Validating the Electronic Cardiac Arrest Risk Triage (eCART) Score for Risk Stratification of Surgical Inpatients in the Postoperative Setting:

Retrospective Cohort Study

Bartlomiej Bartkowiak, PhD[1], **Ashley M. Snyder, MPH**[1], **Andrew Benjamin, MD**[2], **Andrew Schneider, MD**[2], **Nicole M. Twu, MS**[1], **Matthew M. Churpek, MD, MPH, PhD**[1], **Kevin K. Roggin, MD**[2,*], and **Dana P. Edelson, MD, MS**[1,*]

[1]Department of Medicine, University of Chicago Medicine, Chicago, IL

[2]Department of Surgery, University of Chicago Medicine, Chicago, IL

Abstract

Objective—Assess the accuracy of three early warning scores for predicting severe adverse events in postoperative inpatients.

Summary Background Data—Postoperative clinical deterioration on inpatient hospital services is associated with increased morbidity, mortality, and cost. Early warning scores have been developed to detect inpatient clinical deterioration and trigger rapid response activation, but knowledge regarding the application of early warning scores to postoperative inpatients is limited.

Methods—This was a retrospective cohort study of adult patients hospitalized on the wards following surgical procedures at an urban academic medical center from 11/2008 to 1/2016. The accuracies of the Modified Early Warning Score (MEWS), National Early Warning Score (NEWS), and the electronic Cardiac Arrest Risk Triage (eCART) score were compared in predicting severe adverse events (ICU transfer, ward cardiac arrest, or ward death) in the postoperative period using the area under the receiver operating characteristic curve (AUC).

Results—Of the 32,537 patient admissions included in the study, 3.8% (n=1,243) experienced a severe adverse outcome following the procedure. The accuracy for predicting the composite outcome was highest for eCART (AUC 0.79 [95% CI: 0.78–0.81]), followed by NEWS (AUC 0.76 [95% CI: 0.75–0.78]), and MEWS (AUC 0.75 [95% CI: 0.73–0.76]). Of the individual vital signs and labs, maximum respiratory rate was the most predictive (AUC 0.67) and maximum temperature was an inverse predictor (AUC 0.46).

Corresponding Author: Kevin K. Roggin; The University of Chicago Medicine, 5841 S. Maryland Avenue, MC 4052, Chicago, IL 60637. Phone: (773) 834-3524. Fax: (773) 702-6120. kroggin@surgery.bsd.uchicago.edu.
*Both senior authors contributed equally to this work

Conclusions—Early warning scores are predictive of severe adverse events in postoperative patients. eCART is significantly more accurate in this patient population than both NEWS and MEWS.

MINI-ABSTRACT

Early warning scores are useful in the detection of clinical deterioration, but knowledge regarding their application to postoperative inpatients is limited. Comparison of three early warning scores finds them to be predictive of severe adverse events in postoperative patients and reveals eCART to be more accurate than NEWS and MEWS.

INTRODUCTION

There are several validated perioperative risk assessment tools which are widely used in surgical decision-making and risk modification, including the American Society of Anesthesiologists (ASA) Physical Status classification system, the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) surgical risk calculator, and the surgical Apgar score (SAS) (1–3). Although these tools have demonstrated value in the perioperative setting, they are either designed for preoperative (ASA and ACS-NSQIP) risk assessment or based entirely on intraoperative data (SAS). Consequently they do not account for postoperative physiology. Therefore, there are many circumstances where their utility for guiding postoperative care on the medical-surgical wards in the hospital is limited (1,4).

Early warning scores, on the other hand, such as the Modified Early Warning Score (MEWS) (5) and National Early Warning Score (NEWS) (6), are increasingly being used to dynamically risk stratify general ward patients (7). Recent studies in patients admitted to surgical services have found them to have good predictive accuracy for adverse events, including cardiac arrest, unplanned ICU transfer, and death (8–13). More recently, electronic databases and advanced statistics have enabled increasingly complex early warning scores, such as the electronic Cardiac Arrest Risk Triage (eCART) score (14–17), which utilizes 33 time-varying parameters including both vital signs and laboratory data and has been shown to be more accurate than MEWS in the general inpatient population (16). However, the predictive ability of eCART has not been investigated specifically in surgical patients. We therefore sought to determine the ability of eCART to predict severe adverse events in the postoperative period and to compare its accuracy to that of MEWS and NEWS.

METHODS

Study Population and Data Collection

We conducted a retrospective cohort study of all adult patients admitted to inpatient surgical service following an operative procedure at the University of Chicago from November 2008 to January 2016. Patients with operating room (OR) procedural events were identified via EPIC OpTime. Postoperative ward stay was defined by the presence of at least one postoperative ward vital sign following the first OR visit. Surgical procedures were defined through a reproducible, text-based, selection criteria applied to OR procedure descriptions obtained from EPIC OpTime. The selection criteria were reviewed by three surgeons [K.R.,

A.B., and A.S.] and were designed to remove minor endoscopic and interventional radiology procedures (e.g. endoscopic procedures, uteroscopies, pacemaker insertions, and interventional radiology procedures) (Appendix 1).

Patient demographic information was obtained from administrative databases, and time- and location-stamped vital sign and laboratory results were obtained from electronic health record data (EPIC, Verona, WI), as described previously (14–16). Cardiac arrest was identified using a previously described quality improvement database (14). The study protocol was designated as "not human subject's research" by the University of Chicago Institutional Review Board (IRB #15-0195).

Analysis

The primary outcome was a composite of ward cardiac arrest, ward to ICU transfer, or ward death during the postoperative period. Only the first postoperative outcome of each encounter was included in the analysis. To ensure that the outcome was related to a single procedure, any observations following a subsequent return to the operating room were censored from the analysis.

The early warning score algorithms used were the previously validated MEWS (5), NEWS (6), and eCART (15,16) scores. Scores were calculated each time a new predictor variable measurement became available. For time points that were missing predictor values, previous values were carried forward. If no previous values were available, the median value for that variable was imputed (16).

Area under the receiver operating characteristic curves (AUC) were calculated for the combined outcome for each early warning score using the maximum ward score value during the first postoperative period for each encounter. Early warning scores were not calculated 30 minutes prior to an outcome occurring due to the limitations in the accuracy of the location data within this timeframe.

Patient characteristics of the outcome and no outcome groups were compared using t-tests, Wilcoxon rank sum tests, and Chi-squared tests as appropriate. All analyses were performed using Stata version 14.1 (StataCorp, College Station, TX), with a p value less than 0.05 denoting statistical significance.

RESULTS

A total of 39,009 patients experienced an OR stay and a subsequent admission to the general ward. Of those, 6,472 did not undergo a previously specified surgical procedure and were excluded from the analysis, resulting in a final population of 32,537 patients. The final population contained 1,243 (3.8%) patients who experienced a clinical deterioration event during the postoperative period, consisting of 1,189 unplanned ICU transfers, 29 ward cardiac arrests, and 25 ward deaths (Appendix 2).

Compared to the patients that did not experience an adverse outcome, patients who experienced an adverse outcome were more likely to have been admitted to the hospital prior to the operation (64% vs 32%, p<0.001) and have longer preoperative stays (median 3.0

[IQR, 1.2–6.8] vs 2.0 [IQR, 0.9–4.7] days, p<0.001) (Table 1). Patients who experienced an outcome were also more likely to spend their initial postoperative period in the ICU, prior to transfer to the wards (31% vs 20%, p<0.001). Additionally, patients who experienced an outcome were older on average (mean age 60 years [SD=16] vs 54 years [SD=17], p<0.001).

Evaluation of the predictive accuracy of select vital signs and labs revealed that several single variables, such as respiratory rate (AUC 0.67 [95% CI: 0.65–0.69]), heart rate (AUC 0.66 [95% CI: 0.64–0.68]), and blood urea nitrogen (BUN) (0.65 [95% CI: 0.63–0.67]) had moderate accuracy (Figure 1). Oppositely, the maximum postoperative temperature was found to be an inverse predictor (AUC 0.46 [95% CI: 0.44–0.48]). Although some individual parameters had moderate accuracy, the early warning scores were more accurate than any individual parameter. Accuracy for predicting the composite outcome was highest for eCART (AUC 0.79 [95% CI: 0.78–0.81]), followed by NEWS (AUC 0.76 [95% CI: 0.75–0.78]), and MEWS (AUC 0.75 [95% CI: 0.73–0.76]) (p<0.001 for all comparisons). A NEWS cut-off of 7 resulted in a sensitivity of 75% and a specificity of 64% (Table 2); at the same sensitivity, eCART (outcome probability cut-off 2.1%) had a specificity of 72%. Therefore, in comparison to NEWS, eCART (at a sensitivity of 75%) would have resulted in 2,504 fewer patient calls over the study period (9,694 vs 12,198 calls) while detecting the same number of outcomes. Comparison of eCART to MEWS at a sensitivity of 63% (specificities of 84% for eCART [cut-off 3.3%] and 78% for MEWS [cut-off 4]) results in 1,878 fewer patient calls over the study period.

Patients who experienced an outcome did so at a median of 48 [IQR: 6–135] hours post-operatively. They had a higher median eCART probability of having an adverse event immediately following surgery (1.2% [IQR: 0.6–2.3%] vs 0.7% [IQR: 0.4–1.2%], p<0.001) and reached higher maximum probabilities during their postoperative ward stays (median 5.5% [IQR: 2.0–14.4%] vs 1.2% [IQR: 0.7–2.2%], P<0.001; see table 2 for NEWS and MEWS values) (Table 3 and Figure 2). Further, the risk of an adverse event increased in the 24 hours leading up to the outcome, in contrast to the decreasing probability over time seen in randomly selected 24-hour time periods from patients who did not experience an outcome (Figure 3). Although no predetermined alarm threshold was set a priori; at an alarm threshold of eCART probability of 2.1% (which has the same sensitivity as the NEWS alarm threshold of 7, commonly used in the general inpatient setting), the threshold is crossed at a median of 24.4 hours [IQR: 4–78 hours] prior to event occurrence.

DISCUSSION

In this study of over thirty thousand post-operative inpatients, we demonstrated that the risk of severe adverse events, is dynamic and can be predicted with general early warning scores. Further, we found that eCART was more accurate than either MEWS or NEWS for determining postoperative risk, allowing for higher detection rates with fewer false alarms. This is consistent with prior studies across the population at large in which eCART has been shown to outperform contemporary risk stratification tools (15).

The accuracy of the eCART algorithm is likely attributable to the increase in variables (33 vs 5 for MEWS and 7 for NEWS) and the increased complexity with which they are modeled

(cubic spline logistic regression vs expert opinion or linear logistic regression) (14–16). Subsequent versions of eCART rely on machine learning analytics which have even higher accuracy but are more difficult to implement in real time (17). Although not directly assessed in this study, eCART's accuracy for predicting the first postoperative ward cardiac arrest, unplanned ICU transfer, or ward death in surgical patients is similar to its reported accuracy in predicting the same composite outcome in general hospital inpatients (AUC of 0.79 vs 0.77 (16)). This is in agreement with a recent analysis from the UK by Kovacs and colleagues of surgical admissions which demonstrated that NEWS had similar performance in non-elective surgical and medical admissions; however, the study did not identify which of the surgical admissions underwent a surgical procedure (12). Although the AUCs were comparable to the general population in our prior work, the sensitivities and specificities for eCART, MEWS and NEWS were lower in post-operative patients suggesting that higher thresholds may be needed to avoid alarm fatigue. This may be due to the physiological perturbation brought on by the postoperative inflammatory state, volume shifts, and pain (10).

As might be expected, patients who experienced the outcome had longer preoperative and postoperative hospital stays and were more likely to have undergone procedures resulting in an immediate postoperative ICU transfer. Compared to patients who did not have an outcome, patients who did had an elevated eCART score immediately following their procedure and had higher eCART scores throughout their hospital stay. Comparison of median eCART scores from patients with and without the outcome as a function of time reveals that the scores of the two groups trend in opposite directions and that the interquartile ranges of the two populations diverge at about 8 hours prior to the outcome. This suggests an adequate window for assessment and an appropriate response such as rapid response team activation or ICU transfer.

Evaluation of the predictive accuracy of select vital signs and labs revealed that several single variables, such respiratory rate, heart rate, and BUN also had high accuracies in outcome prediction when compared to that of other single parameters. However, as expected, these accuracies were significantly lower than that of the early warning scores. Surprisingly the maximum postoperative temperature was found to be an inverse predictor in this patient population (AUC 0.46 [95% CI: 0.44–0.48]). Therefore, fever in a postoperative patient is at best non-predictive, and may even indicate a slightly lower probability of experiencing a severe adverse event.

Our analysis has several limitations. This was a single center study, which limits the generalizability of the findings. Similarly, despite the fact that we used ICU transfer as a broad marker for clinical decline, it is important to note that individual hospitals can have variable protocols regarding ICU transfer. Our data is limited with regards to verifiable documentation of other specific outcome measures such as intubation, DVT formation, or specific indication for return to the OR (planned operation vs. adverse event). Furthermore, in order to eliminate endoscopic and interventional radiology procedures, we employed intraoperative notes and a conservative text based exclusion criteria. Although this approach was regimented in its application, it did allow for the possibility of misclassification of some procedures. Finally, it should be noted that the dataset used in the development and

validation of the eCART algorithm did include some of the surgical patients in this study. However the accuracy of eCART was similar when only using patients not included in the original development/validation cohort (data not shown). Our study, however, has several key strengths, including its large sample size of over 32,000 surgical patients collected over an eight year period, whereas most previous studies were limited to less than 600 patients (8–11,13). Additionally, our study compares the accuracy of multiple validated early warning scores, whereas all but one previous study assessed a single early warning score (8–12). Consequently, this is the first study of its size to both verify postoperative status and compare multiple, validated, postoperative risk assessment tools.

In conclusion, we find that early warning scores are useful for risk stratification of postoperative surgical patients. The eCART score was more accurate than both NEWS and MEWS while individual vital signs and lab values were limited in their predictive ability. Therefore, the application of eCART to the postoperative setting has the potential for increased efficiency and decreased clinical workload via evidence based allocation of scarce resources, increased patient satisfaction through fewer intrusions on low risk patients, and the potential for improved outcomes for high risk patients. Future studies to confirm the generalizability of our findings and examine the effect of eCART monitoring on patient outcomes are warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEGMENTS

REFERENCES

1. Shah N, Hamilton M. Clinical review: Can we predict which patients are at risk of complications following surgery? Crit Care. 2013;17:226. [PubMed: 23672931]

2. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg. 2013 11;217(5):833–842–3. [PubMed: 24055383]

3. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar Score for Surgery. J Am Coll Surg. 2007 2 1;204(2):201–8. [PubMed: 17254923]

4. Koperna T, Semmler D, Marian F. Risk stratification in emergency surgical patients: Is the apache ii score a reliable marker of physiological impairment? Arch Surg. 2001 1 1;136(1):55–9. [PubMed: 11146778]

5. Morgan RJM, Williams F, Wright M. An early warning scoring system for detecting developing critical illness[J]. Clin Intensive Care. 1997;8:100–14.

6. 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 1;84(4):465–70. [PubMed: 23295778]

7. Churpek MM, Yuen TC, Edelson DP. Risk Stratification of Hospitalized Patients on the Wards. Chest. 2013 6;143(6):1758–65. [PubMed: 23732586]

8. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The Value of Modified Early Warning Score (MEWS) in Surgical In-Patients: A Prospective Observational Study. Ann R Coll Surg Engl. 2006 10;88(6):571–5. [PubMed: 17059720]

9. Smith T, Den Hartog D, Moerman T, Patka P, Van Lieshout EMM, Schep NWL. Accuracy of an expanded early warning score for patients in general and trauma surgery wards. Br J Surg. 2012 2 1;99(2):192–7. [PubMed: 22183685]

10. Hollis RH, Graham LA, Lazenby JP, Brown DM, Taylor BB, Heslin MJ, et al. A Role for the Early Warning Score in Early Identification of Critical Postoperative Complications: Ann Surg. 2016 5;263(5):918–23.

11. Stark AP, Maciel RC, Sheppard W, Sacks G, Hines OJ. An Early Warning Score Predicts Risk of Death after In-hospital Cardiopulmonary Arrest in Surgical Patients. Am Surg. 2015 10;81(10): 916–21. [PubMed: 26463280]

12. Kovacs C, Jarvis SW, Prytherch DR, Meredith P, Schmidt PE, Briggs JS, et al. Comparison of the National Early Warning Score in non-elective medical and surgical patients. Br J Surg. 2016 9;103(10):1385–93. [PubMed: 27487317]

13. Cuthbertson BH, Boroujerdi M, McKie L, Aucott L, Prescott G. Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient? Crit Care Med. 2007 2;35(2):402–9. [PubMed: 17205002]

14. Churpek MM, Yuen TC, Huber MT, Park SY, Hall JB, Edelson DP. Predicting Cardiac Arrest on the Wards. Chest. 2012 5;141(5):1170–6. [PubMed: 22052772]

15. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using Electronic Health Record Data to Develop and Validate a Prediction Model for Adverse Outcomes on the Wards. Crit Care Med. 2014 4;42(4):841–8. [PubMed: 24247472]

16. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, et al. Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients. Am J Respir Crit Care Med. 2014 9 15;190(6):649–55. [PubMed: 25089847]

17. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. Crit Care Med. 2016 2;44(2):368–74. [PubMed: 26771782]
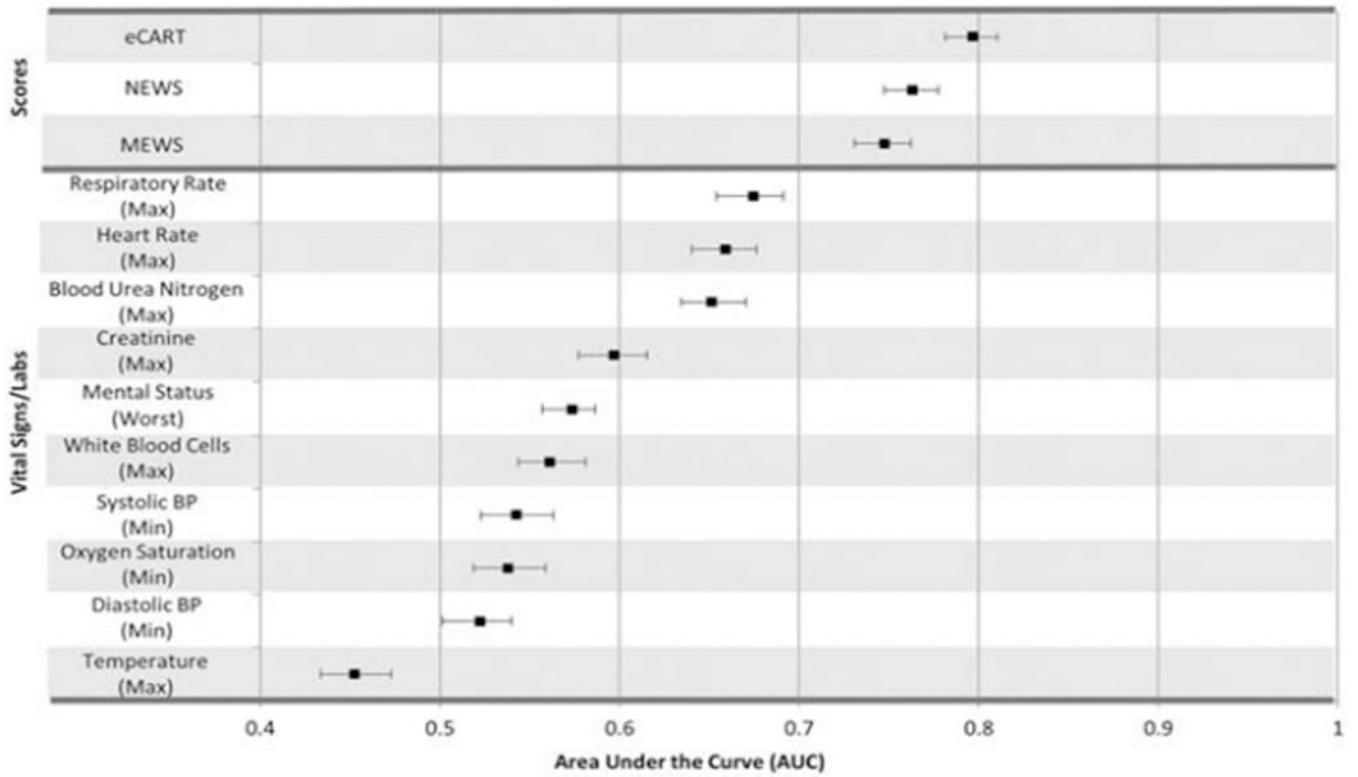
Figure 1.
AUROCs for the maximum postoperative electronic Cardiac Arrest Risk Triage score (eCART), National Early Warning Score (NEWS), Modified Early Warning Score (MEWS), and select vital signs/labs for prediction of the first postoperative ICU transfer, ward cardiac arrest, or ward death. Minimum postoperative values were used for blood pressures (BPs), and oxygen saturation. Error bars represent 95% confidence intervals.
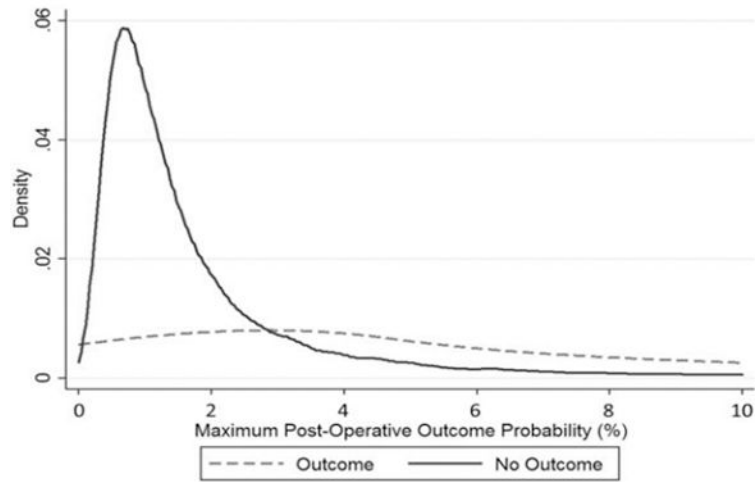
Figure 2.
Density distribution of the maximum postoperative probability of an outcome occurring as calculated by eCART (eCART score/1000), for patients who did and did not experience an outcome.

Figure 3.
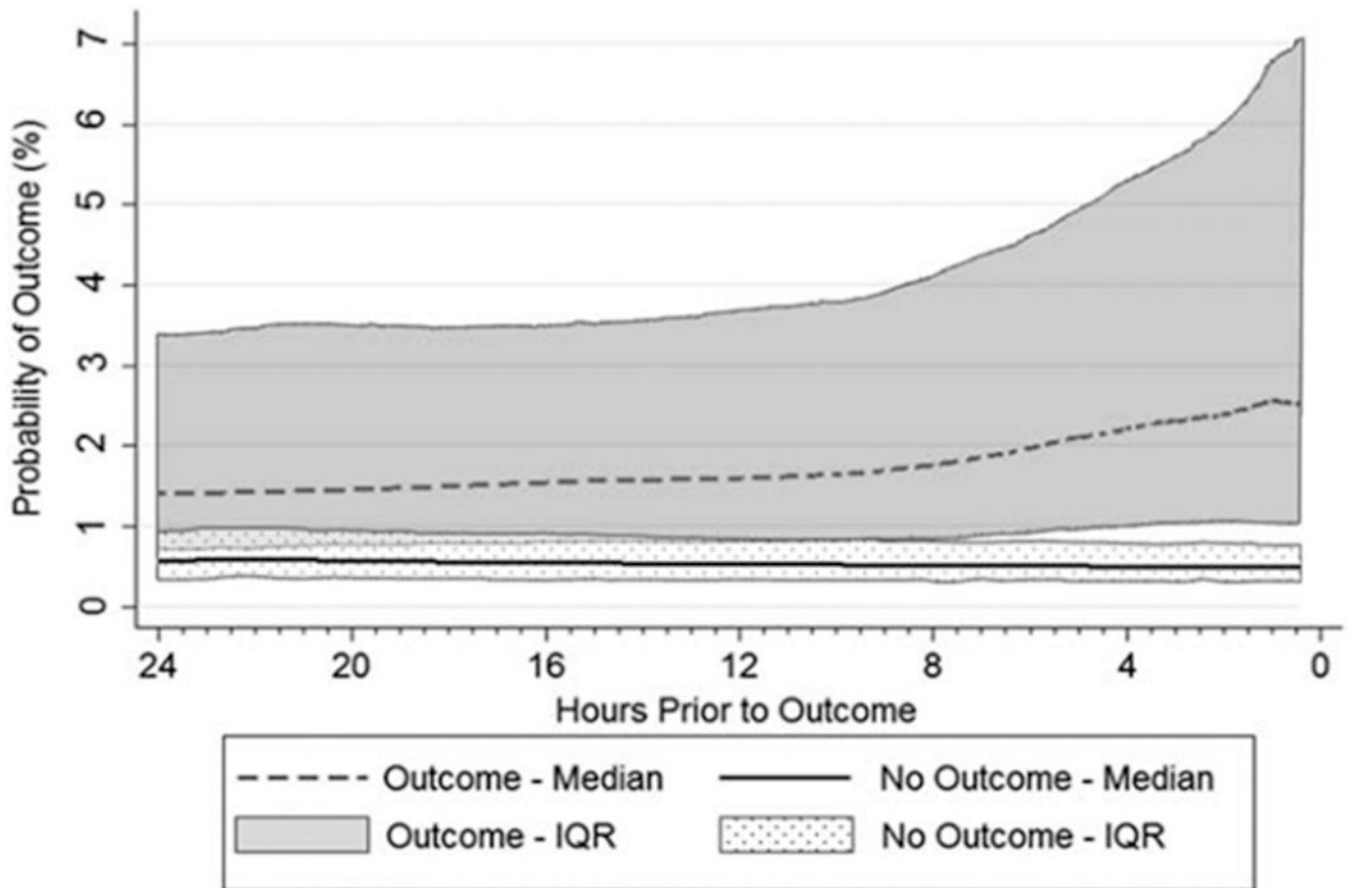Change over time of the probability of an outcome occurring as calculated by eCART (eCART score/1000), in the 24 hours prior to the first postoperative outcome with 0 on the x axis representing the time of the outcome. "No Outcome" represents a sample of 31,294 postoperative, ward location, time intervals from patients who did not experience the outcome. Probability is shown up to 30 minutes prior to outcome occurrence.

Table 1.

Patient Population Characteristics

	No Outcome (n=31,294)	Outcome (n=1,243)	P-Value
Age in years, mean (SD)	54 (17)	60 (16)	<0.001
Male, n (%)	13,465 (43%)	640 (51%)	<0.001
Race, n (%)			0.022
Black	12,295 (39%)	511 (41%)	
White	16,157 (52%)	611 (49%)	
Other/Unknown	2,842 (9%)	121 (10%)	
Planned post-op ICU transfer, n (%)	6,202 (20%)	380 (31%)	<0.001
Pre-op length of stay in days, median (IQR)	2.0 (0.9–4.7)	3.0 (1.2–6.8)	<0.001
Location Prior to OR, n (%)			<0.001
Ward	6,490 (21%)	636 (51%)	
ICU	1,013 (3%)	117 (9%)	
ER	690 (2%)	27 (2%)	
Other	1,978 (6%)	13 (1%)	
Direct Admission to OR	21,123 (68%)	450 (36%)	
Post-op length of stay in days, median (IQR)	2.9 (1.6–5.7)	11.8 (6.8–20.8)	<0.001

Definition of Abbreviations: IQR = InterQuartile Range; ICU = Intensive Care Unit; ER = Emergency Room; OR = Operating Room

Table 2.

Sensitivity, Specificity, PPV, and NPV of MEWS, NEWS, and eCART Score Cut-offs for Patients Experiencing an Outcome Compared to Those Not Experiencing Any Event

Model Cut-off	Sensitivity (%)	Specificity (%)	PPV	NPV
MEWS				
3	81	50	0.06	0.99
4	63	78	0.10	0.98
5	44	91	0.16	0.98
6	25	97	0.25	0.97
NEWS				
5	90	32	0.05	0.99
6	85	48	0.06	0.99
7	75	64	0.08	0.98
8	64	77	0.10	0.98
9	53	87	0.14	0.98
10	38	93	0.18	0.97
11	25	97	0.25	0.97
eCART probability				
0.83%	91	32	0.05	0.99
1.15%	87	48	0.06	0.99
1.20%	86	50	0.06	0.99
1.64%	80	64	0.08	0.99
2.10%	75	72	0.10	0.99
2.40%	71	77	0.11	0.99
2.50%	69	78	0.11	0.98
3.30%	63	84	0.13	0.98
3.81%	59	87	0.15	0.98
5.00%	52	91	0.19	0.98
6.11%	47	93	0.21	0.98
11.48%	30	97	0.28	0.97

NPV and PPV are calculated assuming prevalence of 3.8%

eCART probability = (eCART score)/1000

Table 3.

Score and Individual Vital Sign Characteristics

	No Outcome (n=31,294)	Outcome (n=1,243)	P-Value
First post-op eCART probability, median (IQR) - %	0.7 (0.4–1.2)	1.2 (0.6–2.3)	<0.001
Maximum post-operative value, median (IQR)			
eCART probability - %	1.2 (0.7–2.2)	5.5 (2.0–14.4)	<0.001
MEWS	3 (2–3)	4 (3–5)	<0.001
NEWS	6 (4–7)	9 (7–10)	<0.001
Respiratory Rate - breaths/min	20 (20–22)	24 (20–29)	<0.001
Heart Rate - beats/min	100 (90–111)	113 (97–130)	<0.001
Blood Urea Nitrogen - mg/dL	14 (14–19)	21 (14–36)	<0.001
Creatinine - mg/dL	0.9 (0.9–1.1)	1.1 (0.8–2.0)	<0.001
Mental Status - AVPU	A (A-V)	V (A-V)	<0.001
White Blood Cells - $\times 10^9$/L	10.8 (9.3–14.2)	12.5 (9.3–17)	<0.001
Temperature - °C	37.3 (36.9–37.7)	37.2 (36.6–37.8)	<0.001
Minimum post-operative, median (IQR)			
Diastolic BP – mm Hg	53 (47–60)	50 (43–61)	<0.001
Systolic BP – mm Hg	101 (93–111)	98 (85–115)	<0.001
Oxygen Saturation - %	93 (92–95)	93 (89–95)	<0.001

Definition of Abbreviations: IQR = InterQuartile Range; BP = Blood Pressure; AVPU = Alert, Voice, Pain, Unresponsive; eCART probability = (eCARTscore/1000)