Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients

Matthew M. Churpek[1,2], Trevor C. Yuen[1], Christopher Winslow[3], Ari A. Robicsek[3], David O. Meltzer[1], Robert D. Gibbons[2], and Dana P. Edelson[1]

[1]Department of Medicine and [2]Department of Health Studies, University of Chicago, Chicago, Illinois; and [3]Department of Medicine, NorthShore University HealthSystem, Evanston, Illinois

Abstract

Rationale: Most ward risk scores were created using subjective opinion in individual hospitals and only use vital signs.

Objectives: To develop and validate a risk score using commonly collected electronic health record data.

Methods: All patients hospitalized on the wards in five hospitals were included in this observational cohort study. Discrete-time survival analysis was used to predict the combined outcome of cardiac arrest (CA), intensive care unit (ICU) transfer, or death on the wards. Laboratory results, vital signs, and demographics were used as predictor variables. The model was developed in the first 60% of the data at each hospital and then validated in the remaining 40%. The final model was compared with the Modified Early Warning Score (MEWS) using the area under the receiver operating characteristic curve and the net reclassification index (NRI).

Measurements and Main Results: A total of 269,999 patient admissions were included, with 424 CAs, 13,188 ICU transfers, and 2,840 deaths occurring during the study period. The derived model was more accurate than the MEWS in the validation dataset for all outcomes (area under the receiver operating characteristic curve, 0.83 vs. 0.71 for CA; 0.75 vs. 0.68 for ICU transfer; 0.93 vs. 0.88 for death; and 0.77 vs. 0.70 for the combined outcome; *P* value < 0.01 for all comparisons). This accuracy improvement was seen across all hospitals. The NRI for the electronic Cardiac Arrest Risk Triage compared with the MEWS was 0.28 (0.18–0.38), with a positive NRI of 0.19 (0.09–0.29) and a negative NRI of 0.09 (0.09–0.09).

Conclusions: We developed an accurate ward risk stratification tool using commonly collected electronic health record variables in a large multicenter dataset. Further study is needed to determine whether implementation in real-time would improve patient outcomes.

Keywords: heart arrest; hospital rapid response team; decision support techniques; early diagnosis; statistical models

In-hospital cardiac arrest causes a significant healthcare burden, and its incidence is increasing in the United States (1). Hospitalized patients outside of intensive care units (ICUs) are particularly vulnerable because of infrequent physiologic measurements, variable ICU bed availability, and errors in medical decision-making that may cause patients to be left in less closely monitored ward beds too long, resulting in an arrest (2, 3). Even a delay of a few hours in transferring critically ill patients to the ICU results in increased morbidity and mortality (4). Previous studies have demonstrated that vital sign abnormalities occur hours before cardiac arrest (2, 3, 5, 6), which suggests that many of these events could be detected

At a Glance Commentary

Scientific Knowledge on the Subject: Early warning scores designed to detect physiologic decline in ward patients are typically based only on vital sign abnormalities and most of these scores were developed subjectively in individual hospitals. Development of a risk stratification tool using commonly collected electronic health record data from multiple hospitals may result in a more accurate and generalizable risk score.

What This Study Adds to the Field: Our risk score, developed from a dataset of more than 250,000 patient admissions across five hospitals, more accurately detected cardiac arrest, intensive care unit transfer, and death than the commonly used Modified Early Warning Score, and this improvement in accuracy was seen across all hospitals for all outcomes. Implementation of our risk score in real-time could help identify high-risk ward patients and improve intensive care unit triage decisions.

using physiologic data. In addition, early intervention improves outcomes in common causes of arrest, such as sepsis (7), myocardial infarction (8), and respiratory failure (9). Thus, the development of an accurate risk-stratification tool could improve the identification of high-risk ward patients resulting in earlier interventions and improved patient outcomes.

A variety of different risk stratification tools for ward patients, such as the Modified Early Warning Score (MEWS), are in use today (10–12). However, most in widespread use only use vital signs and were created using subjective opinion in individual hospitals (13, 14). This limits their accuracy and generalizability, resulting in inefficient resource use and missed opportunities to improve patient outcomes. Recently, several studies have been performed to improve on current systems, although these have been limited to single centers or have included predictors that are not available in many hospital electronic health records (EHRs) (15–20). The aim of this study was to develop an accurate and generalizable

risk-stratification tool using commonly collected EHR data from multiple hospitals. Some of the results have been previously reported in the form of an abstract (21).

Methods

Study Population
All patients hospitalized on the wards at the University of Chicago and four NorthShore University HealthSystem hospitals (Evanston, Glenbrook, Highland Park, and Skokie) from November 2008 to January 2013 were included in the study (*see* Table E1 in the online supplement). All hospitals had rapid response teams in place during the study period. These were nurse-led at the University of Chicago and physician-led at the NorthShore hospitals. No specific vital sign triggers were used to activate the teams. The study protocol was approved by the University of Chicago Institutional Review Board, and a waiver of consent was granted based on minimal harm and general impracticability (IRB #16995A).

Data Collection
Patient demographic information was obtained from administrative databases. Time- and location-stamped vital sign (temperature, heart rate, blood pressure, respiratory rate, oxygen saturation) and mental status (coded as alert, responds to voice, responds to pain, or unresponsive [AVPU] [17]), and laboratory results (white cell count, hemoglobin, platelets, sodium, potassium, chloride, bicarbonate, anion gap, blood urea nitrogen, creatinine, glucose, calcium, total protein, albumin, total bilirubin, aspartate aminotransferase, alanine aminotransferase, and alkaline phosphatase) were obtained from the Electronic Data Warehouse at NorthShore and the EHR (EPIC, Verona, WI) at University of Chicago.

Outcomes
The primary outcome was ward cardiac arrest, defined as the loss of a palpable pulse with attempted resuscitation. Cardiac arrest was identified at the University of Chicago using a prospectively validated quality improvement database, as previously described (5), and at NorthShore HealthSystem hospitals via a prospectively collected cardiac arrest log. All cardiac arrest events underwent manual chart review to ensure data quality. Secondary

outcomes were ICU transfer, death on the ward without activation of the cardiac arrest team, and the combined outcome of all three events. To make the individual outcomes mutually exclusive, ICU transfers or deaths on the wards occurring within 24 hours of a cardiac arrest were only counted as a cardiac arrest. ICU transfers occurring immediately after an intervention in a procedural suite were not counted as an event, because these might have been expected transfers. These patients were censored at the time of their last ward vital sign before the procedure and then reentered the study after returning to the ward from the ICU.

Model Development
The dataset was split into development (60%) and validation (40%) sections by date at each hospital to simulate a prospective validation of the derived model. Discrete-time survival analysis was used to develop the prediction models (16, 22–24). This method involved separating time into discrete intervals and using the predictor variable values nearest to each time interval cut-off to predict whether the combined outcome occurred within that time-block. Vital signs, laboratory values, age, number of previous ICU stays, and time since ward admission were used as potential predictor variables. In addition, pulse-pressure index (systolic minus diastolic divided by systolic blood pressure) and blood urea nitrogen to creatinine ratio were also considered. Preliminary screening for collinearity was performed using pairwise correlations between all predictors, and one of the two predictors was dropped if the correlation was greater than 0.75. When a predictor value was missing for a time interval, the previous value was carried forward. If no previous value was available, the median value for that variable was imputed, as performed in similar studies, because of the fact that these values are likely to be normal (20, 25, 26). Mental status was modeled as AVPU using indicator variables and the other continuous variables were modeled as linear splines, to allow for nonlinear effects, with knots placed *a priori* based on prior literature (10, 26, 27) and visual inspection of risk of cardiac arrest over different values of the variables in the derivation dataset. A squared term was added for the time since ward admission variable to allow the baseline risk to change over time nonlinearly. Backward stepwise

selection was used for variable selection using a *P* value of 0.157, which approximates the Akaike information criteria for variables with one degree of freedom (28).

Several modeling approaches were compared in the development sample by using 10-fold cross-validation to calculate the area under the receiver operating characteristic curve (AUC) for cardiac arrest. First, interactions were tested between the predictors in the model and the predefined variables of age, number of previous ICU stays, and time since ward admission. If all the above models had similar accuracy, the most parsimonious model would be selected. In addition, because cardiac arrest comprises a small proportion of all patients in the combined outcome, the importance of these patients in model estimation was then increased by using frequency weights by factors of 10, 25, 50, 75, and 100 to determine if this further improved model accuracy for cardiac arrest without significantly decreasing accuracy for ICU transfer (29). Finally, the time-block length was changed from 8 hours to 4 hours to determine if this affected model accuracy.

Model Validation
Model accuracy was determined by calculating the predicted probability of having an event using the final derived model and the MEWS for every observation in the validation dataset. AUCs were calculated and compared for each outcome using whether an event occurred within 24 hours of an observation in the validation dataset, because this method is commonly used for model comparisons in prior literature (30). Additionally, each patient's highest predicted probability of the event and MEWS from the period from ward admission until 30 minutes before the event was also used to calculate AUCs. A sensitivity analyses was performed by comparing the AUCs between the derived model and the MEWS by omitting death without attempted resuscitation from the combined outcome. Additionally, accuracy was compared between the models for detecting cardiac arrest within the first 8 hours of admission to determine if they differed in terms of detecting deterioration early in the admission. Finally, a net reclassification index (NRI) was calculated to estimate the clinical benefit of the derived model by comparing the

proportion of patients correctly reclassified to the MEWS at previously published cutpoints used to denote intermediate risk (3–4) and high risk (>4) for deterioration (31–33). After model validation, regression coefficients for the final model were reestimated using the entire dataset, as recommended by Steyerberg (34). All analyses were performed using Stata version 12.1 (StataCorps, College Station, TX), with a *P* value less than 0.05 denoting statistical significance.

Results

Study Population Characteristics
A total of 269,999 patient admissions occurred with at least one vital sign on the ward and all were included in the study, with 13,188 ICU transfers, 424 cardiac arrests, and 2,840 deaths on the ward without attempted resuscitation occurring during the study period. Hospital and patient characteristics in the derivation, validation, and total population are shown in Table 1. Characteristics of the individual hospitals can be found in Table E1. The overall study population had a mean age of 60 years, 60% were female, and 52% were white. The derivation and validation cohorts were similar but a greater proportion of patients were white in the validation cohort (59% vs. 47%), and the rates of adverse events were lower in the validation cohort (1.6 vs. 1.5 cardiac arrests per 1,000 admissions, 50 vs. 47 ICU transfers per 1,000 admissions, and 12 vs. 9 deaths on the wards without a cardiac arrest or ICU transfer per 1,000 admissions).

Model Development
Aspartate aminotransferase and alanine aminotransferase had a correlation of 0.79 so alanine aminotransferase was dropped from consideration in the model-building process. All other pairwise correlations were less than 0.75. Variable missingness differed by data type, with vital signs having the least percent missing (all <1% except oxygen saturation [10%] and AVPU [19%]), followed by complete blood count (7–8%), electrolytes and renal function tests (11–16%), and liver function tests (48–50%). In the development dataset, interaction terms did not improve model accuracy so they were removed. Changing the time-block length did not improve model accuracy so 8-hour intervals were used for the final model. Weighting cardiac arrest patients by a factor of 25 in model estimation improved accuracy for cardiac arrest in the development dataset (AUC, 0.85 vs. 0.83 without weights) without a substantial decrement in the AUC for ICU transfer (AUC, 0.77 vs. 0.77). The predictor variable regression coefficients for the final model are shown in Table E2.

Model Validation
The predicted probability of an event calculated from the regression equation estimated from the development cohort was multiplied by 1,000 to create a score that we termed the electronic Cardiac Arrest Risk Triage (eCART) score. During model validation, the eCART score was more accurate than the MEWS for detecting all outcomes using whether an event occurred within 24 hours of an observation (AUC,

Table 1. Hospital and Patient Characteristics for the Derivation, Validation, and Total Population

	Derivation Cohort (n = 162,088)	Validation Cohort (n = 107,911)	Total (n = 269,999)
Age, mean (SD), yr	60 (20)	60 (20)	60 (20)
Female sex, n (%)	97,978 (60)	64,303 (60)	162,281 (60)
Race			
Black, n (%)	27,870 (17)	21,800 (20)	49,670 (18)
White, n (%)	76,237 (47)	64,193 (59)	140,430 (52)
Other/unknown, n (%)	57,981 (36)	21,918 (20)	79,899 (30)
Cardiac arrests, n (rate per 1,000 admissions)	264 (1.6)	160 (1.5)	424 (1.6)
ICU transfers, n (rate per 1,000 admissions)	8,143 (50)	5,045 (47)	13,188 (49)
Deaths, n (rate per 1,000 admissions)	1,902 (12)	938 (9)	2,840 (11)

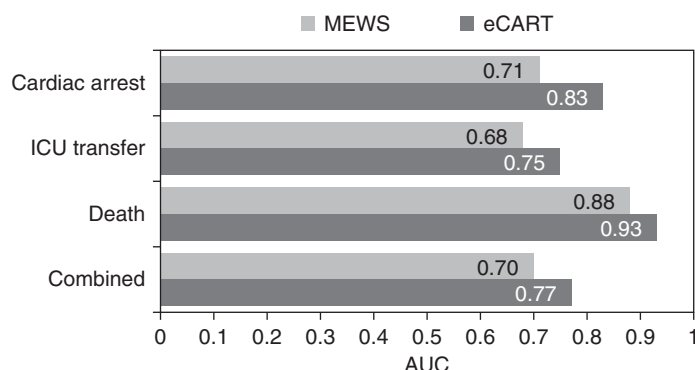*Definition of abbreviation*: ICU = intensive care unit.

Figure 1. Areas under the receiver operating characteristic curves (AUCs) for the Modified Early Warning Score (MEWS) and electronic Cardiac Arrest Risk Triage (eCART) score for whether an event occurred within 24 hours of an observation. ICU = intensive care unit.

0.83 [95% confidence interval (CI), 0.82–0.83] vs. 0.71 [95% CI, 0.70–0.73] for cardiac arrest; 0.75 [95% CI, 0.74–0.75] vs. 0.68 [95% CI, 0.68–0.68] for ICU transfer; 0.93 [95% CI, 0.93–0.93] vs. 0.88 [95% CI, 0.88–0.88] for death; and 0.77 [95% CI, 0.76–0.77] vs. 0.70 [95% CI, 0.70–0.70] for the combined outcome; *P* value < 0.01 for all comparisons), as shown in Figure 1. A sensitivity analysis was performed by excluding patients who died without attempted resuscitation from the combined outcome, because these would likely be palliative care patients, and the AUC difference between the eCART score and the MEWS was similar (0.75 [0.75–0.75] vs. 0.68 [0.68–0.68]; *P* < 0.001). At a specificity of 90%, the eCART score had a sensitivity of 54% for cardiac arrest within 24 hours compared with 39% for the MEWS

(Table 2). For patients detected by both risk scores, the eCART score first detected the arrest earlier than the MEWS (median, 37 h vs. 22 h), although this difference was not statistically significant (*P* = 0.15). Conversely, at a similar sensitivity (65% and 67% for the eCART score and MEWS), our model had a specificity of 85% versus 71% for the MEWS. Accuracy results were similar when using each patient's highest value during their ward admission (AUC, 0.83 vs. 0.74 for cardiac arrest; 0.79 vs. 0.73 for ICU transfer; 0.94 vs. 0.90 for death; and 0.82 vs. 0.76 for the combined outcome; *P* value < 0.01 for all comparisons) (*see* Figure E1). This improvement in accuracy was consistent across hospitals for all outcomes (*see* Figures E2–E5). A sensitivity analysis performed by comparing the accuracy of the two models for cardiac

arrest within the first 8 hours of ward admission demonstrated similar accuracy improvement with the eCART (AUC, 0.77 vs. 0.63; *P* = 0.02). In addition, the NRI for the eCART was 0.28 (0.18–0.38), with a positive NRI of 0.19 (0.09–0.29) and a negative NRI of 0.09 (0.09–0.09) (reclassification index table shown in Table E3).

A graph of the proportion of the entire population and events detected for each eCART score cut-off is shown in Figure 2. For example, a score cut-off of greater than 50 would detect 51% of cardiac arrests, 44% of ICU transfers, and 83% of deaths by identifying the highest risk 9% of the ward population. To determine the timing of when critically ill ward patients first demonstrate signs of physiologic decline, a graph of the mean eCART score over time in the 24 hours before the event estimated using a polynomial model with a quadratic term is shown in Figure 3, with a random 24-hour time period selected for patients who did not experience an event during the ward admission. As illustrated, the mean eCART score was highest for patients who died on the wards, followed by cardiac arrest, ICU transfer, and then those patients who did not experience an event, and scores increased leading up to the adverse events.

Discussion

We developed and validated the eCART score, a risk-stratification tool for ward patients that uses commonly collected EHR data. It was more accurate than the MEWS for identifying cardiac arrest, ICU transfer, and death on the wards, and this improved accuracy was seen across all hospitals. Importantly, by calculating net reclassification indices using previously cited MEWS cut-points, we demonstrated that 19% of the cardiac arrest patients and 9% of the nonevent patients were reclassified into a more appropriate risk group by the eCART score. These findings suggest improvements in both the detection of critical illness and resource use that deserve future validation studies. Implementation of the eCART score in real-time would accurately identify critically ill ward patients and provide important information for ICU triage decisions. This could result in improved patient outcomes and reallocation of scarce resources to the patients who need them most.

Table 2. Sensitivity and Specificity of Different MEWS and eCART Cut-offs for Patients Suffering a Ward Cardiac Arrest Compared with Those Not Experiencing Any Event

Model Cut-off	Sensitivity (%)	Specificity (%)
MEWS		
≥2	67 (65–68)	71 (71–71)
≥3	39 (37–41)	90 (90–90)
≥4	20 (18–22)	96 (96–96)
≥5	8 (7–9)	99 (99–99)
eCART score		
≥6	89 (88–91)	52 (52–52)
≥9	78 (76–79)	73 (73–73)
≥13	65 (63–67)	85 (85–85)
≥17	54 (52–56)	90 (90–90)
≥23	39 (37–41)	94 (94–94)
≥27	33 (31–35)	96 (95–96)
≥46	20 (18–21)	98 (98–98)
≥56	16 (14–17)	99 (98–99)

Definition of abbreviations: eCART = electronic Cardiac Arrest Risk Triage; MEWS = Modified Early Warning Score.
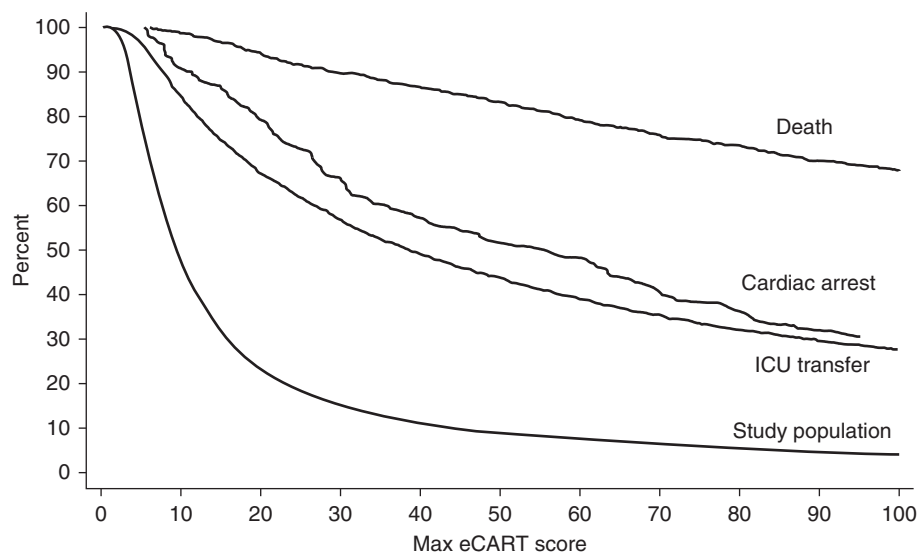
Figure 2. Shown is the percentage of patients who reached each electronic Cardiac Arrest Risk Triage (eCART) score during their ward admission in the validation cohort, with each outcome and the entire study population shown as a separate line. ICU = intensive care unit.

Several previous studies have sought to improve on current methods of ward risk-stratification using EHR data. For example, Escobar and colleagues (15) developed a prediction model for the combined outcome of ICU transfer and death on the wards without a do-not-resuscitate order in a multicenter case-control study, which was more accurate than the MEWS (AUC, 0.78 vs. 0.70). However, their model uses data not commonly found in many EHRs (e.g., longitudinal disease burden scores). Hackmann and colleagues (20) also developed a prediction model for ICU transfer using EHR data in a single-center study, finding an AUC of 0.73 in a real-time
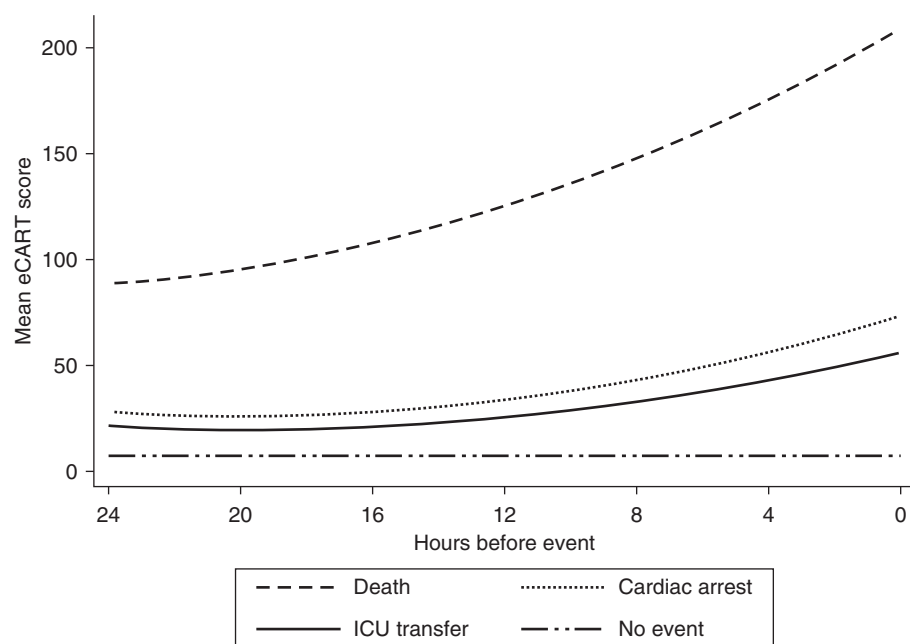
simulation. In addition, we previously published a single-center study using demographic, vital sign, and laboratory data to predict cardiac arrest and ICU transfer (16). Finally, several other authors have investigated laboratory values alone or adding specific variables, such as lactate, to previously published vital-sign–based risk scores to improve accuracy over current systems (35–38). These studies were mostly single-center and did not use other data, such as previous ICU admission, to improve model accuracy.

We developed our model for the combined outcome of cardiac arrest, ICU transfer, and death, and then we further improved its accuracy for cardiac arrest by weighting these patients more heavily when estimating the coefficients. This was done for several reasons. Most importantly, we believe that cardiac arrest is the most meaningful outcome to predict because most of these patients are inappropriately left on the ward too long instead of being transferred to the ICU, and survival to discharge for such patients is less than 25% (39). A prediction model would provide additional useful information in this setting by suggesting that these patients should be transferred to a higher level of care. ICU transfer is a recognized event, because caregivers make this triage decision, so predicting this outcome may have less value than cardiac arrest. However, some patients are transferred too late, resulting in increased morbidity and mortality (4), and this outcome is more common than ward cardiac arrest. Therefore, predicting this event may have some value, as long as it is not at the expense of predicting cardiac arrest. In addition, one study found that real-time implementation of an early warning score designed to detect ICU transfer did not improve ICU transfer rates or mortality (40). Finally, we included ward death without a cardiac arrest or ICU transfer as an outcome because it is possible that, although most of these would have been comfort care patients, some may have desired life-saving care earlier in their hospital course. Death is also commonly used as an outcome in other investigations of ward risk scores so it was included to allow for comparisons with other studies (30, 41, 42).

To estimate the impact of our risk score on the detection of cardiac arrest and resource use, we calculated an event, nonevent, and total NRI. To do this, we used

Figure 3. Change over time of the mean electronic Cardiac Arrest Risk Triage (eCART) score in the 24 hours before cardiac arrest, ICU transfer, and death and a random 24-hour period for those patients who experienced neither event. ICU = intensive care unit.

previously published cut-points used to denote intermediate risk (3–4) and high risk (>4) for deterioration (31). Our finding of an improved NRI for detecting cardiac arrest is important given that up to 80% of these patients die before leaving the hospital. In addition, the 9% reclassification for those who did not experience an event is notable given that more than 200,000 patients were admitted to the ward and did not experience an event during their hospital stay in our study. Finally, although not statistically significant, the fact that the eCART score detected arrests a median of 15 hours earlier than the MEWS suggests ample time to intervene on these critically ill patients. Thus, the eCART score has the potential to detect more events with fewer resources than the MEWS. Future clinical trials are needed to determine if implementation of our risk score in real-time would improve patient outcomes.

Our study has several limitations. First, our risk score is complex, and so calculation is best done electronically. Paper records are becoming less common the United States, but the ability to implement an electronic system varies across hospital systems and countries. Future clinical trials and cost-effectiveness analyses need to determine the trade-offs between implementing a complex score with electronic calculation compared with a simpler score, such as the MEWS.

Second, some predictors, such as oxygen delivery, comorbidities, and longitudinal chronic disease burden scores, were not available so we could not compare our model with the VitalPAC Early Warning Score (30) or the model published by Escobar and colleagues (15). Third, it is possible that some of the ICU transfers were planned transfers. We did exclude transfers that occurred immediately after an operation to try to ameliorate this issue. In addition, this would not change the main conclusions of our study because the eCART score and MEWS were compared on the same patients. Fourth, we compared the eCART and MEWS in the latter 40% of the dataset to estimate their comparative accuracy in a setting that simulates a prospective validation. However, it is possible that there may be delays or changes in the order of variables being available in real-time that could alter the accuracy of these models. In addition, external validation in a different setting, with potentially differing patient populations and laboratory collection practices, could result in decreased accuracy compared with our study. Finally, it is possible that the prospective logs missed some cardiac arrests. However, we used prospectively collected data and manually reviewed the cases at all hospitals. These patients would have experienced either death or ICU transfer after the arrest and thus would be included in the combined outcome. In addition, misclassified patients would affect the accuracy of both the eCART score and MEWS.

In conclusion, we developed and validated a risk-stratification tool for ward patients. The eCART score, which uses routinely available data, was more accurate than the MEWS for all measured outcomes and this improvement was seen across all hospitals in a diverse multicenter dataset comprised of urban tertiary, suburban teaching, and community hospitals. Implementation of our risk score in the EHR could improve detection of high-risk patients, ICU triage decisions, and identify low-risk patients who may require less frequent monitoring. ■

References

1. Merchant RM, Yang L, Becker LB, Berg RA, Nadkarni V, Nichol G, Carr BG, Mitra N, Bradley SM, Abella BS, et al.; American Heart Association Get With The Guidelines-Resuscitation Investigators. Incidence of treated cardiac arrest in hospitalized patients in the United States. Crit Care Med 2011;39:2401–2406.
2. Berlot G, Pangher A, Petrucci L, Bussani R, Lucangelo U. Anticipating events of in-hospital cardiac arrest. Eur J Emerg Med 2004;11:24–28.
3. Hodgetts TJ, Kenward G, Vlackonikolis I, Payne S, Castle N, Crouch R, Ineson N, Shaikh L. Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. Resuscitation 2002;54:115–123.
4. Young MP, Gooder VJ, McBride K, James B, Fisher ES. Inpatient transfers to the intensive care unit: delays are associated with increased mortality and morbidity. J Gen Intern Med 2003;18:77–83.
5. Churpek MM, Yuen TC, Huber MT, Park SY, Hall JB, Edelson DP. Predicting cardiac arrest on the wards: a nested case-control study. Chest 2012;141:1170–1176.
6. Schein RM, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. Chest 1990;98:1388–1392.
7. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, Peterson E, Tomlanovich M; Early Goal-Directed Therapy Collaborative Group. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001;345:1368–1377.
8. Boersma E, Maas AC, Deckers JW, Simoons ML. Early thrombolytic treatment in acute myocardial infarction: reappraisal of the golden hour. Lancet 1996;348:771–775.
9. Plant PK, Owen JL, Elliott MW. Early use of non-invasive ventilation for acute exacerbations of chronic obstructive pulmonary disease on general respiratory wards: a multicentre randomised controlled trial. Lancet 2000;355:1931–1935.
10. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalized patients on the wards. Chest 2013;143:1758–1765.
11. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted "track and trigger" systems. Resuscitation 2008;77:170–179.
12. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI, Higgins B. A review, and performance evaluation, of single-parameter "track and trigger" systems. Resuscitation 2008;79:11–21.
13. 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.
14. Cuthbertson BH, Smith GB. A warning on early-warning scores! Br J Anaesth 2007;98:704–706.
15. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D. Early detection of impending physiologic deterioration among patients who are not in intensive care: development of predictive models using data from an automated electronic medical record. J Hosp Med 2012;7:388–395.
16. 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 in the wards. Crit Care Med 2014;42:841–848.

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

18. 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; 35:402–409.

19. Cuthbertson BH, Boroujerdi M, Prescott G. The use of combined physiological parameters in the early recognition of the deteriorating acute medical patient. *J R Coll Physicians Edinb* 2010;40:19–25.

20. Hackmann G, Chen M, Chipara O, Lu C, Chen Y, Kollef M, Bailey TC. Toward a two-tier clinical warning system for hospitalized patients. *AMIA Annu Symp Proc* 2011;2011:511–519.

21. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Gibbons RD, Edelson DP. Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med* 2014; 189:A3626.

22. Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. *J Am Stat Assoc* 1988;83:414.

23. Gibbons RD, Duan N, Meltzer D, Pope A, Penhoet ED, Dubler NN, Francis C, Gill B, Guinan E, Henderson M, *et al*.; Institute of Medicine Committee. Waiting for organ transplantation: results of an analysis by an Institute of Medicine Committee. *Biostatistics* 2003;4:207–222.

24. Singer JD, Willett JB. Its about time: using discrete-time survival analysis to study duration and the timing of events. *J Educ Stat* 1993; 18:155–195.

25. van den Boogaard M, Pickkers P, Slooter AJ, Kuiper MA, Spronk PE, van der Voort PH, van der Hoeven JG, Donders R, van Achterberg T, Schoonhoven L. Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study. *BMJ* 2012;344:e420.

26. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, *et al*. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619–1636.

27. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13: 818–829.

28. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester, England: John Wiley; 2008.

29. Kuhn M, Johnson K. Applied predictive modeling. New York, NY: Springer; 2013.

30. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS— Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010;81:932–937.

31. 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;58:797–802.

32. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med* 2014;160:122–131.

33. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; 27:157–172, discussion 207–212.

34. Steyerberg EW. Clinical prediction models a practical approach to development, validation, and updating. Statistics for biology and health. New York: Springer; 2009.

35. Kho A, Rotz D, Alrahi K, Cárdenas W, Ramsey K, Liebovitz D, Noskin G, Watts C. Utility of commonly captured data from an EHR to identify hospitalized patients at risk for clinical deterioration. *AMIA Annu Symp Proc* 2007;2007:404–408.

36. Jo S, Lee JB, Jin YH, Jeong TO, Yoon JC, Jun YK, Park BY. Modified early warning score with rapid lactate level in critically ill medical patients: the ViEWS-L score. *Emerg Med J* 2013;30: 123–129.

37. 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;84:280–285.

38. Jarvis SW, Kovacs C, Badriyah T, Briggs J, Mohammed MA, Meredith P, Schmidt PE, Featherstone PI, Prytherch DR, Smith GB. Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions. *Resuscitation* 2013;84:1494–1499.

39. Girotra S, Nallamothu BK, Spertus JA, Li Y, Krumholz HM, Chan PS; American Heart Association Get with the Guidelines–Resuscitation Investigators. Trends in survival after in-hospital cardiac arrest. *N Engl J Med* 2012;367:1912–1920.

40. Bailey TC, Chen Y, Mao Y, Lu C, Hackmann G, Micek ST, Heard KM, Faulkner KM, Kollef MH. A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. *J Hosp Med* 2013;8:236–242.

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

42. Bleyer AJ, Vidya S, Russell GB, Jones CM, Sujata L, Daeihagh P, Hire D. Longitudinal analysis of one million vital signs in patients in an academic medical center. *Resuscitation* 2011;82:1387–1392.