nursing ........................................................................................................................... 15

DISCUSSION .......................................................................................................................... 22
   Conclusion .......................................................................................................................... 24

REFERENCES .......................................................................................................................... 25

FIGURES
   Figure 1. Analytic framework ......................................................................................... 3
   Figure 2. Literature flow diagram .................................................................................... 6

TABLES
   Table 1. Parameters used in studies of the predictive ability and effectiveness of early warning scoring systems for clinical deterioration in medical and surgical inpatients ................................ 7
   Table 2. Observational studies of the predictive value of EWS scores ............................. 10
   Table 3. Assessment of study methods for potential sources of bias in studies of the predictive value of EWS scores ................................................................................. 12
   Table 4. Studies of the impact of EWS interventions on patient outcomes and resource utilization .......................................................... 16

APPENDIX A. SEARCH STRATEGIES .................................................................................... 33

APPENDIX B. INCLUSION/EXCLUSION CRITERIA .................................................................. 35

APPENDIX C. PEER REVIEW COMMENTS AND RESPONSES .................................................. 36
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. 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. 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.

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

- **EWS score**
  - Low risk
  - High risk

- **KQ1**
  - \( KQ_{2A} \)
  - \( KQ_{2B} \)

- **Modifying factors**
  - patient subgroup
  - vital signs
  - timing of use
  - hospital characteristics

- **Intermediate actions**
  - RRT
  - ICU RN call
  - MD call
  - increased surveillance
  - resource utilization
  - decisional conflict

- **Outcomes**
  - mortality
  - CV outcomes
  - respiratory failure
  - functional status
  - discharged to location
  - utilization
    - length of stay
    - hospital, ICU
  - number of ventilator days

EWS in patients hospitalized on general medicine and surgical wards
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. 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. 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. 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. 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

<table>
<thead>
<tr>
<th>Study Country</th>
<th>N parameters; name of scoring system</th>
<th>Parameters used in the system scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rothschild, 2010&lt;sup&gt;8&lt;/sup&gt; USA</td>
<td>Single items, not combined</td>
<td>Heart rate</td>
</tr>
<tr>
<td>Churpek, 2012&lt;sup&gt;9&lt;/sup&gt; USA</td>
<td>4-item CART</td>
<td>X</td>
</tr>
<tr>
<td>Maupin, 2009&lt;sup&gt;10&lt;/sup&gt; USA</td>
<td>5-item MEWS</td>
<td>X</td>
</tr>
<tr>
<td>Jones, 2011&lt;sup&gt;11&lt;/sup&gt; UK</td>
<td>Patientrack EWS</td>
<td>X</td>
</tr>
<tr>
<td>Subbe, 2003&lt;sup&gt;12&lt;/sup&gt; UK</td>
<td>5-item MEWS</td>
<td>X</td>
</tr>
<tr>
<td>Churpek, 2012&lt;sup&gt;13&lt;/sup&gt; USA</td>
<td>5-item MEWS</td>
<td>X</td>
</tr>
<tr>
<td>O’Dell, 2002&lt;sup&gt;14&lt;/sup&gt; UK</td>
<td>5-item MEWS</td>
<td>X</td>
</tr>
<tr>
<td>DeMeester, 2012&lt;sup&gt;15&lt;/sup&gt; Belgium</td>
<td>6-item MEWS</td>
<td>X</td>
</tr>
<tr>
<td>Smith, 2006&lt;sup&gt;16&lt;/sup&gt; UK</td>
<td>6-item EWS</td>
<td>X</td>
</tr>
<tr>
<td>Patel, 2011&lt;sup&gt;17&lt;/sup&gt; UK</td>
<td>6-item MEWS</td>
<td>X</td>
</tr>
<tr>
<td>Kellett, 2012&lt;sup&gt;18&lt;/sup&gt; Canada</td>
<td>6-item ViEWS</td>
<td>X</td>
</tr>
<tr>
<td>Mitchell, 2010&lt;sup&gt;19&lt;/sup&gt;</td>
<td>7-item MEWS</td>
<td>X</td>
</tr>
<tr>
<td>Moon, 2011&lt;sup&gt;20&lt;/sup&gt; UK</td>
<td>7-item MEWS</td>
<td>X</td>
</tr>
<tr>
<td>Green, 2006&lt;sup&gt;21&lt;/sup&gt; Australia</td>
<td>7-item clinical marker tool</td>
<td>X</td>
</tr>
<tr>
<td>Smith, 2013&lt;sup&gt;22&lt;/sup&gt; UK</td>
<td>7-item NEWS</td>
<td>X</td>
</tr>
<tr>
<td>Prytherch, 2010&lt;sup&gt;23&lt;/sup&gt; UK</td>
<td>7-item ViEWS</td>
<td>X</td>
</tr>
<tr>
<td>Albert, 2011&lt;sup&gt;24&lt;/sup&gt; USA</td>
<td>12-item MEWS</td>
<td>X</td>
</tr>
</tbody>
</table>
PREDICTIVE VALUE OF EWS

Six observational studies (4 prospective cohort, 2 case-control) met our inclusion criteria (Table 2).\textsuperscript{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\textsuperscript{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,\textsuperscript{13,18} 4 were derivation studies.\textsuperscript{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.\textsuperscript{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).\textsuperscript{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). 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 within 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. 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. 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. 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

<table>
<thead>
<tr>
<th>Author, Year; Model; Setting</th>
<th>Population; Age; Gender</th>
<th>Comparison</th>
<th>N</th>
<th>Predictive measures of mortality occurring within 72 hours of EWSS data collection/analysis</th>
<th>Predictive measures of cardiac arrest occurring within 72 hours of EWSS data collection/analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Churpek, 2012&lt;sup&gt;a&lt;/sup&gt;; predictive model, retrospective cohort; US academic tertiary care hospital 500 beds</td>
<td>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%</td>
<td>Pts without an event (cardiac arrest or ICU transfer)</td>
<td>See previous Cardiac arrest: 88 pts ICU Txf: 2820 pts Control: 44,519 pts</td>
<td>NR</td>
<td>AUROC: 0.84 (vs 0.78 for MEWS), 95% CI NR</td>
</tr>
<tr>
<td>Churpek, 2012&lt;sup&gt;b&lt;/sup&gt;; case-control; US academic tertiary care hospital</td>
<td>73% medical, 27% surgical; Mean age: Cases 64 Controls 58 (p=0.002); Male: Cases 43% Controls 49% (P=ns)</td>
<td>Control pts without cardiac arrest</td>
<td>88 cases, 352 controls</td>
<td>NR</td>
<td>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 &gt;3 of possible 14, the specificity for cardiac arrest was 87% with a sensitivity of 51%.</td>
</tr>
<tr>
<td>Kellett, 2012&lt;sup&gt;c&lt;/sup&gt;; predictive model; Canada urban academic hospital 375 beds</td>
<td>Med/Surg inpts; Mean age 63; 48.9% male</td>
<td>None</td>
<td>75,419 total Medical, non-ICU: 43693 Surg: 30485 (presumably includes some ICU pts)</td>
<td>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 &gt;11 died within 48 hours (86% false positives).</td>
<td></td>
</tr>
<tr>
<td>Author, Year; Model; Setting</td>
<td>Population; Age; Gender</td>
<td>Comparison</td>
<td>N</td>
<td>Predictive measures of mortality occurring within 72 hours of EWSS data collection/analysis</td>
<td>Predictive measures of cardiac arrest occurring within 72 hours of EWSS data collection/analysis</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------------------</td>
<td>------------</td>
<td>---</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Prytherch, 2010; prospective UK Hospital</td>
<td>General Medical Emergency patients consecutively admitted; Mean age 67.7 Median age 72.6; 47.5% male</td>
<td>Pts who were alive at 24 hours following observation</td>
<td>35,585 pt episodes</td>
<td>AUROC = 0.888 (95%CI 0.885-0.895) for death within 24 hrs Sensitivity was approximately 67% at a specificity of 90%.</td>
<td>NR</td>
</tr>
<tr>
<td>Rothschild, 2010 case-control, predictive model; US urban academic 745 beds</td>
<td>Medical and medicine subspeciality inpts; Mean age: 61; 49.6% male</td>
<td>Control patients matched on day of admit</td>
<td>262 pts 318 controls</td>
<td>NR</td>
<td>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.</td>
</tr>
<tr>
<td>Smith, 2013 UK predictive model; prospective Hospital</td>
<td>General Medical patients; Mean age 67.7 Median age 72.6; 94,376/198,755 observations from males</td>
<td>Pts without an event</td>
<td>See previous</td>
<td>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.</td>
<td>Cardiac arrest within 24hrs AUROC: 0.857 (95% CI 0.847-0.868)</td>
</tr>
</tbody>
</table>
Table 3. Assessment of study methods for potential sources of bias in studies of the predictive value of EWS scores

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Population: study sample adequately represents population of interest</th>
<th>Study Attrition: study data available adequately represent study sample</th>
<th>Prognostic Factor Measurement: measured in a similar way for all participants</th>
<th>Outcomes Measurement: outcome of interest measured in similar way for all participants</th>
<th>Study Confounding: Important potential confounding factors are appropriately accounted for</th>
<th>Statistical Analysis: statistical analysis is appropriate</th>
<th>Reporting: All primary outcomes reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Churpek, 2012\textsuperscript{13}</td>
<td>Yes; &quot;consecutive&quot;</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unclear: Cases might be significantly different from controls; not enough information provided</td>
<td>Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all (all AUCs include a 95% CI)</td>
<td>Yes</td>
</tr>
<tr>
<td>Kellett, 2012\textsuperscript{18}</td>
<td>Yes; &quot;every patient&quot;</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all (all AUCs include a 95% CI)</td>
<td>Yes</td>
</tr>
<tr>
<td>Prytherch, 2010\textsuperscript{23}</td>
<td>Yes; &quot;consecutive&quot;</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all (all AUCs include a 95% CI)</td>
<td>Yes</td>
</tr>
<tr>
<td>Rothschild, 2010\textsuperscript{8}</td>
<td>Unclear: No statement of “all” etc.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unclear: Cases might be significantly different from controls; not enough information provided</td>
<td>Yes but they conducted multiple comparisons and it doesn't look like they adjusted analyses at all (all AUCs include a 95% CI)</td>
<td>Yes</td>
</tr>
<tr>
<td>Smith, 2013\textsuperscript{22}</td>
<td>Yes; &quot;consecutive&quot;</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>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 &lt; .001</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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).\textsuperscript{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,\textsuperscript{11,12,17,20} but only one study found this to be statistically significant.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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).\textsuperscript{21}

Cardiac arrest

Three studies evaluating the effects of EWS implementation on cardiac arrests met inclusion criteria and found mixed results.\textsuperscript{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%)\textsuperscript{21} and 2 studies found no difference in the length of the ICU stay.\textsuperscript{12,21}

**Use of rapid response and code teams**

Four studies evaluated the impact of EWS on rapid response and code teams.\textsuperscript{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.\textsuperscript{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%.\textsuperscript{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\textsuperscript{15} and accuracy being as high as 81-100% with the use of electronic calculations.\textsuperscript{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%).\textsuperscript{16} The number of observations and clinical attention by nursing increased with the use of the EWS with greater attention for EWS scores >5.\textsuperscript{11} One study found that the frequency of patient observations per nursing shift increased during the daytime but not at night.\textsuperscript{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

<table>
<thead>
<tr>
<th>Author Year; Setting</th>
<th>Study design</th>
<th>Observation period</th>
<th>Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)</th>
<th>Score that triggers activation of response; Description of Response</th>
<th>In-hospital outcomes:</th>
<th>Resource utilization:</th>
<th>Resource utilization:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing</td>
</tr>
<tr>
<td>Albert, 2011&lt;sup&gt;24&lt;/sup&gt;; US 550-bed tertiary care academic hospital</td>
<td>Retrospective cohort; comparing MEWS score 4-8hr prior to event and at time of emergent event</td>
<td>6mo pilot study</td>
<td>N=140; age: RRT 64.7 v code 65.8; telemetry ward</td>
<td>≥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</td>
<td>NR</td>
<td>NR</td>
<td>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</td>
</tr>
<tr>
<td>De Meester, 2012&lt;sup&gt;15&lt;/sup&gt;; Belgium tertiary care hospital, 14 medical and surgical wards</td>
<td>Pre-retrospective and post-prospective</td>
<td>5dy period after ICU discharge</td>
<td>N=1039; age: 59; % male: 60; MEWS at ICU discharge: 2 5-days post ICU-discharge to medical and surgical wards</td>
<td>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</td>
<td>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%)</td>
<td>NR</td>
<td>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 1.1036-1.1101); 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)</td>
</tr>
<tr>
<td>Author Year; Setting</td>
<td>Study design</td>
<td>Observation period</td>
<td>Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)</td>
<td>Score that triggers activation of response; Description of Response</td>
<td>In-hospital outcomes: • Mortality (30-day) • Cardiovascular events</td>
<td>Resource utilization: • length of hospital stay; • ICU</td>
<td>Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-----------------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>---------------------------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Green, 2006; Australia 323 bed tertiary care academic hospital surgical and medical ward</td>
<td>Pre-retrospective and post-prospective</td>
<td>Pre: February 2002-January 2003; post: February 2003-January 2005</td>
<td>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</td>
<td>NR; activate ICU Liaison Team</td>
<td>Mortality, %: pre 33.9 vs post 34.5</td>
<td>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 readmissions; LOS: median 2.6 (1.2-6.4)</td>
<td>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&lt;0.0002</td>
</tr>
<tr>
<td>Author Year; Setting</td>
<td>Study design</td>
<td>Observation period</td>
<td>Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)</td>
<td>Score that triggers activation of response; Description of Response</td>
<td>In-hospital outcomes: • Mortality (30-day) • Cardiovascular events</td>
<td>Resource utilization: • length of hospital stay; • ICU</td>
<td>Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Jones, 2011</strong>;&lt;br&gt; UK academic hospital medical assessment unit</td>
<td>Pre-post with 3 phases: baseline, PDA with downloading to ward-view, Patient Track; hospital and clinical outcomes and LOS collected retrospectively</td>
<td>Pre: 47dy November-December, 2007; alert 38dy August-September, 2008</td>
<td>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</td>
<td>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</td>
<td>Mortality, %, pre 9.5 vs post 7.6; cardiac arrest, %: pre 0.4 vs 0</td>
<td>Hospital: 9.7dy vs 6.9dy, p&lt;0.001 ICU: pre 51dy vs 26dy, p=0.04</td>
<td>Clinical attendance with EWS 3-5, %: pre 29 vs post 78, p&lt;.001; clinical attendance with EWS &gt;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&lt;0.001; for EWS &gt;5 clinical attendance increased from 67% to 95%, p&lt;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&lt;0.001; for EWS &gt;5 clinical attendance increased from 67% to 95%, p&lt;0.001</td>
</tr>
<tr>
<td><strong>Maupin, 2009</strong>;&lt;br&gt; US 200 bed general hospital</td>
<td>Pre-post; RRT calls per 1000 patient days increased from 7.8 to 16.4 after implementation (Figure 3)</td>
<td>Pre: 1yr (2007) pilot for 2mo post: 1yr</td>
<td>N=NR; general medical, surgical, tele floors</td>
<td>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</td>
<td>NR, the cardiac arrest was listed as either outside ICU or not</td>
<td>NR</td>
<td>RRT calls per 1000 patient days increased from 7.8 to 16.4 after implementation</td>
</tr>
<tr>
<td>Author Year; Setting</td>
<td>Study design</td>
<td>Observation period</td>
<td>Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)</td>
<td>Score that triggers activation of response; Description of Response</td>
<td>In-hospital outcomes: • Mortality (30-day) • Cardiovascular events</td>
<td>Resource utilization: • length of hospital stay; • ICU</td>
<td>Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>---------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Mitchell 2010&lt;sup&gt;19&lt;/sup&gt;; Australia 2 teaching hospitals</td>
<td>Before-after; Before: 4mo; implementation: 8mo; after: 4mo</td>
<td>Before: n=1157, age 58.6 +/-19.7, male 55.7%; after: n=985, age 57.4 +/-19.8, male 54.8%</td>
<td>≥4</td>
<td>NR</td>
<td>ICU admissions: 1.6% 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</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Moon, 2011&lt;sup&gt;20&lt;/sup&gt;; UK academic tertiary care hospital</td>
<td>Pre-post retrospective analysis 8yr: pre: 2002-2005 post: 2006-2009</td>
<td>Pre: 213,117 post: 235,516; demographics NR; all patients admitted to the hospital</td>
<td>NR (total 21); response NR</td>
<td>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=0.001; cardiac arrests, %: 0.35 to 0.25</td>
<td>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</td>
<td>Decrease in proportion of code blue calls: pre 723/213,117 vs 668/235,516; p&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Odell, 2002&lt;sup&gt;14&lt;/sup&gt;; UK surgical wards</td>
<td>Pre-retrospective vs post-prospective 7mo pre with 3mo pilot</td>
<td>N=NR; surgical patients</td>
<td>3 (total 15); referral to patient’s medical team and the critical care outreach team</td>
<td>NR</td>
<td>NR</td>
<td>Calls to outreach increased from 432 in 7mo to 231 in 3mo; 70% found the outreach service “beneficial”</td>
<td></td>
</tr>
<tr>
<td>Author Year; Setting</td>
<td>Study design</td>
<td>Observation period</td>
<td>Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)</td>
<td>Score that triggers activation of response; Description of Response</td>
<td>In-hospital outcomes: • Mortality (30-day) • Cardiovascular events</td>
<td>Resource utilization: • length of hospital stay; • ICU</td>
<td>Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Patel, 2011; UK trauma and orthopedic wards</td>
<td>Retrospective cohort - pre-post</td>
<td>7yr: pre: 2002-2005; post: 2006-12/2009</td>
<td>N=32149 admissions; % male: 55; trauma and orthopedic ward patients</td>
<td>&gt;4 (total 21); score &gt;4: nursing to seek senior medical advice, referral to RRT (critical care outreach team)</td>
<td>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</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Smith, 2006; UK general hospital</td>
<td>Retrospective cohort with matched control (Legionnaire negative patients)</td>
<td>August 3-22, 2002 during outbreak</td>
<td>N=89 Legionnaires positive, 2045 record sets vs 100 negative, 1700 record sets; median age: 64.7 vs 61.0, p&lt;0.03; ward patients with suspected Legionnaires disease during an outbreak</td>
<td>3 (total 17); referral for critical care advice</td>
<td>NR</td>
<td>NR</td>
<td>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</td>
</tr>
<tr>
<td>Author Year; Setting</td>
<td>Study design</td>
<td>Observation period</td>
<td>Population N; mean age; % male; other characteristics (e.g., ward; diagnosis)</td>
<td>Score that triggers activation of response; Description of Response</td>
<td>In-hospital outcomes:</td>
<td>Resource utilization:</td>
<td>Resource utilization: Use of rapid response teams, Critical Care Outreach Teams or MET (medical emergency team); Nursing</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>Subbe, 200312; UK general hospital 56-bed medical admissions unit</td>
<td>Pre-retrospective vs post-prospective; patients admitted to same unit in February of the previous year served as pre-intervention comparison</td>
<td>February-April, 2001</td>
<td>Study group n=1695 vs control n=659; age: 64 vs 63; % male: 4 vs 45</td>
<td>&gt;4 (total 15); doctors examined patients within 1hr</td>
<td>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&lt;0.016</td>
<td>Hospital: NR ICU: LOS in ICU was 2dy in study group vs 4dy in control, p=0.3</td>
<td>NR</td>
</tr>
</tbody>
</table>
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. 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. 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. 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
Early Warning System Scores: A Systematic Review

Evidence-based Synthesis Program

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. 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. Other studies also reveal that accuracy and compliance of scoring is low particularly with respiratory rate, urinary output and more subjective elements such as level of consciousness. Calculation errors are common and although improved with an electronic system, responses to the triggering score remain variable.

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


