Journal of HOSPITAL MEDICINE www.journalofhospitalmedicine.com

Measuring the Modified Early Warning Score and the Rothman Index: Advantages of Utilizing the Electronic Medical Record in an Early Warning System

G. Duncan Finlay, MD[1,2*], Michael J. Rothman, PhD[2], Robert A. Smith, PhD, FCCM[1]

[1]F.A.R. Institute, Sarasota, Florida; [2]PeraHealth, Inc., Charlotte, North Carolina.

Early detection of an impending cardiac or pulmonary arrest is an important focus for hospitals trying to improve quality of care. Unfortunately, all current early warning systems suffer from high false-alarm rates. Most systems are based on the Modified Early Warning Score (MEWS); 4 of its 5 inputs are vital signs. The purpose of this study was to compare the accuracy of MEWS against the Rothman Index (RI), a patient acuity score based upon summation of excess risk functions that utilize additional data from the electronic medical record (EMR). MEWS and RI scores were computed retrospectively for 32,472 patient visits. Nursing assessments, a category of EMR inputs only used by the RI, showed sharp differences 24 hours before death. Receiver operating characteristic curves for 24-hour mortality demonstrated superior RI performance with c-statistics, 0.82 and 0.93, respectively. At the point where MEWS triggers an alarm, we identified the RI point corresponding to equal sensitivity and found the positive likelihood ratio (LR+) for MEWS was 7.8, and for the RI was 16.9 with false alarms reduced by 53%. At the RI point corresponding to equal LR+, the sensitivity for MEWS was 49% and 77% for RI, capturing 54% more of those patients who will die within 24 hours. *Journal of Hospital Medicine* 2014;9:116–119. 2013 The Authors. Journal of Hospital Medicine published by Wiley Periodicals, Inc. on behalf of Society of Hospital Medicine

Bedside calculation of early warning system (EWS) scores is standard practice in many hospitals to predict clinical deterioration. These systems were designed for periodic hand-scoring, typically using a half-dozen variables dominated by vital signs. Most derive from the Modified Early Warning Score (MEWS).[1,2] Despite years of modification, EWSs have had only modest impact on outcomes.[3,4] Major improvement is possible only by adding more information than is contained in vital signs. Thus, the next generation of EWSs must analyze electronic medical records (EMRs). Analysis would be performed by computer, displayed automatically, and updated whenever new data are entered into the EMR. Such systems could deliver timely, accurate, longitudinally trended acuity information that could aid in earlier detection of declining patient condition as well as improving sensitivity and specificity of EWS alarms.

Advancing this endeavor along with others,[5,6] we previously published a patient acuity metric, the Rothman Index (RI), which automatically updates when asynchronous vital signs, laboratory test results, Braden Scale,[7] cardiac rhythm, and nursing assessments are entered into the EMR.[8] Our goal was to enable clinicians to visualize changes in acuity by simple line graphs personalized to each patient at any point in time across the trajectory of care. In our model validation studies,[8] we made no attempt to identify generalizable thresholds, though others[9] have defined decision cut points for RI in a nonemergent context. To examine decision support feasibility in an emergent context, and to compare RI with a general EWS standard, we compare the accuracy of the RI with the MEWS in predicting hospital death within 24 hours.

METHODS

Site Description and Ethics

The institutional review board of Abington Memorial Hospital (Abington, PA) approved collection of retrospective data obtained from their 665-bed, regional referral center and teaching hospital. Handling of patient information complied with the Health Insurance Portability and Accountability Act of 1996 regulations.

Patient Inclusion

The analysis included all patients, aged 18 years or older, admitted from July 2009 through June 2010, when there were sufficient data in the EMR to compute the RI. Obstetric and psychiatric patients were excluded because nursing documentation is insufficient in this dataset.

Data Collection/Data Sources

Clinical variables were extracted from the EMR (All-Scripts Sunrise Clinical Manager, Chicago, IL) by SQL query and placed into a database. RI[8] and MEWS[1] were computed according to published methods. Table 1 shows definitions of standards for each nursing assessment,[8] and Table 2 identifies all clinical

*Address for correspondence and reprint requests: G. Duncan Finlay, MD, 5019 Kestral Park Dr., Sarasota, FL 34231; Telephone: 866-794-0837; Fax: 866-255-0783; E-mail: duncan.finlay@FARinstitute.org

Additional Supporting Information may be found in the online version of this article.

Received: August 24, 2013; **Revised:** November 19, 2013; **Accepted:** November 20, 2013

Published 2013. The Authors Journal of Hospital Medicine published by Wiley Periodicals, Inc. on behalf of Society of Hospital Medicine DOI 10.1002/jhm.2132

TABLE 1. Nursing Assessments

Cardiac	Pulse regular, rate 60–100 bpm, skin warm and dry. Blood pressure <140/90 and no symptoms of hypotension.
Food/nutrition	No difficulty with chewing, swallowing, or manual dexterity. Patient consuming >50% of daily diet ordered as observed or stated.
Gastrointestinal	Abdomen soft and nontender. Bowel sounds present. No nausea or vomiting. Continent. Bowel pattern normal as observed or stated.
Genitourinary	Voids without difficulty. Continent. Urine clear, yellow to amber as observed or stated. Urinary catheter patent if present.
Musculoskeletal	Independently able to move all extremities and perform functional activities as observed or stated (includes assistive devices).
Neurological	Alert and oriented to person, place, time, situation. Speech is coherent.
Peripheral-vascular	Extremities are normal or pink and warm. Peripheral pulses palpable. Capillary refill <3 seconds. No edema, numbness or tingling.
Psychosocial	Behavior appropriate to situation. Expressed concerns and fears being addressed. Adequate support system.
Respiratory	Respiration 12–24/minute at rest, quiet and regular. Bilateral breath sounds clear. Nail beds and mucous membranes pink. Sputum clear, if present.
Safety/fall risk	Safety/fall risk factors not present. Not a risk to self or others.
Skin/tissue	Skin clean, dry, and intact with no reddened areas. Patient is alert, cooperative and able to reposition self independently. Braden Scale >15.

NOTE: Nursing assessment data are collected in the course of head-to-toe patient examinations performed once each shift and recorded in structured data fields within the electronic medical record. For hospitals that do not use these standards, Rothman Index input variables are derived from nursing observations (eg, nail beds pink).

TABLE 2. Comparison of Input Variables Used to Derive Modified Early Warning Score and Rothman Index Risk Scores

Input Variable	A: Alive in 24 Hours, Mean (SD)	B: Dead Within 24 Hours, Mean (SD)	*P* Value
Diastolic blood pressure, mm Hg	66.8 (13.5)	56.6 (16.8)	<0.0001
Systolic blood pressure, mm Hg*	127.3 (23.8)	105.2 (29.4)	<0.0001
Temperature, °F*	98.2 (1.1)	98.2 (2.0)	0.1165
Respiration, breaths per minute*	20.1 (4.7)	23.6 (9.1)	<0.0001
Heart rate, bpm*	81.1 (16.5)	96.9 (22.2)	<0.0001
Pulse oximetry, % O_2 saturation	96.3 (3.3)	93.8 (10.1)	<0.0001
Creatinine, mg/dL	1.2 (1.2)	1.8 (1.5)	<0.0001
Blood urea nitrogen, mg/dL	23.9 (17.9)	42.1 (26.4)	<0.0001
Serum chloride, mmol/L	104.3 (5.4)	106.9 (9.7)	<0.0001
Serum potassium, mmol/L	4.2 (0.5)	4.4 (0.8)	<0.0001
Serum sodium, mmol/L	139.0 (4.1)	140.7 (8.5)	<0.0001
Hemoglobin, gm/dL	11.2 (2.1)	10.6 (2.1)	<0.0001
White blood cell count, 10^3 cell/μL	9.9 (6.3)	15.0 (10.9)	<0.0001
Braden Scale, total points	17.7 (3.4)	12.2 (3.1)	<0.0001
NURSING ASSESSMENTS	A: Alive in 24 Hours and Failed Standard	B: Dead Within 24 Hours and Failed Standard	*P* Value
Neurological	38.7%	91.4%	<0.0001
Genitourinary	46.6%	90.0%	<0.0001
Respiratory	55.6%	89.0%	<0.0001
Peripheral vascular	54.1%	86.9%	<0.0001
Food	28.3%	80.6%	<0.0001
Skin	56.3%	75.0%	<0.0001
Gastrointestinal	49.3%	75.0%	<0.0001
Musculoskeletal	50.3%	72.4%	<0.0001
Cardiac	30.4%	59.8%	<0.0001
Psychosocial	24.6%	40.9%	<0.0001
Safety	25.5%	29.0%	<0.0001
A/V/P/U score*	96.3/2.1/1.4/0.2%	88.6/21.6/4.6/5.3%	<0.0001
Sinus rhythm (absent)†	34.9%	53.3%	<0.0001

NOTE: Each observation is classified according to 24-hour mortality: column A = this patient will live at least for the next 24 hours; column B = this patient will die within the next 24 hours. The dataset consisted of 32,472 patients with a total of 1,794,910 observations: 12,514 in the last 24 hours before death and 1,782,396 for patients who did not die within the next 24 hours. In the latter group are 1,708,434 observations for patients who survived and 73,962 for patients who later died (after the 24-hour window that defined a true positive). *P* values for continuous variables use the *t* test with Cochran and Cox approximation for unequal variance. *P* values for discrete variables are from the χ^2 test (each nursing assessment is mapped to binary pass or fail). **Abbreviations:** A/V/P/U, alert/voice/pain/unresponsive; SD, standard deviation.

*Modified Early Warning Score uses these 5 variables; Rothman Index uses 26 variables (all the variables in this table except A/V/P/U score).

†Sinus rhythm is the normal heart pattern; when absent the Rothman Index associates risk with 8 abnormal patterns.

variables employed for each system. Briefly, RI utilizes 26 variables related to clinical care and routinely available in the EMR. These include vital signs, laboratory results, cardiac rhythms, and nursing assessments. Excess risk associated with any value of a variable is defined as percent absolute increase in 1-year mortality relative to minimum 1-year mortality identified for that variable. Excess risk is summed on a linear scale to reflect cumulative risk for individual patients at any given time. RI was computed at every

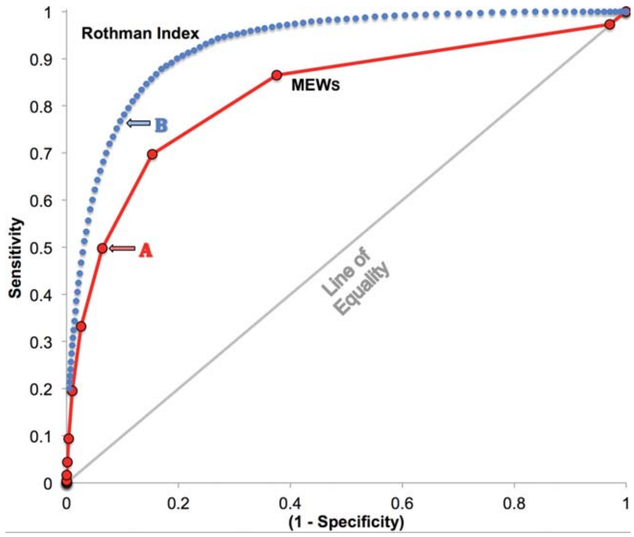

FIG. 1. Modified Early Warning Score (MEWS) and Rothman Index (RI). Shown are receiver operating characteristic curves for 24-hour hospital mortality of general medical-surgical unit patients (N = 32,472); area under the curve is MEWS = 0.82, RI = 0.93. (A) An alarm at MEWS = 4 corresponds to the cut point of RI = 16 for similar sensitivity (49.8%, 48.9%), resulting in 1 true positive for 18 false positives by MEWS, and for 8 false positives by RI. (B) Cut point at RI = 30 provides a positive predictive value (PPV) similar to MEWS = 4; these points of PPV (5.3%, 5.2%) result in 49% sensitivity by MEWS and 77% sensitivity by RI.

new observation during a patient visit, when input values were available. Laboratory results are included when measured, but after 24 hours their weighting is reduced by 50%, and after 48 hours they are excluded. Data input intervals were a function of institutional patient care protocols and physician orders. All observations during a patient's stay were included in the analysis, per the method of Prytherch et al.[4] Because data did not contain the simplified alert/voice/pain/unresponsive (A/V/P/U) score, computation of MEWS used appropriate mapping of the Glasgow Coma Scale.[10] A corresponding MEWS was calculated for each RI. The relationship between RI and MEWS is inverse. RI ranges from −91 to 100, with lower scores indicating increasing acuity. MEWS ranges from 0 to 14, with higher scores indicating increasing acuity.

Outcome Ascertainment

In-hospital death was determined by merging the date and time of discharge with clinical inputs from the hospital's EMR. Data points were judged to be within 24 hours of death if the timestamp of the data point collection was within 24 hours of the discharge time with "expired" as the discharge disposition.

Statistical Methods

Demographics and input variables from the 2 groups of observations, those who were within 24 hours of death and those who were not, were compared using a t test with a Cochran and Cox[11] approximation of the probability level of the approximate t statistic for unequal variances. Mean, standard deviation, and P values are reported. Discrimination of RI and MEWS to predict 24-hour mortality was estimated using area under the receiver operating characteristic (ROC) curve (AUC), and null hypothesis was tested using χ^2. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive and negative likelihood ratios (LR+, LR−) were computed. Analyses were performed with SAS 9.3 (procedures ttest, freq, logistic, nlmixed; SAS Institute, Cary, NC). Typically MEWS = 4 triggers a protocol to increase level of assessment and/or care, often a transfer to the intensive care unit (ICU). We denoted the point on ROC curve where MEWS = 4 and identified an RI point of similar LR− and sensitivity to compare false alarm rate. Then we identified an RI point of similar LR+ for comparison of LR− and sensitivity.

TABLE 3. Accuracy of the Modified Early Warning Score Versus the Rothman Index to Predict 24-Hour Mortality (N = 1,794,910)

Cut Points	MEWS = 4	RI = 16*	MEWS = 4	RI = 30†
Likelihood ratio, positive	7.8	16.9	7.8‡	7.9‡
Likelihood ratio, negative	0.54‡	0.53‡	0.54	0.26
Sensitivity	49.8%	48.9%	49.8%	76.8%
Specificity	93.6%	97.1%	93.6%	90.4%
Positive predictive value	5.2%	10.6%	5.2%	5.3%
Negative predictive value	99.6%	99.6%	99.6%	99.8%

NOTE: An alarm at MEWS = 4 corresponds to a cut point of RI = 16 at similar LR− (and similar sensitivity) and to a cut point of RI = 30 at similar LR+ (and similar positive predictive value). Dataset contained 1,794,910 observations of 32,472 patients. Of the patients, 98.1% survived (n = 31,855; mean age, 65.0 years; SD = 18.6 years) and 1.9% died (n = 617; mean age, 75.7 years; SD = 13.9 years). **Abbreviations:** CI, confidence interval; LR, likelihood ratio; MEWS, Modified Early Warning Score; RI, Rothman Index; SD, standard deviation.

*LRs $P < 0.0001$ for all individual points. LR+ in first pair of columns is significantly different (95% CI: 7.68-7.97; 16.6–17.3), whereas the LR− is virtually the same (95% CI: 0.528-0.546; 0.517-0.535).

†LR− in second pair of columns is significantly different (95% CI: 0.528-0.546; 0.517-0.535), while the LR+ is virtually the same (95% CI: 7.68-7.97; 7.90-8.07).

‡LRs were used to select the nearest RI cut point for performance comparisons with MEWS at the times when an alarm was being triggered.

RESULTS

A total of 1,794,910 observations during 32,472 patient visits were included; 617 patients died (1.9%). Physiological characteristics for all input variables used by RI or MEWS are shown in Table 2, comparing observations taken within 24 hours of death to all other observations.

RI versus MEWS demonstrated superior discrimination of 24-hour mortality (AUC was 0.93 [95% confidence interval {CI}: 0.92-0.93] vs 0.82 [95% CI: 0.82-0.83]; difference, 0.11 [95% CI: 0.10-0.11]; $P < 0.0001$). ROC curves for RI and MEWS are shown in Figure 1; the MEWS is subsumed by RI across the entire range. Further, paired comparisons at points of clinical importance are presented in Table 3 for LR+, LR−, sensitivity, specificity, PPV, and NPV. In the first pair of columns, MEWS = 4 (typical trigger point for alarms) is matched to RI using sensitivity or LR−; the corresponding point is RI = 16, which generates twice the LR+ and reduces false alarms by 53%. In the second pair of columns, MEWS = 4 is matched to RI using PPV or LR+; the corresponding point is RI = 30, which captures 54% more of those patients who will die within 24 hours.

DISCUSSION

We have shown that a general acuity metric (RI) computed using data routinely entered into an EMR outperforms MEWS in identifying hospitalized patients likely to die within 24 hours. At similar sensitivity, RI yields an LR+ more than 2-fold greater, at a value often considered conclusive. MEWS is derived using 4 vital signs and a neurologic assessment. Such a focus on vital signs may limit responsiveness to changes in acuity, especially during early clinical deterioration. Indeed, threshold breach tools may inadvertently induce a false sense of an individual patient's condition and safety.[12] The present findings suggest the performance of RI over MEWS may be due to inclusion of nursing assessments, laboratory test results, and heart rhythm. Relative contributions of each category are: vital signs (35%), nursing assessments (34%), and laboratory test results (31%). We found in previous work that failed nursing assessments strongly correlate with mortality,[13] as illustrated in Table 2 by sharp differences between patients dying within 24 hours and those who did not.

Sensitivity to detect early deterioration, especially when not evidenced by compromised vital signs, is crucial for acuity vigilance and preemptive interventions. Others[14] have demonstrated that our approach to longitudinal modeling of the acuity continuum is well positioned to investigate clinical pathophysiology preceding adverse events and to identify actionable trends in patients at high risk of complications and sepsis after colorectal operations. Future research may reveal both clinical and administrative advantages to having this real-time acuity measure available for all patients during the entire hospital visit, with efficacy in applications beyond use as a trigger for EWS alarms.

Study limitations include retrospective design, single-center cohort, no exclusion of "expected" hospital deaths, and EMR requirement. For MEWS, the Glasgow Coma Scale was mapped to A/V/P/U, which does not appear to affect results, as our c-statistic is identical to the literature.[4] Any hospital with an EMR collects the data necessary for computation of RI values. The RI algorithms are available in software compatible with systems from numerous EMR manufacturers (eg, Epic, Cerner, McKesson, Siemens, AllScripts, Phillips).

The advent of the EMR in hospitals marries well with an EWS that leverages from additional data more information than is contained in vital signs, permitting complex numeric computations of acuity scores, a process simply not possible with paper systems. Further, the automatic recalculation of the score reduces the burden on clinicians, and broadens potential use over a wide range, from minute-by-minute recalculations when attached to sensors in the ICU, to comparative metrics of hospital performance, to nonclinical financial resource applications. This new information technology is guiding methods to achieve a significant performance increment over current EWS and may assist earlier detection of deterioration, providing a chance to avoid medical crises.[15]

Acknowledgements

The authors express their appreciation to Abington Memorial Hospital. Particular thanks are extended to Steven I. Rothman, MSEM, for extensive discussions and technical support. The authors thank Alan Solinger, PhD, for his assistance in reviewing the manuscript.

Disclosures: One author (RAS) declares no conflict of interest. Two authors (GDF, MJR) are employees and shareholders in PeraHealth, Inc. of Charlotte, North Carolina, a health information technology company that offers products utilizing the Rothman Index. All of the original research defining the Rothman Index was performed prior to the formation of the company and is now published in peer-reviewed journals. The index is freely available to all qualified researchers and is currently installed at several major medical research centers and hospital systems. This present work is under the auspices and partly funded by an independent foundation, F.A.R. Institute of Sarasota, Florida. Early research defining the Rothman Index was funded by grants from Sarasota Memorial Healthcare Foundation and the Goldsmith Fund of Greenfield Foundation. Continuing research has been funded by the F.A.R. Institute.

References

1. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. *QJM Mon J Assoc Physicians*. 2001;94:521–526.
2. Kyriacos U, Jelsma J, Jordan S. Monitoring vital signs using early warning scoring systems: a review of the literature. *J Nurs Manag*. 2011;19:311–330.
3. Kirkland LL, Malinchoc M, O'Byrne M, et al. A clinical deterioration prediction tool for internal medicine patients. *Am J Med Qual*. 2013; 28:135–142.
4. 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.
5. 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.
6. Alvarez CA, Clark CA, Zhang S, et al. Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak*. 2013;13:28.
7. Bergstrom N, Braden BJ, Laguzza A, Holman V. The Braden Scale for predicting pressure sore risk. *Nurs Res*. 1987;36:205–210.
8. Rothman MJ, Rothman SI, Beals J IV. Development and validation of a continuous measure of patient condition using the electronic medical record. *J Biomed Inform*. 2013;46:837–848.
9. Bradley EH, Yakusheva O, Horwitz LI, Sipsma H, Fletcher J. Identifying patients at increased risk for unplanned readmission. *Med Care*. 2013;51:761–766.
10. Kelly CA, Upex A, Bateman DN. Comparison of consciousness level assessment in the poisoned patient using the alert/verbal/painful/unresponsive scale and the Glasgow Coma Scale. *Ann Emerg Med*. 2004; 44:108–113.
11. Cochran W, Cox, GM. *Experimental Design*. New York, NY: John Wiley & Sons; 1950.
12. Lynn LA, Curry JP. Patterns of unexpected in-hospital deaths: a root cause analysis. *Patient Saf Surg*. 2011;5:3.
13. Rothman MJ, Solinger AB, Rothman SI, Finlay GD. Clinical implications and validity of nursing assessments: a longitudinal measure of patient condition from analysis of the Electronic Medical Record. *BMJ Open*. 2012;2(4):pii: e000646.
14. Tepas JJ III, Rimar JM, Hsiao AL, Nussbaum MS. Automated analysis of electronic medical record data reflects the pathophysiology of operative complications. *Surgery*. 2013;154:918–926.
15. Subbe CP, Thorpe CM, Hancock C. Not getting better means getting worse—trends in Early Warning Scores suggest that there might only be a short time span to rescue those threatening to fall off a "physiological" cliff? *Resuscitation*. 2013;84:409–410.