



HHS Public Access

Author manuscript

Resuscitation. Author manuscript; available in PMC 2019 June 09.

Published in final edited form as:

Resuscitation. 2018 February ; 123: 86–91. doi:10.1016/j.resuscitation.2017.10.028.

Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients.

Malcolm Green,

Clinical Excellence Commission, Level 17 McKell Building, 2-24 Rawson Place, Sydney 2000, New South Wales, Australia

Dr Harvey Lander,

Clinical Excellence Commission, Level 17 McKell Building, 2-24 Rawson Place, Sydney 2000, New South Wales, Australia

Ashley Snyder,

Department of Medicine, University of Chicago, 5841 South Maryland Avenue, MC 6076, Chicago, IL 60637

Paul Hudson,

Clinical Excellence Commission, Level 17 McKell Building, 2-24 Rawson Place, Sydney 2000, New South Wales, Australia

Dr Matthew Churpek, and

Department of Medicine, University of Chicago, 5841 South Maryland Avenue, MC 6076, Chicago, IL 60637

Dr Dana Edelson

Department of Medicine, University of Chicago, 5841 South Maryland Avenue, MC 6076, Chicago, IL 60637

Abstract

Introduction: Traditionally, paper based observation charts have been used to identify deteriorating patients, with emerging recent electronic medical records allowing electronic algorithms to risk stratify and help direct the response to deterioration.

Objective(s): We sought to compare the Between the Flags (BTF) calling criteria to the Modified Early Warning Score (MEWS), National Early Warning Score (NEWS) and electronic Cardiac Arrest Risk Triage (eCART) score.

Design and Participants: Multicenter retrospective analysis of electronic health record data from all patients admitted to five US hospitals from November 2008-August 2013.

corresponding author Malcolm Green Clinical Excellence Commission, Level 17 McKell Building, 2-24 Rawson Place, Sydney 2000, New South Wales, Australia, malcolm.green1@health.nsw.gov.au, Phone: +61 2 9269 5587.

Conflicts of Interest

Main outcome measures: Cardiac arrest, ICU transfer or death within 24 hours of a score

Results: Overall accuracy was highest for eCART, with an AUC of 0.801 (95% CI 0.799–0.802), followed by NEWS, MEWS and BTF respectively (0.718 [0.716–0.720]; 0.698 [0.696–0.700]; 0.663 [0.661–0.664]). BTF criteria had a high risk (Red Zone) specificity of 95.0% and a moderate risk (Yellow Zone) specificity of 27.5%, which corresponded to MEWS thresholds of ≥ 4 and ≥ 2 , NEWS thresholds of ≥ 5 and ≥ 2 , and eCART thresholds of ≥ 12 and ≥ 4 , respectively. At those thresholds, eCART caught 22 more adverse events per 10,000 patients than BTF using the moderate risk criteria and 13 more using high risk criteria, while MEWS and NEWS identified the same or fewer.

Conclusion(s): An electronically generated eCART score was more accurate than commonly used paper based observation tools for predicting the composite outcome of in-hospital cardiac arrest, ICU transfer and death within 24 hours of observation. The outcomes of this analysis lend weight for a move towards an algorithm based electronic risk identification tool for deteriorating patients to ensure earlier detection and prevent adverse events in the hospital.

Introduction

Failure to recognise and appropriately manage deteriorating patients is a contributing factor in many adverse events in hospitals and health care organisations around the world,^{1–4} which has led to the widespread adoption of *patient safety net systems*.^{5–8} These systems, which have their origins in pioneering work at Liverpool Hospital in New South Wales (NSW), Australia in the early 1990's,⁹ are designed to help clinicians recognise deterioration in their patients and enable them to initiate an appropriate response.¹⁰

Essential to these systems is a set of predetermined criteria as indicators for the need to escalate monitoring or call for help. In recent years, there is an increasing focus on the use of observation charts to assist in the identification of patients who are deteriorating. This can be seen in efforts internationally and within Australia to revise and improve charts, and to incorporate specific features in them (such as early warning scores) to support this identification process.^{11–13}

In 2010, the Clinical Excellence Commission's (CEC) *Between the Flags* (BTF) patient safety net system introduced a standardised colour coded observation chart (incorporating standard calling criteria) in over 200 hospitals across New South Wales (NSW) Australia.¹⁴ Unlike the modified early warning score (MEWS)¹⁵ and the National Early Warning Score (NEWS),¹⁶ the BTF chart is a single parameter track and trigger system.

Whilst there have been several studies examining design, usability and health professional's perception of observation charts^{11, 17–18} there are no studies which have compared the BTF calling criteria with other commonly used track and trigger systems. Therefore, the aim of this study was to compare the accuracy of the BTF calling criteria to the MEWS and NEWS, commonly used risk tools in identifying adult patients on the ward for predicting adverse outcomes. We also aimed to compare BTF to the eCART (electronic Cardiac Arrest Risk Triage) score,¹⁹ which is a previously derived machine learning algorithm that utilizes vital signs and laboratory values to identify high risk patients.

MATERIALS AND METHODS

Study Population and Data collection

All patients hospitalized on the wards at the University of Chicago and four Northshore University Health System hospitals (Evanston, Glenbrook, Highland Park, and Skokie) from November 2008 to August 2013 were eligible for inclusion in the study. All of the participating hospitals had Rapid Response Teams (RRT) in place but did not have specific calling criteria to activate the RRT. Because one goal of this study was to compare a previously developed machine learning algorithm to other early warning scores, admissions included in the derivation phase of the algorithm were excluded from this analysis. Patient demographics and time and location stamped vital signs and laboratory results were obtained from the Electronic Data Warehouse at Northshore and the Electronic Health Record (EHR) at the University of Chicago. The study protocol was approved by the University of Chicago Institutional Review Board and a waiver of consent was granted (IRB #16995A).

Outcomes

The primary outcome was defined as a cardiac arrest, ICU transfer, or death on the ward occurring within 24 hours of an observation. Cardiac arrests were defined as loss of a palpable pulse with attempted resuscitation, and quality checks were performed as previously described. ICU transfers were defined as direct transfers from the wards and were determined using location data from the admission-discharge-transfer feed.

Score calculation

The accuracy of four early warning scores: Between the Flags (BTF), Modified Early Warning Score (MEWS), National Early Warning Score (NEWS), and electronic Cardiac Arrest Risk Triage (eCART), were investigated in this study. MEWS¹⁵ and NEWS¹⁶ are commonly used vital sign based aggregated weighted scores and eCART score is a machine learning random forest algorithm that includes laboratory values and patient demographics, in addition to vital signs, and has been previously shown to outperform MEWS.¹⁹ The laboratory values in eCART are white cell count, haemoglobin, platelets, sodium, potassium, chloride, bicarbonate, anion gap, blood urea nitrogen, creatinine, glucose, calcium, total protein, albumin, total bilirubin, aspartate aminotransferase, and alkaline phosphatase. The vital sign observation frequency is every four hours for all the hospitals participating in the study. Data at the time of observation were used to calculate each score. Only scores calculated on the wards were included in the analyses.

To be considered high risk by BTF criteria (ie Red Zone), a patient must have met any of the following criteria: a respiratory rate ≥ 30 or ≤ 5 ; oxygen saturation $< 90\%$; systolic blood pressure ≥ 200 or ≤ 90 ; heart rate ≥ 140 or ≤ 40 ; or an unresponsive or responsive only to pain mental status. To be considered moderate risk by BTF criteria (ie Yellow Zone), a patient must have met any of the following criteria: a respiratory rate ≥ 25 or ≤ 10 ; oxygen saturation $< 95\%$; systolic blood pressure ≥ 180 or ≤ 100 ; heart rate ≥ 120 or ≤ 50 ; less than alert mental status; or temperature ≥ 38.5 or ≤ 35.5 degrees Celsius. For comparison purposes, the specificity of both the yellow and red zone BTF criteria were matched with the

closest specificity for MEWS, NEWS, and eCART scores, resulting in the identification of similar “yellow” and “red” zones for all scores.

Calculations of accuracy, sensitivity, and specificity were assessed using the calculated vital sign score at each time point included in the analysis. Calculations of estimated patient saves were assessed using the maximum vital sign score within the 24 hours preceding an outcome. Non-physiologic values were converted to missing, as previously described.¹⁹ At each time with a new vital sign or a laboratory result, prior values were pulled forward for the other variables not collected simultaneously, as needed. If no prior value was available, then a median value was imputed.

Statistical analysis

Patient characteristics were conducted at the patient level and were compared between those who experienced an outcome and those who did not using t-tests, Wilcoxon rank sum tests, and chi-squared tests as appropriate, depending on the distribution of the data. Accuracy comparisons were performed using sensitivity, specificity, and false positive rates. The area under the receiver operating curve (AUROC) was used to evaluate score discrimination with vital sign observations treated as if they are independent as per previous studies.²⁰ A two-tailed p-value of less than 0.05 was considered statistically significant. All analyses were performed using Stata version 14.1 (Stata Corps; College Station, TX).

RESULTS

Study population

A total of 107,868 patient admissions occurred during the study period. The included population experienced 160 cardiac arrests, 938 deaths, and 5,044 ICU transfers during the study period, with 5.1% (n=5,485) of patients experiencing any adverse outcome. Compared to patients who did not experience an outcome, those who did were older (71.3 vs. 61.4 years, p<0.001), more likely to be male (52% vs. 40%, p<0.001) and more likely to be white (65% vs. 61%, p<0.001). Patients experiencing an adverse event were also more likely to meet moderate and high risk criteria of all four tools (Table 1). However, patients with adverse events were most likely to meet high and moderate criteria with eCART (98% and 71%, respectively), while those not experiencing an event were least likely to meet MEWS criteria (73% and 15%, respectively).

Accuracy comparisons

Using the calculated scores at the time of each observation, the moderate risk threshold for BTF had a sensitivity of 64.9% and a specificity of 59.8% for the combined outcome occurring within the next 24 hours. At a similar specificity, NEWS₂ had a sensitivity of 69.8% and eCART₄ had the highest sensitivity of 81.4%. The high risk threshold for BTF had a sensitivity of 27.5% and a specificity of 95.0%, while at a similar specificity, NEWS₅ had a sensitivity of 30.9% and eCART₁₂ had a sensitivity of 36.9% (Table 2).

When compared to BTF, both MEWS and NEWS high risk criteria identified fewer patients with a subsequent adverse event, resulting in 55 fewer catches per 10,000 patients for

MEWS and 13 fewer for NEWS. For the moderate risk criteria, MEWS identified 14 fewer adverse outcomes per 10,000 patients than BTF, while eCART caught 22 more adverse events per 10,000 patients than BTF using the moderate risk criteria and 13 more using high risk criteria (Figure 1). All three comparator tools resulted in fewer false positives than BTF, with MEWS identifying the least number of false positives per 10,000 patients (Figure 2).

Overall accuracy was highest for eCART, with an AUC of 0.801 (95% CI 0.799–0.802), followed by NEWS with an AUC of 0.718 (0.716–0.720), MEWS with an AUC of 0.698 (0.696–0.700), and then BTF, with an AUC of 0.663 (0.661–0.664) (Figure 3).

Time to Outcome

The moderate risk criteria for BTF, MEWS, NEWS, and eCART were all first met at a median of at least 40 hours prior to an adverse event, while the high risk criteria were all first met at a median of at least 20 hours prior to the event. (Table 3).

We found similar trends to the combined outcome results when the AUC was calculated for each outcome separately, as well as for the patient characteristics and time to outcome analyses. (see supplementary materials).

Discussion

The challenge with setting vital sign thresholds is that there is no science that can inform precisely where an escalation threshold should be drawn. There is evidence regarding the relationship between risk of death and adverse events for vital sign values^{21, 22} but this evidence is not sufficient to inform when escalation should occur, and, therefore, where escalation thresholds should be drawn on an observation chart. Our study demonstrated that commonly used aggregated early warning scores such as NEWS and MEWS are more accurate than a single trigger system for predicting in-hospital cardiac arrest, mortality and ICU transfer in adult ward patients within 24 hours. We found that the eCART risk stratification tool is the most accurate in predicting adverse outcomes with an AUROC of 0.80.

Not only is there no science that can give precision to where these thresholds are drawn, but the logic of the track and trigger system for recognising deterioration is that the pattern of all vital signs is considered together and not individually; so, individual precision is even less important. Ultimately, the degree of acceptance among clinicians of these thresholds depends on gaining their understanding of the principle that precision is not warranted by any available evidence and is much less important than the fact that there is a set of thresholds which are based on the judgement of experienced clinicians and which collectively provide a patient safety net.^{23, 24}

Between the Flags is a multi-valent patient safety net system developed for over 200 hospitals by the Clinical Excellence Commission in New South Wales, Australia to improve the early recognition and management of patients who are clinically deteriorating.^{14, 25} One element of this system was the introduction of the standard adult general observation (SAGO) chart with standard calling criteria, using a ‘track and trigger’ design, and including

two coloured zones (yellow and red) representing early and late warning signs for deterioration respectively, which help clinicians recognise when the threshold has been reached to trigger a system response. Any breach of a threshold (single-trigger) requires an appropriate prescribed action (which is defined on the observation chart). Use of the SAGO charts was mandated in the state-wide policy.

We found that the Yellow Zone calling criteria in the BTF observation charts had a lower sensitivity and specificity than the other tools in predicting the combined outcome of in-hospital cardiac arrest, mortality and transfer to ICU within 24 hours. A key decision for BTF was whether to require an earlier clinical response to early warning signs. The evidence from the SOCCER Study²⁶ had demonstrated that even moderate deviation of vital sign observations from normal was predictive of adverse events. These were called ‘early warning signs’ by the SOCCER study authors. This provided the rationale for the introduction of earlier thresholds and these were incorporated in the Standard Observation Charts. This required the introduction of a new system for clinical response to these patients called the Clinical Review. Activation of this response is discretionary using clinical judgement, and is intended to be provided by the patient’s medical team.

The Red Zone thresholds are based on the pre-existing MET criteria²² which when breached mandates a call to the Rapid Response team. Comparing the different tools across 95% specificity levels found that the BTF Red Zone thresholds were more sensitive than MEWS but less sensitive than NEWS and eCART. Breaching these thresholds requires a mandatory Rapid Response call in hospitals across NSW and therefore there is a trade-off between need and resources as the number of calls has risen over the program’s implementation.^{10, 25}

A recently published independent evaluation of the impact of implementing BTF has shown significant improvements in patient outcomes including a 46% reduction in in-hospital cardiac arrest rates; a 54% reduction in cardiac arrest related mortality rates; a 19% reduction in hospital mortality; and a 35% decrease in failure to rescue rates over seven-years. In addition, there was a new 20% mortality reduction among low mortality diagnostic related group patients.²⁷ Our study showed that 94% of patients who experienced an adverse event met the moderate risk (Yellow Zone) criteria (Table 1) and this finding supports the decision to introduce the BTF Yellow Zone thresholds as earlier recognition results in better outcomes.

One of the challenges with any deteriorating patient system relates to the expected increase in Rapid Response calls and the time and human resources needed to operationalize the system. Our study found that the BTF thresholds identified more patients with an outcome, resulting in more lives saved compared to MEWS and NEWS per 10,000 patients. However, when compared to BTF, MEWS and NEWS identified less false positives per 10,000 patients. Compared to the single parameter and aggregated scoring systems the eCART tool identified more patients with an outcome resulting in more lives saved per 10,000 patients. This suggests that the use of eCART could improve the stratification of patients enabling more immediate action for those known to be at higher risk to prevent further deterioration and the reduction of ‘unnecessary’ Rapid Responses calls as compared to BTF.

The combination of multiple laboratory tests and vital sign observations to develop predictive models has been investigated in a number of previous studies dealing with ward and emergency department patients.^{28–30} We believe that moving towards the electronic medical record will overcome previously described issues with aggregated scoring systems³¹ and allow an electronic algorithm to increase the positive predictive value of the afferent arm and assist in the risk stratification of patients so that resources can be used appropriately.

Finally, our study found that the least accurate tool identified patients as early as 25 hours and the most accurate up to 45 hours prior to an adverse event. Whilst it is logical to identify at risk patients as early as possible because most have antecedents before adverse events,^{32, 33} and early intervention is associated with decreased mortality and morbidity³⁴ the ideal time to alert clinicians before an adverse event is not known.

In some jurisdictions, there are examples of efficient afferent arms whilst others have highly developed efferent arms – very few have mastered both as part of a system of care. The afferent arm of a RRS could be improved by moving from paper based observation charts to using an automatic electronic algorithm combining patient characteristics, observations and laboratory parameters. Ultimately, we need to improve both the afferent and efferent arms to maximise the effectiveness of the system and outcomes for patients. The potential generation of an optimal algorithm based electronic risk identification tool for deteriorating patients will assist health systems as they move from paper based tools to the electronic health record.

This study has several strengths. It involved more than 107,000 patients and more than 4,000,000 single observations. It is the first study to compare the BTF calling criteria with MEWS, NEWS and eCART risk tools. However, this was an investigation which compared international calling criteria in five U.S. hospitals and the results may not be generalizable to other settings. In addition, there is no gold standard to determine when current clinical parameters indicate the beginning of deterioration. In this retrospective study we only compared objective criteria omitting the subjective criteria ‘worried’ which is an important part of the BTF calling criteria and serves as an additional RRT activation criteria for all warning scores. Further comparative studies are required to determine the impact of subjective and objective criteria in reducing adverse patient outcomes. There are many other things to consider when implementing calling criteria, such as the ability to see trends of vital signs, to change or suspend a calling criterion, and the simplicity and transparency of the tool used. We do not have information on ‘do not resuscitate’ orders for this dataset. This is important because previous studies have shown limitations of medical therapy arise in up to 30% of Rapid Response calls, that Rapid Response Teams (RRT) issue more DNR orders than conventional cardiac arrest teams and RRTs can improve end of life care as well.^{35–37}

Whilst we have only examined one single parameter track and trigger system (BTF) and two aggregated scoring systems (MEWS and NEWS), there are over 100 published early warning scores in the literature.³⁸ Given that BTF, MEWS and NEWS are commonly used tools we believe these results would be of value. Lastly, optimising the afferent arm of the

deteriorating patient system will not guarantee better clinical outcomes unless the efferent arm, and the rest of the system, is effective.

Conclusions

We conducted a study with more than 4 million vital sign observations from more than 107,000 patients and found that the eCART tool was more accurate than commonly used paper based observation tools for predicting the composite outcome of in-hospital cardiac arrest, ICU transfer and death within 24 hours of observation. This is an important implication for health systems worldwide, for as they move towards the introduction of electronic health records the use of statistically derived algorithms combining patient demographics, vital signs and laboratory values becomes easier and therefore could improve the early identification of patients at risk of cardiac arrest, ICU admission and death. Future clinical trials will need to assess the comparative ability of these tools to improve patient outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data from this study were provided by the Clinical Research Data Warehouse (CRDW) maintained by the Center for Research Informatics (CRI) at University of Chicago. The Center for Research Informatics is funded by the Biological Sciences Division, the Institute for Translational Medicine/CTSA (NIH UL1 TR000430) at the University of Chicago.

Dr. Churpek is supported by a career development award from the National Heart, Lung, and Blood Institute (K08 HL121080) and has received honoraria from Chest for invited speaking engagements. Drs. Churpek and Edelson have a patent pending (ARCD. P0535US.P2) for risk stratification algorithms for hospitalized patients. In addition, Dr. Edelson has received research support and honoraria from Philips Healthcare (Andover, MA), and Early Sense (Tel Aviv, Israel). She has ownership interest in Quant HC (Chicago, IL), which is developing products for risk stratification of hospitalized patients.

References

1. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalised patients: results of the Harvard Medical Practice Study I. *N Engl J Med* 1991; 324: 370–77. [PubMed: 1987460]
2. Wilson RM, Runciman WB, Gibberd RW, et al. The quality in Australian Health Care Study. *Med J Aust* 1995; 163: 458–76. [PubMed: 7476634]
3. Davis P, Lay-Yee R, Briant R, et al. Adverse events in New Zealand public hospitals I: occurrence and impact *N Z Med J* 2002; 115: 271.
4. Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ* 2004; 170: 1678–86. [PubMed: 15159366]
5. Ball C, Kirby K, Williams S. Effect of the critical care outreach team on patient survival to discharge from hospital and readmission to critical care: non-randomised population based study. *BMJ*. 2003 11 1; 327(7422): 1014. [PubMed: 14593033]
6. England K, Bion JF. Introduction of medical emergency teams in Australia and New Zealand: a multicentre study. *Crit Care* 2008; 12:151. [PubMed: 18505599]
7. Steel AC, Reynolds SF. The growth of rapid response systems. *Jt Comm J Qual Patient Saf* 2008; 34:489–95. [PubMed: 18714752]

8. Winters BD, Pham J, Pronovost PJ. Rapid response teams — walk, don't run. *JAMA* 2006; 296:1645–7. [PubMed: 17018807]
9. Lee A, Bishop G, Hillman KM, Daffurn K. The medical emergency team. *Anaesth Intensive Care* 1995; 23: 183–186. [PubMed: 7793590]
10. DeVita MA, Bellomo R, Hillman K, et al. Findings of the first consensus conference on medical emergency teams. *Crit Care Med* 2006;34:2463–78. [PubMed: 16878033]
11. Chatterjee MT, Moon JC, Murphy R, McCrea D. The “OBS” chart: an evidence based approach to re-design of the patient observation chart in a district general hospital setting. *Postgraduate Medical Journal*. 2005;81:663–6. [PubMed: 16210466]
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*. 2003;58:775803.
13. Van Leuvan C, Mitchell I. Missed opportunities? An observational study of vital sign measurements. *Critical Care and Resuscitation*. 2008;10(2):111–5. 12 [PubMed: 18522524]
14. Hughes C, Pain C, Braithwaite J, & Hillman K (2014). ‘Between the Flags’: implementing a rapid response system at scale. *BMJ quality & safety*, 23(9), 714–717.
15. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. *Qjm*. 2001 10 1;94(10):521–6. [PubMed: 11588210]
16. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation*. 2013 4 30;84(4):465–70. [PubMed: 23295778]
17. Preece MH, Hill A, Horswill MS, Watson MO. Supporting the detection of patient deterioration: observation chart design affects the recognition of abnormal vital signs. *Resuscitation*. 2012 9 30;83(9):1111–8. [PubMed: 22353643]
18. Elliott D, Allen E, Perry L, Fry M, Duffield C, Gallagher R, Iedema R, McKinley S, Roche M. Clinical user experiences of observation and response charts: focus group findings of using a new format chart incorporating a track and trigger system. *BMJ quality & safety*. 2015 1 1;24(1):65–75.
19. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, Edelson DP. Multicenter development and validation of a risk stratification tool for ward patients. *American journal of respiratory and critical care medicine*. 2014 9 15;190(6):649–55. [PubMed: 25089847]
20. Jarvis SW, Kovacs C, Briggs J, Meredith P, Schmidt PE, Featherstone PI, Prytherch DR, Smith GB. Are observation selection methods important when comparing early warning score performance?. *Resuscitation*. 2015 5 31;90:1–6. [PubMed: 25668311]
21. Kause J, Smith G, Prytherch D, Parr M, Flabouris A, Hillman K. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom—the ACADEMIA study. *Resuscitation*. 2004 9 30;62(3):275–82. [PubMed: 15325446]
22. Cretikos M, Chen J, Hillman K, Bellomo R, Finfer S, Flabouris A, Merit Study Investigators. The objective medical emergency team activation criteria: a case-control study. *Resuscitation*. 2007 4 30;73(1):62–72. [PubMed: 17241732]
23. Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, Esmonde L, Goldhill DR, Parry GJ, Rashidian A, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med* 2007;33:667–679. [PubMed: 17318499]
24. Cuthbertson BH, Smith GB. A warning on early-warning scores! *Br J Anaesth* 2007;98:704–706. [PubMed: 17519260]
25. Pain C, Green M, Duff C et al. Between the Flags: Implementing A Safety Net System At Scale To Recognise And Manage Deteriorating Patients In The New South Wales Public Health System. *International Journal of Quality and Safety*. 2016.
26. Jacques T, Harrison GA, McLaws M, Kilborn G Signs of critical conditions and emergency responses (SOCCER): a model for predicting adverse events in the inpatient setting. *Resuscitation*. 2006; 69(2): 175–83 [PubMed: 16497427]

27. Chen J, Ou L, Flabouris A, Hillman K, Bellomo R, Parr M. Impact of a standardized rapid response system on outcomes in a large healthcare jurisdiction. *Resuscitation*. 2016 10 31;107:47–56. [PubMed: 27507434]
28. Loekito E, Bailey J, Bellomo R, Hart GK, Hegarty C, Davey P, Bain C, Pilcher D, Schneider H. Common laboratory tests predict imminent death in ward patients. *Resuscitation*. 2013 3 31;84(3):280–5. [PubMed: 22863543]
29. Loekito E, Bailey J, Bellomo R, Hart GK, Hegarty C, Davey P, Bain C, Pilcher D, Schneider H. Common laboratory tests predict imminent medical emergency team calls, intensive care unit admission or death in emergency department patients. *Emergency Medicine Australasia*. 2013 4 1;25(2):132–9. [PubMed: 23560963]
30. Khurana HS, Groves RH, Simons MP, Martin M, Stoffer B, Kou S, Gerkin R, Reiman E, Parthasarathy S. Real-Time Automated Sampling of Electronic Medical Records Predicts Hospital Mortality. *The American journal of medicine*. 2016 3 24.
31. Prytherch DR, Smith GB, Schmidt P, Featherstone PI, Stewart K, Knight D, Higgins B. Calculating early warning scores—a classroom comparison of pen and paper and hand-held computer methods. *Resuscitation*. 2006 8 31;70(2):173–8. [PubMed: 16806641]
32. Schein RM, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest*. 1990 12 31;98(6):1388–92. [PubMed: 2245680]
33. Buist MD, Jarmolowski E, Burton PR, Bernard SA, Waxman BP, Anderson J. Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care. A pilot study in a tertiary-care hospital. *The Medical Journal of Australia*. 1999 7;171(1):22–5. [PubMed: 10451667]
34. Chen J, Bellomo R, Flabouris A, Hillman K, Finfer S, MERIT Study Investigators for the Simpson Centre, ANZICS Clinical Trials Group. The relationship between early emergency team calls and serious adverse events. *Critical care medicine*. 2009 1 1;37(1):148–53. [PubMed: 19050625]
35. Jones DA, Bagshaw SM, Barrett J, et al. The role of the medical emergency team in end-of-life care: a multicenter, prospective, observational study. *Crit Care Med* 2012; 40: 98–103. [PubMed: 21926596]
36. Chen J, Flabouris A, Bellomo R, Hillman K, Finfer S. The medical emergency team system and not-for-resuscitation orders: results from the MERIT study. *Resuscitation* 2008; 79:391–7. [PubMed: 18952354]
37. Vazquez R, Gheorghe C, Grigoriyan A, Palvinskaya T, Amoateng-Adjepong Y and Manthous CA (2009), Enhanced end-of-life care associated with deploying a rapid response team: A pilot study. *J. Hosp. Med.*, 4: 449–452. [PubMed: 19753581]
38. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalized patients on the wards. *CHEST Journal*. 2013 6 1;143(6):1758–65.

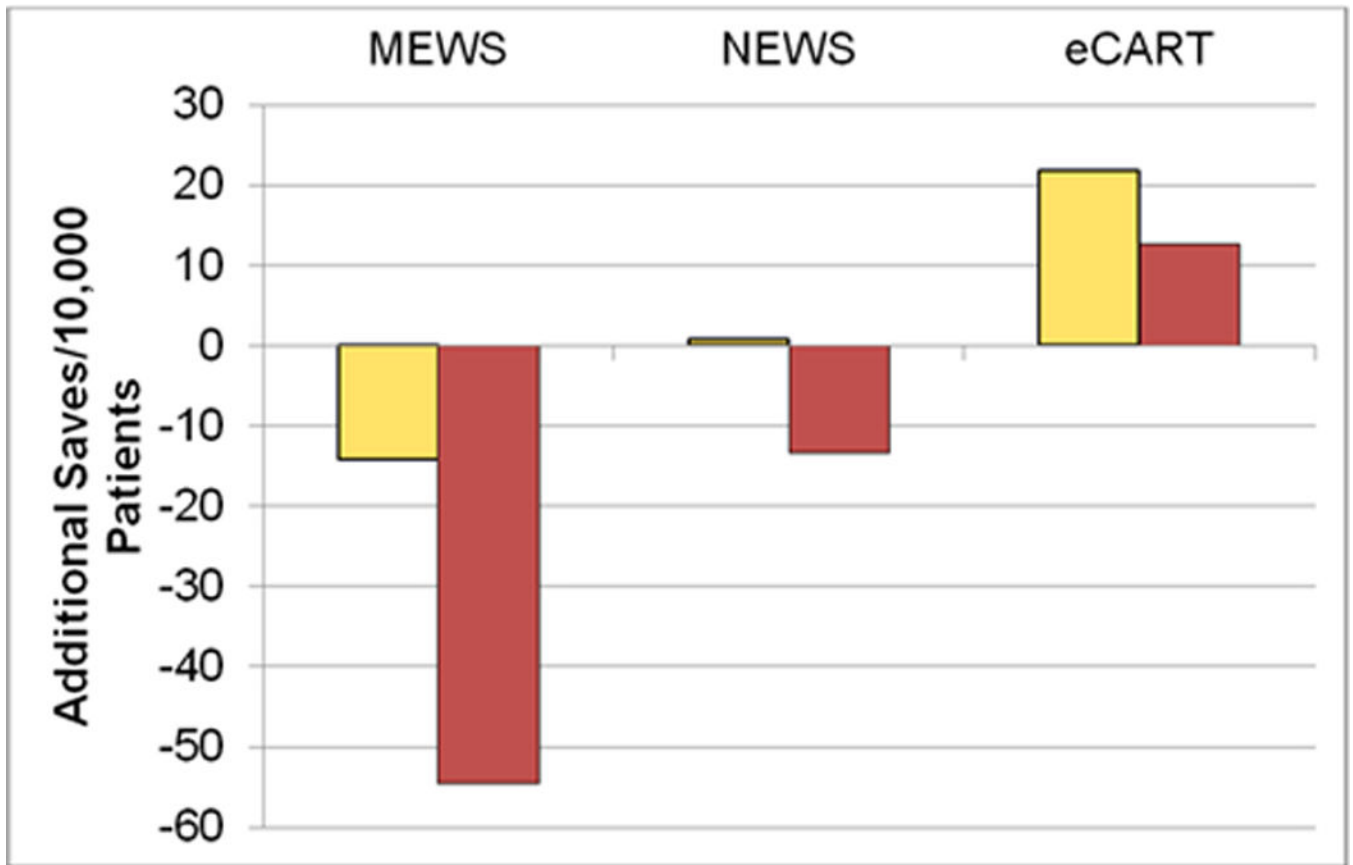


Fig. 1.
Additional Saves Compared to BTF.

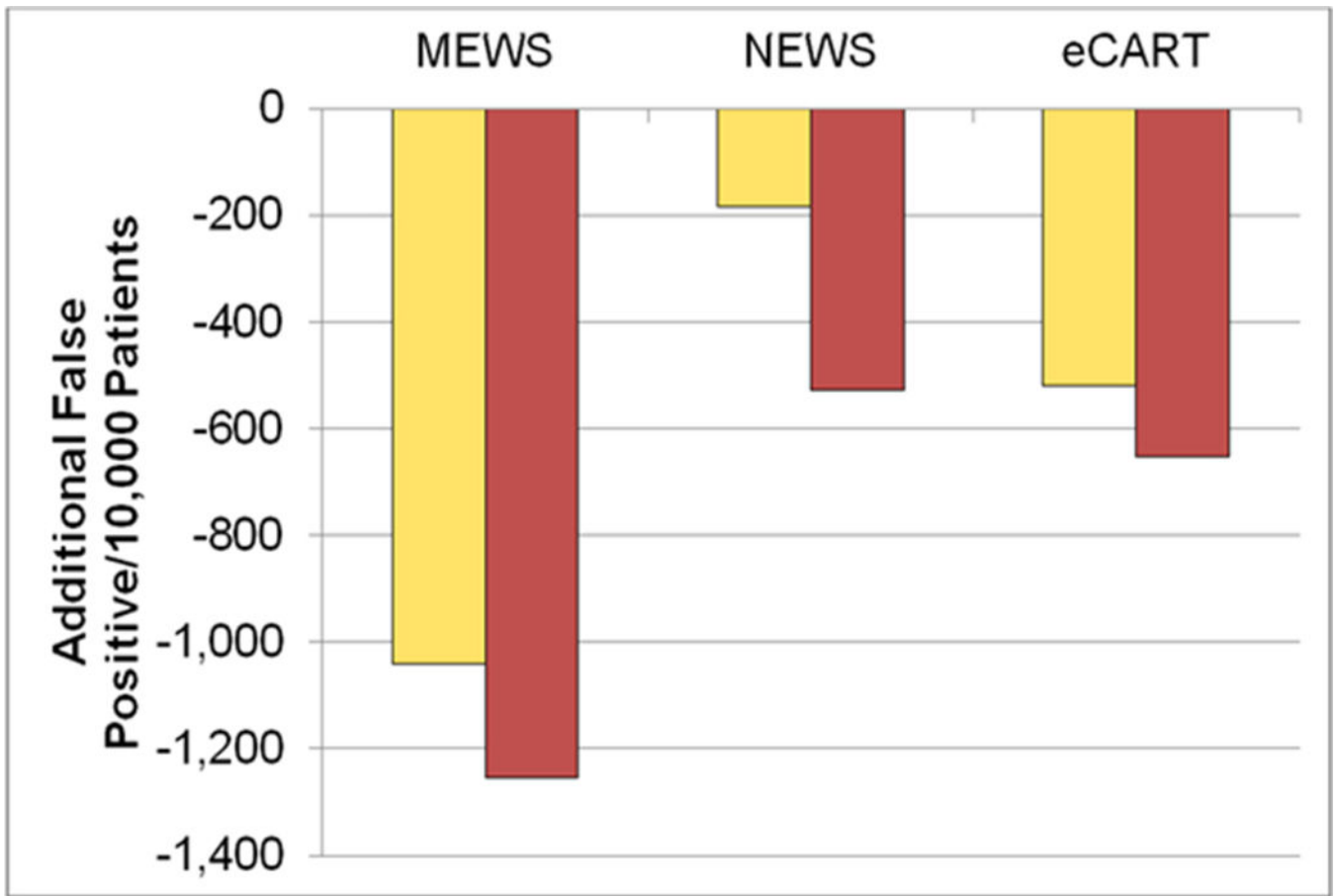


Fig. 2.
Additional False Positives Compared to BTF.

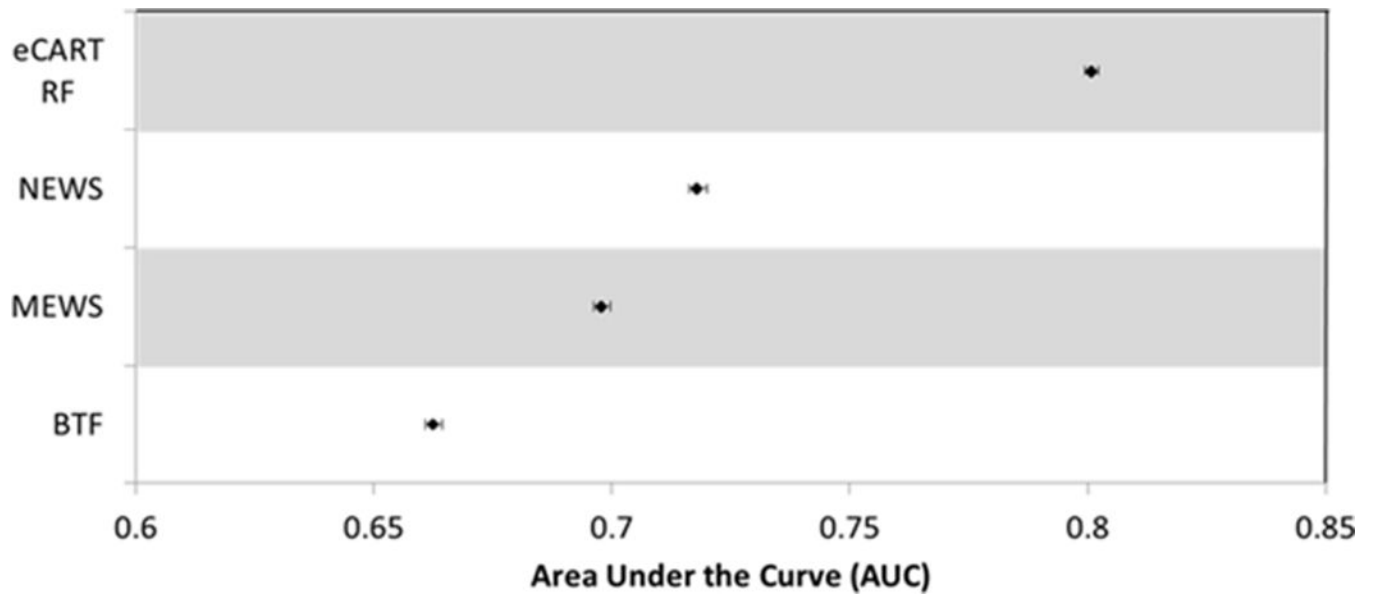


Fig. 3.
Accuracy Comparisons.

Table 1.

Patient Characteristics

Patient Characteristics	No Outcome (n=102,383)	Outcome (n=5,485)	p-value
Age in years, median (IQR) *	61.4 (43.9–77)	71.3 (58–82.7)	p<0.001
Male, n (%)	40,717 (40)	2,848 (52)	p<0.001
Race, n (%)			p<0.001
Black/African-American	20,684 (21)	1,116 (21)	
White	60,729 (61)	3,464 (65)	
Other/unknown	18,509 (19)	788 (15)	
Proportion of patients who met criteria at any time during ward segment	No Outcome (n=102,383)	Outcome (n=5,485)	p-value
BTF, n (%)			
Moderate risk (Yellow Zone)	86,341 (84)	5,162 (94)	p<0.001
High risk (Red Zone)	28,479 (28)	3,771 (69)	p<0.001
MEWS, n (%)			
Moderate risk (2)	75,125 (73)	5,009 (91)	p<0.001
High risk (4)	14,960 (15)	3,183 (58)	p<0.001
NEWS, n (%)			
Moderate risk (2)	84,349 (82)	5,171 (94)	p<0.001
High risk (5)	22,790 (22)	3,627 (66)	p<0.001
eCART, n (%)			
Moderate risk (4)	80,736 (79)	5,398 (98)	p<0.001
High risk (12)	21,442 (21)	3,908 (71)	p<0.001

* n=43 patients excluded from the analysis due to missing age

Table 2.

Sensitivity and Specificity of Combined Outcome within 24 hours

Model	Cutoff	Sensitivity	Specificity
BTF	Yellow Zone	64.9%	59.8%
	Red Zone	27.5%	95.0%
MEWS	1	97.3%	1.3%
	2	62.5%	71.8%
	3	40.2%	90.6%
	4	24.9%	96.9%
	5	14.0%	99.0%
NEWS	1	85.1%	32.3%
	2	69.8%	61.0%
	3	56.6%	79.0%
	4	42.7%	89.7%
	5*	42.1%	90.0%
	5	30.9%	95.1%
	6	21.9%	97.7%
eCART	4	81.4%	59.9%
	6	64.9%	78.1%
	9	50.0%	89.9%
	12	36.9%	95.0%
	15	27.5%	97.3%
	21	16.0%	99.0%

Table 3.

Time to Outcome

Score, median (IQR)	Time before outcome (hours)	
	Yellow Zone	Red Zone
BTF	42.2 (14.9–105.3)	25.2 (6.0–82.5)
MEWS	40.6 (13.2–103.8)	21.0 (5.0–75.7)
NEWS	42.7 (14.7–106.7)	27.4 (6.9–85.8)
eCART	44.9 (16.5–108.5)	29.1 (7.6–88.4)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript