

# Evaluating Automated Rules for Rapid Response System Alarm Triggers in Medical and Surgical Patients

Santiago Romero-Brufau, MD<sup>1</sup>, Bruce W. Morlan, MS<sup>2</sup>, Matthew Johnson, MPH<sup>1</sup>, Joel Hickman<sup>2</sup>, Lisa L. Kirkland, MD<sup>3</sup>, James M. Naessens, ScD<sup>1,2</sup>, Jeanne Huddleston, MD<sup>1,3</sup>

<sup>1</sup>Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota; <sup>2</sup>Health Care Policy and Research, Mayo Clinic, Rochester, Minnesota; <sup>3</sup>Department of Medicine, Division of Hospital Internal Medicine, Mayo Clinic, Rochester, Minnesota.

**BACKGROUND:** The use of rapid response systems (RRS), which were designed to bring clinicians with critical care expertise to the bedside to prevent unnecessary deaths, has increased. RRS rely on accurate detection of acute deterioration events. Early warning scores (EWS) have been used for this purpose but were developed using heterogeneous populations. Predictive performance may differ in medical vs surgical patients.

**OBJECTIVE:** To evaluate the performance of published EWS in medical vs surgical patient populations.

**DESIGN:** Retrospective cohort study.

**SETTING:** Two tertiary care academic medical center hospitals in the Midwest totaling more than 1500 beds.

**PATIENTS:** All patients discharged from January to December 2011.

**INTERVENTION:** None.

**MEASUREMENTS:** Time-stamped longitudinal database of patient variables and outcomes, categorized as surgical or medical. Outcomes included unscheduled transfers to the intensive care unit, activation of the RRS, and calls for cardiorespiratory resuscitation (“resuscitation call”). The EWS were calculated and updated with every new patient variable entry over time. Scores were considered accurate if they predicted an outcome in the following 24 hours.

**RESULTS:** All EWS demonstrated higher performance within the medical population as compared to surgical: higher positive predictive value ( $P < .0001$  for all scores) and sensitivity ( $P < .0001$  for all scores). All EWS had positive predictive values below 25%.

**CONCLUSIONS:** The overall poor performance of the evaluated EWS was marginally better in medical patients when compared to surgical patients. *Journal of Hospital Medicine* 2017;12:217-223. © 2017 Society of Hospital Medicine

Patients typically show signs and symptoms of deterioration hours to days prior to cardiorespiratory arrest.<sup>1,2</sup> The rate of in-hospital cardiorespiratory arrest (CRA) requiring cardiopulmonary resuscitation is estimated to be 0.174 per bed per year in the United States.<sup>3</sup> After CRA, survival to discharge is estimated to be as low as 18%.<sup>3,4</sup> Efforts to predict and prevent arrest could prove beneficial.<sup>1,2</sup>

Rapid response systems (RRS) have been proposed as a means of identifying clinical deterioration and facilitating a timely response. These systems were designed to bring clinicians with critical care expertise to the bedside to prevent unnecessary deaths. They typically include an afferent limb (detects deteriorating patients), an efferent limb (responds to calls and acts to avoid further deterioration), and administrative and data analysis limbs.<sup>5,6</sup> Automatic provision of recommendations and computer-based systems are desirable components of the afferent limb of the detection system.<sup>6</sup> Both are independent predictors of improved clinical practices for clinical decision support systems.<sup>7</sup> However, the existing early warning scores (EWS) may not be ready for

automation due to low positive predictive values (PPV) and sensitivities.<sup>8</sup>

It is possible that the low discriminatory accuracy of the published EWS may be secondary to the use of aggregate patient populations for derivation of scores. We hypothesized that these EWS perform differently in medical and in surgical subpopulations. Also, the EWS need to be tested in a time-dependent manner to serve as a realistic clinical support tool for hospitalized patients.

## STUDY AIM

The aim of this study was to evaluate the differential performance of widely used EWS in medical vs surgical patients.

## METHODS

### Site

The study was conducted in an academic center with 2 hospitals in Southeastern Minnesota totaling approximately 1500 general care nonintensive care unit (ICU) beds. The Mayo Clinic Institutional Review Board approved the research proposal.

### Subjects

Our retrospective cohort was comprised of all adult inpatients discharged from 2 academic hospitals between January 1, 2011 and December 31, 2011 who spent any time in a general care (non-ICU) unit. We excluded patients younger

\*Address for correspondence and reprint requests: Lisa L. Kirkland MD, 200 First Street SW, Rochester, MN 55905; Telephone: 507-255-8715; Fax: 507-255-9189; E-mail: kirkland.lisa@mayo.edu

Received: October 26, 2015; Revised: August 31, 2016; Accepted: September 5, 2016

2017 Society of Hospital Medicine DOI 10.12788/jhm.2712

than 18 years, psychiatric or rehabilitation inpatients, those without research authorization, and patients admitted for research purposes.

Study patients were divided into medical and surgical cohorts. Hospitalizations were considered surgical if patients had surgery at any time during their hospital stay according to billing data. A trigger was an instance in which a patient met the conditions of a specific rule (score/vital sign exceeded the published/defined threshold).

A resuscitation call was defined as a call for cardiopulmonary resuscitation when a patient has a CRA.

An event was an occurrence of 1 of the following in a general care setting: unplanned transfer to the ICU, resuscitation call, or RRS activation.

The RRS activation criteria consisted of an “acute and persistent change” in any 1 or more of the following: oxygen saturations less than 90%, heart rate less than 40 or greater than 130 beats/minute, systolic blood pressure less than 90 mm Hg, or respiratory rate less than 10 or greater than 28 breaths/minute. The RRS activation requires health provider action; they are not electronically generated. Nurses and physicians may also activate the RRS if they are concerned about a patient, even if calling criteria are not met. This is in contrast to the EWS analyzed, which are aggregate composites of multiple parameters. However, whether or not a derangement in vital signs is considered an “acute and persistent change” still involves clinical judgment. Any movement from a general care bed to an ICU bed, or from a general care bed to a procedure area, and from there to an ICU, was considered unplanned. Transfers to the ICU directly from the emergency department or operating room (OR) were not considered as an unplanned transfer and were not included in the analyses.

Coverage time was the period observed for events after a rule was triggered. In this analysis, a coverage time of 24 hours was considered, with a 1-hour look-back. A trigger was counted as a true positive if an event occurred during the following 24 hours. The 1-hour look-back was included to take into account the nursing clinical process of prioritizing a call to the RRS followed by documentation of the altered vital signs that prompted the call.

An episode was the continuous time on the general care floor within a hospitalization, excluding times when a patient was in the OR or ICU. For example, if a patient was admitted to a general bed on a surgery floor, subsequently went to the OR, and then returned to the surgery floor, the 2 episodes were considered separate: the time on the floor before surgery, and the time on the floor after surgery.

Assessment of implementation of RRS in our hospitals showed a significant drop in the failure-to-rescue rate (issues considered related to delay or failure to identify or intervene appropriately when a patient was deteriorating, as identified through mortality review) and a decrease in non-ICU mortality.<sup>9,10</sup> This suggests that our current process captures many of the relevant episodes of acute deterioration when a rapid response team is needed and supports using RRS activation as outcomes.

## Data Sources

We developed a time-stamped longitudinal database of patient data from the electronic health record, including vital signs, laboratory test results, demographics (age, sex), administrative data (including length of stay), comorbidities, resuscitation code status, location in hospital, and at the minute level throughout each patient’s hospital stay. Physiologically impossible values (eg, blood pressures of 1200 mm Hg) were considered entered in error and eliminated from the database. Time spent in the OR or ICU was excluded because RRS activation would not be applied in these already highly monitored areas. SAS Statistical software (SAS Institute Inc. Cary, North Carolina) was used for database creation.

We applied the current RRS calling criteria in our institution and calculated the Kirkland score,<sup>11</sup> along with some of the most widely used early warning scores:<sup>12</sup> Modified Early Warning System (MEWS),<sup>13</sup> Standardized Early Warning Scoring System (SEWS),<sup>14</sup> Global Modified Early Warning Score (GMEWS),<sup>15</sup> Worthing physiologic scoring system,<sup>16</sup> National Early Warning Score (NEWS),<sup>17</sup> and VitaPAC Early Warning Score (ViEWS).<sup>18</sup> Published thresholds for these scores were used to create rule triggers in the data. Once a trigger was created to calculate the number of false positives and true positives, all subsequent triggers were ignored until the end of the episode or until 24 hours elapsed. We calculated triggers in a rolling fashion throughout the episodes of care. The EWS score was updated every time a new parameter was entered into the analytical electronic health record, and the most recent value for each was used to calculate the score. SAS statistical software was used for calculation of scores and identification of outcomes.

For our analysis, events were treated as dependent variables, and triggers were independent variables. We calculated the score for each EWS to the minute level throughout our retrospective database. If the score for a specific EWS was higher than the published/recommended threshold for that EWS, an alert was considered to have been issued, and the patient was followed for 24 hours. If the patient had an event in the subsequent 24 hours, or 1 hour before (1-hour look-back), the alert was considered a true positive; if not, a false positive. Events that were not preceded by an alert were false negatives, and 24-hour intervals without either an alert or an event were considered true negatives. This simulation exercise was performed for each EWS in both subcohorts (medical and surgical). Clusters of RRS calls followed by transfers to the ICU within 3 hours were considered as a single adverse event (RRS calls, as it was the first event to occur) to avoid double counting. We have described how well this simulation methodology,<sup>8</sup> correlates with results from prospective studies.<sup>19</sup>

## Statistical Analysis

To calculate whether results were statistically significant for subgroups, a jackknife method of calculating variance<sup>20</sup> was used. The jackknife method calculates variance by repeating

**TABLE 1. Patient Characteristics, Events, and Triggers**

	All	Surgical	Medical
<b>Patient characteristics</b>			
Total patients (N <sup>a</sup> )	34,898	20,176	14,722
Male patients (N [%])	17,001 (48.7%)	10,368 (51.4%)	6633 (45.1%)
Age (mean)	58.6	58.9	58.2
Total hospitalizations	46,366	23,831	22,535
Total episodes (N)	75,240	46,275	28,965
Total time points (N)	16,780,669	11,878,268	4,902,401
<b>LoS (d)<sup>a</sup></b>			
LOS, hospitalization (median, quartiles)			
25%	2	2	2
50%	3	3	3
75%	6	6	5
<b>LOS, episode. (median, quartiles)</b>			
25%	0	0	1
50%	2	2	2
75%	3	3	4
<b>Events<sup>b</sup></b>			
Total events	3,517	1,820	1,697
RRS calls	1,865	786	1,079
Code 45	203	119	84
Unscheduled transfer to ICU	1,449	915	534
<b>Event rate</b>			
Total events/100 episodes	4.67	3.93	5.86
RRS calls/100 episodes	2.48	1.70	3.73
Code 45/100 episodes	0.27	0.26	0.29
Unscheduled transfers to ICU/100 episodes	1.93	1.98	1.84
<b>Triggers by rule</b>			
RRT	174,014	101,190	72,824
GMEWS	5,043	2,338	2,705
MEWS	51,693	27,386	24,307
SEWS	116,753	63,370	53,383
VEWS	1,375,831	818,695	557,136
Worthing	1,175,736	685,693	490,043
Kirkland	602,441	321,605	280,836
<b>Triggers/d/10 hospital beds</b>			
RRT	3.0	3.1	2.9
GMEWS	0.1	0.1	0.1
MEWS	0.9	0.8	1.0
SEWS	2.0	1.9	2.1
VEWS	23.8	24.9	22.4
Worthing	20.3	20.8	19.7
Kirkland	10.4	9.8	11.3

<sup>a</sup>Patients are classified as surgical if they had at least 1 surgical hospitalization during the study period. All other metrics shown here (eg, events, triggers) are calculated per hospitalization. Hence, the total number of events for the medical group is the total number of events for all medical hospitalizations.

<sup>b</sup>Events and triggers are assigned to the group where a patient was when the event took place. "Triggers" refers to the number of instances when a patient met the EWS' triggering criteria recommended by the authors.

NOTE: Abbreviations: GMEWS, Global Modified Early Warning Score; ICU, intensive care unit; LOS, length of stay; RRS, rapid response systems; RRT, rapid response team calling criteria; SEWS, Standardized Early Warning Scoring System; VEWS, VitaPAC Early Warning Score.

the calculations of the statistic leaving out 1 sample at a time. In our case, we repeated the calculation of sensitivity and PPV leaving out 1 patient at a time. Once the simulation method had been run and the false/true positives/negatives had been assigned, calculation of each metric (PPV and sensitivity) was repeated for *n* subsamples, each leaving out 1 patient. The variance was calculated and 2 Student *t* tests were performed for each EWS: 1 for PPV and another for sensitivity. SAS statistical software v 9.3 was used for the simulation analysis; R statistical software v 3.0.2 (The R Foundation, Vienna, Austria) was used for the calculation of

the statistical significance of results. A univariable analysis was also performed to assess the sensitivity and PPVs for the published thresholds of the most common variables in each EWS: respiratory rate, systolic blood pressure, heart rate, temperature, and mental status as measured by the modified Richmond Agitation Sedation Score.<sup>21</sup>

## RESULTS

The initial cohort included 60,020 hospitalizations, of which the following were excluded: 2751 because of a lack of appropriate research authorization; 6433 because the patients

**TABLE 2.** Comparison of the Predictive Performance of Widely Used EWS in a Surgical and a Medical Population<sup>a</sup>

	Alerts (n)		Sensitivity				PPV			
	Medical	Surgical	Medical		Surgical		Medical		Surgical	
	N	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
GMEWS	688	675	0.10	0.007	0.07	0.006	0.22	0.013	0.18	0.013
KIRKL	14,210	14,560	0.58	0.008	0.49	0.008	0.06	0.002	0.06	0.002
MEWS	4040	3957	0.37	0.009	0.29	0.009	0.14	0.005	0.13	0.005
NEWS	14,269	16,197	0.64	0.007	0.53	0.008	0.07	0.002	0.06	0.002
RRT	13,527	16,120	0.59	0.008	0.53	0.008	0.07	0.002	0.06	0.002
SEWS	7832	9116	0.51	0.008	0.40	0.009	0.10	0.003	0.08	0.003
VEWS	26,034	35,856	0.78	0.005	0.70	0.006	0.05	0.001	0.03	0.001
WRTH	33,602	44,109	0.78	0.005	0.68	0.006	0.04	0.001	0.03	0.001

<sup>a</sup>All analyzed EWS sensitivity and PPV perform better in the medical than in the surgical subpopulation.

NOTE: Abbreviations: EWS, early warning scores; GMEWS, Global Modified Early Warning Score; KIRKL, Kirkland score; MEWS, Modified Early Warning Score; PPV, positive predictive value; RRT, rapid response team calling criteria; SD, standard deviation; SEWS, Standardized Early Warning Scoring System; VEWS, VitaPAC Early Warning Score; WRTH, Worthington physiologic scoring system.

were younger than 18 years; 2129 as psychiatric admissions; 284 as rehabilitation admissions; 872 as research purposes-only admissions; and 1185 because the patient was never in a general care bed (eg, they were either admitted directly to the ICU, or they were admitted for an outpatient surgical procedure and spent time in the postanesthesia care unit).

Table 1 summarizes patient and trigger characteristics, overall and by subgroup. The final cohort included 75,240 total episodes in 46,366 hospitalizations, from 34,898 unique patients, of which 48.7% were male. There were 23,831 medical and 22,535 surgical hospitalizations. Median length of episode was 2 days both for medical and surgical patients. Median length of stay was 3 days, both for medical and for surgical patients.

There were 3332 events in total, of which 1709 were RRS calls, 185 were resuscitation calls, and 1438 were unscheduled transfers to the ICU. The rate of events was 4.67 events per 100 episodes in the aggregate adult population. There were 3.93 events per 100 episodes for surgical hospitalizations, and 5.86 events per 100 episodes for medical hospitalizations ( $P < .001$ ). The number of CRAs in our cohort was 0.27 per 100 episodes, 0.128 per hospital bed per year, or 4.37 per 1000 hospital admissions, similar to other reported numbers in the literature.<sup>3, 22,23</sup>

The total number of EWS triggers varied greatly between EWS rules, with the volume ranging during the study year from 1363 triggers with the GMEWS rule to 77,711 triggers with the ViEWS score.

All scores had PPVs less than 25%. As seen in Table 2 and shown graphically in the Figure, all scores performed better on medical patients (blue) than on surgical patients (yellow). The  $P$  value was  $< .0001$  for both PPV and sensitivity. The Worthing score had the highest sensitivity (0.78 for medical and 0.68 for surgical) but a very low PPV (0.04 for medical and 0.03 for surgical), while GMEWS was the opposite: low sensitivity (0.10 and 0.07) but the highest PPV (0.22 and 0.18).

The results of the univariable analysis can be seen in Table 3. Most of the criteria performed better (higher sensitivity and PPV) as predictors in the medical hospitalizations than in the surgical hospitalizations.

## DISCUSSION

We hypothesized that EWS may perform differently when applied to medical rather than surgical patients. Studies had not analyzed this in a time-dependent manner,<sup>24-26</sup> which limited the applicability of the results.<sup>8</sup>

All analyzed scores performed better in medical patients than in surgical patients (Figure). This could reflect a behavioral difference by the teams on surgical and medical floors in the decision to activate the RRS, or a bias of the clinicians who designed the scores (mostly nonsurgeons). The difference could also mean that physiological deteriorations are intrinsically different in patients who have undergone anesthesia and surgery. For example, in surgical patients, a bleeding episode is more likely to be the cause of their physiological deterioration, or the lingering effects of anesthesia could mask underlying deterioration. Such patients would benefit from scores where variables such as heart rate, blood pressure, or hemoglobin had more influence.

When comparing the different scores, it was much easier for a patient to meet the alerting score with the Worthing score than with GMEWS. In the Worthing score, a respiratory rate greater than 22 breaths per minute, or a systolic blood pressure less than 100 mm Hg, already meet alerting criteria. Similar vital signs result in 0 and 1 points (respectively) in GMEWS, far from its alerting score of 5. This reflects the intrinsic tradeoff of EWS: as the threshold for considering a patient "at risk" drops, not only does the number of true positives (and the sensitivity) increase, but also the number of false positives, thus lowering the PPV.

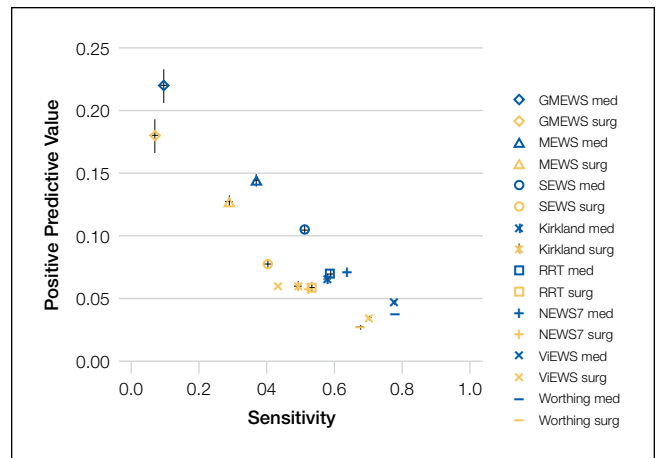
However, none of the scores analyzed were considered to perform well based on their PPV and sensitivity, particularly in the surgical subpopulation. Focusing on another metric,

**TABLE 3.** Univariable analysis in the medical and surgical subpopulations<sup>a</sup>

Criteria	Sensitivity		PPV	
	Medical	Surgical	Medical	Surgical
RR <8	0.018	0.036	0.019	0.030
RR <15	0.218	0.323	0.023	0.016
RR >20	0.621	0.511	0.048	0.045
RR >22	0.546	0.427	0.060	0.055
RR >25	0.414	0.308	0.085	0.079
RR >30	0.229	0.162	0.124	0.110
RR >35	0.137	0.097	0.162	0.133
HR <40	0.078	0.077	0.041	0.003
HR <50	0.131	0.138	0.027	0.026
HR >100	0.599	0.558	0.040	0.031
HR >110	0.478	0.413	0.061	0.046
HR >120	0.332	0.288	0.087	0.066
HR >130	0.229	0.187	0.118	0.088
HR >140	0.137	0.114	0.139	0.103
SBP <70	0.069	0.085	0.162	0.139
SBP <80	0.165	0.175	0.111	0.092
SBP <90	0.290	0.290	0.057	0.048
SBP <100	0.432	0.421	0.031	0.025
SBP >180	0.078	0.091	0.042	0.054
SBP >200	0.034	0.027	0.085	0.090
SBP >220	0.011	0.008	0.131	0.107
SpO <sub>2</sub> <88	0.408	0.353	0.054	0.034
SpO <sub>2</sub> <90	0.522	0.452	0.043	0.028
SpO <sub>2</sub> <91	0.590	0.514	0.037	0.025
SpO <sub>2</sub> <92	0.657	0.605	0.031	0.022
SpO <sub>2</sub> <94	0.795	0.786	0.023	0.018
SpO <sub>2</sub> <95	0.844	0.848	0.020	0.016
SpO <sub>2</sub> <96	0.905	0.909	0.019	0.015
Temp <34.5°	0.008	0.008	0.052	0.076
Temp <35°	0.018	0.013	0.044	0.048
Temp <35.5°	0.032	0.020	0.035	0.038
Temp >38°	0.151	0.110	0.069	0.066
Temp >38.5°	0.103	0.078	0.082	0.110
Temp >39°	0.070	0.050	0.095	0.142
Temp >40°	0.004	0.003	0.291	0.400
mRASS ≠ 0	0.566	0.550	0.030	0.023
mRASS >1 or <-1	0.201	0.159	0.039	0.147
mRASS >2 or <-2	0.081	0.058	0.147	0.114
mRASS >3 or <-3	0.044	0.034	0.191	0.193

<sup>a</sup>Table shows the sensitivity and positive predictive value in the following 24 hours after a patient meets each of the criteria.

NOTE: Abbreviations: HR, heart rate (beats/minute); mRASS, modified Richmond Agitation Sedation Score (+4 to -5); RR, respiratory rate (breaths/minute); PPV, positive predictive value; SBP, systolic blood pressure (mm Hg); SpO<sub>2</sub>, peripheral oxygen saturation (%); temp, temperature (C).

**FIG.** Performance of scores in medical and surgical patients.

NOTE: Abbreviations: GMEWS, Global Modified Early Warning Score; Kirkland, Kirkland score; med, medical; MEWS, Modified Early Warning Score; NEWS 7, National Early Warning Score 7; RRT, rapid response team; SEWS, Standardized Early Warning Scoring System; surg, surgical; VIEW5, VitaPAC Early Warning Score; Worthing, Worthington physiologic scoring system.

the area under the receiver operator curve can give misleadingly optimistic results.<sup>24,27</sup> However, the extremely low prevalence of acute physiological deterioration can produce low PPVs even when specificity seems acceptable, which is why it is important to evaluate PPV directly.<sup>28</sup>

To use EWS effectively to activate RRS, they need to be combined with clinical judgment to avoid high levels of false alerts, particularly in surgical patients. It has been reported that RRS is activated only 30% of the time a patient meets RRS calling criteria.<sup>29</sup> While there may be cultural characteristics inhibiting the decision to call,<sup>30</sup> our study hints at another explanation: if RRS was activated every time a patient met calling criteria based on the scores analyzed, the number of RRS calls would be very high and difficult to manage. So health providers may be doing the right thing when “filtering” RRS calls and not applying the criteria strictly, but in conjunction with clinical judgment.

A limitation of any study like this is how to define “acute physiological deterioration.” We defined an event as recognized episodes of acute physiological deterioration that are signaled by escalations of care (eg, RRS, resuscitation calls, or transfers to an ICU) or unexpected death. By definition, our calculated PPV is affected by clinicians’ recognition of clinical deteriorations. This definition, common in the literature, has the limitation of potentially underestimating EWS’ performance by missing some events that are resolved by the primary care team without an escalation of care. However, we believe our interpretation is not unreasonable since the purpose of EWS is to trigger escalations of care in a timely fashion. Prospective studies could define an event in a way that is less affected by the clinicians’ judgment.

Regarding patient demographics, age was similar between the 2 groups (average, 58.2 years for medical vs 58.9 years for surgical), and there was only a small difference in gender ratios (45.1% male in the medical vs 51.4% in the surgical group). These differences are unlikely to have affected the

results significantly, but unknown differences in demographics or other patient characteristics between groups may account for differences in score performance between surgical and medical patients.

Several of the EWS analyzed had overlapping trigger criteria with our own RRS activation criteria (although as single-parameter triggers and not as aggregate). To test how these potential biases could affect our results, we performed a post hoc sensitivity analysis eliminating calls to the RRS as an outcome (so using the alternative outcome of unexpected transfers to the ICU and resuscitation calls). The results are similar to those of our main analysis, with all analyzed scores having lower sensitivity and PPV in surgical hospitalizations when compared to medical hospitalizations.

Our study suggests that, to optimize detection of physiological deterioration events, EWS should try to take into account different patient types, with the most basic distinction being surgical vs medical. This tailoring will make EWS more complex, and less suited for paper-based calculation, but new electronic health records are increasingly able to incorporate decision support, and some EWS have been developed for electronic calculation only. Of particular interest in this regard is the score developed by Escobar et al,<sup>31</sup> which groups patients into categories according to the reason for admission, and calculates a different subscore based on that category. While the score by Escobar et al. does not split patients based on medical or surgical status, a more general interpretation of our results suggests that a score may be more accurate if it classifies patients into subgroups with different subscores. This seems to be confirmed by the fact that the score by Escobar et al performs better than MEWS.<sup>28</sup> Unfortunately, the paper describing it does not provide enough detail to use it in our database.

A recent systematic review showed increasing evidence that RRS may be effective in reducing CRAs occurring in a non-ICU setting and, more important, overall inhospital mortality.<sup>32</sup> While differing implementation strategies (eg, different length of the educational effort, changes in the frequency of vital signs monitoring) can impact the success of such an initiative, it has been speculated that the afferent limb (which often includes an EWS) might be the most critical part of the system.<sup>33</sup> Our results show that the most widely used EWS perform significantly worse on surgical patients, and suggest that a way to improve the accuracy of EWS would be to tailor the risk calculation to different patient subgroups (eg, medical and surgical patients). Plausible next steps would be to demonstrate that tailoring risk calculation to medical and surgical patients separately can improve risk predictions and accuracy of EWS.

Disclosures: The authors report no financial conflicts of interest.

## References

- 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. *Med J Aust.* 1999; 171(1):22-25.
- Schein RM, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest.* 1990;98(6):1388-1392.
- Peberdy MA, Kaye W, Ornato JP, Larkin GL, Nadkarni V, Mancini ME, et al. Cardiopulmonary resuscitation of adults in the hospital: a report of 14720 cardiac arrests from the National Registry of Cardiopulmonary Resuscitation. *Resuscitation.* 2003; 58(3):297-308.
- Nadkarni VM, Larkin GL, Peberdy MA, Carey SM, Kaye W, Mancini ME, et al. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. *JAMA.* 2006;295(1):50-57.
- Devita MA, Bellomo R, Hillman K, Kellum J, Rotondi A, Teres D, et al. Findings of the first consensus conference on medical emergency teams. *Crit Care Med.* 2006;34(9):2463-2478.
- DeVita MA, Smith GB, Adam SK, Adams-Pizarro I, Buist M, Bellomo R, et al. "Identifying the hospitalised patient in crisis"--a consensus conference on the afferent limb of rapid response systems. *Resuscitation.* 2010;81(4):375-382.
- Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ.* 2005;330(7494):765.
- Romero-Brufau S, Huddleston JM, Naessens JM, Johnson MG, Hickman J, Morlan BW, et al. Widely used track and trigger scores: are they ready for automation in practice? *Resuscitation.* 2014;85(4):549-552.
- Huddleston JM, Diedrich DA, Kinsey GC,ENZler MJ, Manning DM. Learning from every death. *J Patient Saf.* 2014;10(1):6-12.
- Moriarty JP, Schiebel NE, Johnson MG, Jensen JB, Caples SM, Morlan BW, et al. Evaluating implementation of a rapid response team: considering alternative outcome measures. *Int J Qual Health Care.* 2014;26(1):49-57.
- Kirkland LL, Malinchoc M, O'Byrne M, Benson JT, Kashiwagi DT, Burton MC, et al. A clinical deterioration prediction tool for internal medicine patients. *Am J Med Qual.* 2013;28(2):135-142.
- Griffiths JR, Kidney EM. Current use of early warning scores in UK emergency departments. *Emerg Med J.* 2012;29(1):65-66.
- Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. *QJM.* 2001;94(10):521-526.
- Paterson R, MacLeod DC, Theftord D, Beattie A, Graham C, Lam S, et al. Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit. *Clin Med (Lond).* 2006;6(3):281-284.
- Harrison GA, Jacques T, McLaws ML, Kilborn G. Combinations of early signs of critical illness predict in-hospital death--the SOCCER study (signs of critical conditions and emergency responses). *Resuscitation.* 2006;71(3):327-334.
- Duckitt RW, Buxton-Thomas R, Walker J, Cheek E, Bewick V, Venn R, et al. Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study. *Br J Anaesth.* 2007; 98(6):769-774.
- 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;84(4):465-470.
- Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS--Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation.* 2010;81(8):932-937.
- Romero-Brufau S, Huddleston JM. Reply to letter: widely used track and trigger scores: are they ready for automation in practice? *Resuscitation.* 2014;85(10):e159.
- Efron B, Stein C. The jackknife estimate of variance. *Annals of Statistics.* 1981;586-596.
- Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O'Neal PV, Keane KA, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med.* 2002;166(10):1338-1344.
- DeVita MA, Braithwaite RS, Mahidhara R, Stuart S, Foraida M, Simmons RL. Medical Emergency Response Improvement Team (MERIT). Use of medical emergency team responses to reduce hospital cardiopulmonary arrests. *Qual Saf Health Care.* 2004;13(4):251-254.
- Goncalves PD, Polessi JA, Bass LM, Santos Gde P, Yokota PK, Laselva CR, et al. Reduced frequency of cardiopulmonary arrests by rapid response teams. *Einstein (Sao Paulo).* 2012;10(4):442-448.
- 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(2):402-409.
- 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;88(6):571-575.
- Stenhouse C, Coates S, Tivey M, Allsop P, Parker T. Prospective evaluation of a modified Early Warning Score to aid earlier detection of patients developing critical illness on a general surgical ward. *British Journal of Anaesthesia.* 2000;84(5):663-663.

27. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted 'track and trigger' systems. *Resuscitation*. 2008;77(2):170-179.
28. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care*. 2015; 19:285.
29. Hillman K, Chen J, Cretikos M, Bellomo R, Brown D, Doig G, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. *Lancet*. 2005;365(9477):2091-2097.
30. Shearer B, Marshall S, Buist MD, Finnigan M, Kitto S, Hore T, et al. What stops hospital clinical staff from following protocols? An analysis of the incidence and factors behind the failure of bedside clinical staff to activate the rapid response system in a multi-campus Australian metropolitan healthcare service. *BMJ Qual Saf*. 2012;21(7):569-575.
31. 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(5):388-395.
32. Winters BD, Weaver SJ, Pfoh ER, Yang T, Pham JC, Dy SM. Rapid-response systems as a patient safety strategy: a systematic review. *Ann Intern Med*. 2013;158(5 pt 2):417-425.
33. Jones DA, DeVita MA, Bellomo R. Rapid-response teams. *N Engl J Med*. 2011;365(2):139-146.