ORIGINAL RESEARCH

Performance of the Maximum Modified Early Warning Score to Predict the Need for Higher Care Utilization Among Admitted Emergency Department Patients

Corey R. Heitz, MD[1]
John P. Gaillard, MD[1]
Howard Blumstein, MD[1]
Doug Case, PhD[2]
Catherine Messick, MD[3]
Chadwick D. Miller, MD[1]

[1] Department of Emergency Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina.

[2] Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina.

[3] Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina.

BACKGROUND: It is uncertain whether ED-calculated risk scores can predict required intensity of care upon hospital admission. This investigation examines whether versions of the Modified Early Warning Score (MEWS) predict high level of care utilization among patients admitted from the ED.

METHODS: A retrospective chart review of 299 admissions was implemented. Exclusions prior to abstraction included pediatrics, cardiology, or trauma admissions. Using a data-gathering instrument, abstractors recorded physiologic parameters and clinical variables. Risk scores were calculated electronically. In contrast to the original MEWS, the MEWS Max was calculated using data from the entire ED visit. The primary outcome composite included all-cause mortality and higher care utilization within 24 hours.

RESULTS: The final analysis contained 280 participants. 76 (27%) met the composite endpoint of death (n = 1) or higher care utilization (n = 76). The MEWS Max was associated with the composite outcome (OR=l.6 [95% CI 1.3-1.8] for each one point increase). The MEWS Max had moderate predictive ability (C statistic: MEWS Max 0.73 [0.66-0.79]) but classified 82% of participants as intermediate (10-40%) risk. Inclusion of additional variables slightly improved the predictive ability (C statistic 0.76 [0.69–0.82]) and correctly reclassified 17% of patients as <10% risk.

CONCLUSIONS: The MEWS Max has moderate ability to predict the need for higher level of care. Addition of ED length of stay and other variables to MEWS Max may identify patients at both low and high risk of requiring a higher level of care. *Journal of Hospital Medicine* 2010;5:E46–E52. © 2010 Society of Hospital Medicine.

KEYWORDS: care standardization, early warning scores, emergency department, level of care, patient safety, risk scores.

Matching the severity of illness to the appropriate intensity of care is important for the effective delivery of medical care. Overtriage to critical care units results in unnecessary resource consumption. Undertriage to the wards may result in worsening of physiologic parameters[1,2] that often go unnoticed or unaddressed for more than 24 hours.[3] Therefore, it is important for emergency department (ED) admission decisions to be accurate with respect to the level of care. Because of the importance of this decision, objective criteria to aid in this decision process, if accurate, would improve medical care delivery.

Physiologic measurements and procedural interventions appear to predict the need for a higher level of care among inpatients.[2,4,5] This knowledge has led to the development of tools meant to identify inpatients on general wards who are at risk for deterioration. Such tools for identification of inpatients at risk generally use "single threshold" models triggered by a single abnormal physiologic value, or models that combine multiple parameters into a summative score.[6,7] The performance of previously described risk stratification tools has generally been to exhibit high sensitivity at the sacrifice of low specificity and discriminatory value.[8]

The value of these models as they apply in the emergency department is less well characterized. Because derangements in physiologic parameters are common among ED patients, one might expect that "single-threshold" systems would exhibit high sensitivity at the expense of specificity when applied to this population. In contrast, a summative risk score may be better suited for the complexities of illness in undifferentiated ED patients and offer

better discriminatory value in this population. Summative scoring systems have been shown to retain a higher specificity as the score increases compared to single-threshold systems.[8]

The Modified Early Warning Score (MEWS)[9] is a predictive tool for higher level of care that has been tested in the ED setting. This tool produces a summative score using temperature, respiratory rate, heart rate, level of consciousness, and systolic blood pressure. In a single-site study from the United Kingdom, MEWS, when calculated at the time of ED presentation, did not improve decision making over a commonly used triage system, exhibiting inadequate sensitivity in identifying patients who would be admitted to the intensive care unit (ICU).[10] However, as a result of the care delivered in the ED, patients' conditions can change significantly throughout their stay. Therefore we postulate that the MEWS calculated at a single time in the ED (eg, at the time of admission) is not the most accurate predictor of care intensity requirements.

The primary objective of this research was to add to the literature provided by Subbe et al.[10] by describing the performance characteristics and discriminatory ability of the most abnormal MEWS (MEWS Max) score during the entire ED stay in predicting the need for higher levels of care among ED patients presenting to a tertiary care facility in North America.

Patients and Methods
Study Design
To determine the performance characteristics of the MEWS in ED patients, we used a structured explicit retrospective chart review on a random sample of ED patients being admitted to the hospital.

Study Setting
The study was conducted at 1 tertiary care academic medical center in the United States, consisting of 830 beds, approximately 125 of which provide a higher level of care, defined as intensive care, intermediate care, or acute care. The ED volume in 2005 was 75,000 with an admission rate of 20%. In the ED, patients are primarily seen by residents who are supervised by board-certified or board-eligible emergency medicine attendings.

Study Population
All patients presenting to the ED of Wake Forest University Baptist Medical Center in 2005 were considered for inclusion. From these patients, a listing was created of all hospital admissions through the ED in 2005. Because trauma and cardiology patients have disease-specific risk stratification tools that are used to guide admission,[11,12] they were then removed from this list and excluded. Additionally, pediatric patients were excluded from this listing as the MEWS score relies on vital sign abnormalities, which have varying ranges of normal in children. From this list, 500 charts were ran-

domly selected for further review. Additional criteria were applied at the time the charts were reviewed to exclude those: without an ED record matching the date of admission, without 1 complete set of ED vitals, receiving mechanical ventilation at the time of presentation, or patients currently receiving hospice or "comfort care." Charts from the list of 500 were reviewed sequentially until the goal number of charts had been completed. The number of charts reviewed was selected to allow relatively precise 95% confidence intervals [CIs] around sensitivity (±10%) based on the assumptions of 80% sensitivity and a 20% incidence of the primary outcome. Based on this, the intent was to abstract information from 300 patient charts.

Study Protocol
A standardized data abstraction template was created. Data abstractors included 2 physicians and 2 nurses. Group training for the abstractors was provided by the primary investigator and included performance review and feedback until competence was demonstrated. Data abstractors used the paper copy of the ED nursing notes (and physician notes if clarification required) to abstract data from the medical record. Abstractors were not aware of the patient's outcome at the time of data abstraction as this information was contained in a separate database. During the chart review, and blinded to the abstractors, 25 charts were selected for abstraction by all data abstractors to allow calculation of interobserver agreement.

Clinical outcomes were determined by referencing hospital databases and the medical record if clarification was needed. The admission bed location and changes in patient location throughout the hospital stay were used to track the need for a higher level of care. The outcome of death was determined by cross-referencing study participants with hospital mortality data, and the medical record, if needed.

Predictor Score Calculation
Abstracted data were used to calculate the MEWS score according to the criteria specified in Table 1 at the initial ED presentation (MEWS Initial), the maximum during the ED stay (MEWS Max), and prior to admission (MEWS Admit). Parameters not repeated after arrival were carried forward from the most recent recording. An adaptation of the MEWS score was required by replacing the alert/verbal/painful/unresponsive (AVPU) scale to determine the level of consciousness with the Glasgow Coma Scale (GCS), a conversion that has been previously described.[13,14]

Clinical Endpoint Definitions and Outcomes
"Need for higher level of care" was defined as initial admission from the ED or transfer within 24 hours to a nonfloor bed (acute care, intermediate care unit, or critical care unit). Acute care beds at the study hospital have a lower bed-to-nurse ratio and more intensive monitoring (beside vs. radio-telemetry, vitals signs every 2 hours compared to every 4

TABLE 1. Adaptation of the MEWS

	3 Points	2 Points	1 Point	0 Points	1 Point	2 Points	3 Points
Systolic blood pressure	<70	71-80	81-100	101-199		≥200	
Heart rate		<40	41-50	51-100	101-110	111-129	≥130
Respiratory rate		<9		9-14	15-20	21-29	≥30
Temperature		<95		95-101.1		≥101.2	
GCS				15	11-14	7-10	≤6

Abbreviations: GCS, Glascow Coma Scale; MEWS, Modified Early Warning Score.

hours) than floor beds. Intermediate care beds fulfill a gap between these and critical care, with dedicated respiratory therapists, the ability for invasive monitoring, and ventilator management. In addition, the hospital's burn, bone marrow transplant (BMT), and cardiac care units (CCU) are intensive care–level units, and were included when measuring the need for higher level of care. Mortality was defined as death during the index hospitalization. The primary outcome was the composite "need for a higher level of care" or mortality within 24 hours of ED presentation.

Data Analysis
Calculation of interobserver agreement for data obtained from the chart abstraction was performed using Kappa coefficients. Descriptive statistics were used to summarize the patient characteristics separately for those who did and did not need higher levels of care. Fisher exact tests and Wilcoxon rank-sum tests were used to assess group differences in the categorical and continuous patient characteristics, respectively. A frequency table was used to display the cross-tabulation of MEWS Max scores with the need for higher levels of care, and the sensitivity and specificity were calculated for each cutpoint of the predictor scores. These measurements were plotted against one another in receiver-operating characteristic (ROC) curves and the optimal cutpoint chosen as the one that gave the greatest sum of sensitivity and specificity. The area under the ROC curves and approximate 95% CIs were calculated. The Cochran-Armitage trend test[15–17] was used to assess the association between risk score and outcome. Logistic regression was used to model the log odds of needing higher levels of care as a function of the MEWS Max score. Calibration of the model was assessed by analyzing the performance of the MEWS Max score among patient subgroups and comparing observed and expected events. Performance was also assessed among sextiles of risk using the Hosmer and Lemshow[18] goodness-of-fit test.

As a secondary objective, additional covariates were added to the logistic model including MEWS to see if model performance could be improved. First, a simple logistic regression was used to determine the most significant MEWS score measurements among the 3 that were measured (MEWS Initial, MEWS Max, and MEWS Admit). Only 1

MEWS measurement was considered for the final model to avoid colinearity. The selected MEWS measurement was then entered into a multivariable logistic model along with age ≥60 years, gender, race/ethnicity (white, black, Hispanic, other), method of arrival (ambulatory or by ambulance), ED length of stay (recorded to the nearest minute, then converted to hours at the second significant digit), intravenous (IV) antibiotics in the ED, and antibiotics prior to ED arrival. Candidate variables were chosen considering both the plausibility to be associated with the outcome and the reliability of the data elements considering our retrospective methods. Forward selection, stepwise selection, and backward elimination with a significance level of 0.20 to enter and/or stay in the model were used to obtain a predictive model.

In order to assess the risk stratification potential for the MEWS Max model and the exploratory model (MEWS Plus), the ability to classify subjects by their probability of experiencing the outcome was assessed. Because an established consensus does not exist in the literature for these cutoffs, it was hypothesized that 4 risk categories (0-10%, >10-40%, >40-70%, and >70%) would be clinically useful to clinicians allowing categorization into low-, intermediate-, high-, and very-high-risk-groups for requiring a higher level of care.

Results
Complete chart abstraction was performed for 299 patient encounters. After abstraction, 19 charts were excluded from final analysis due to missing outcome data (n = 6) or implausible and/or missing crucial data values (n = 13). Pairwise kappa values for abstraction of the MEWS Max score demonstrated agreement ranging from good to very good (0.67-0.88). Of the 280 analyzed encounters, 76 (27%) met the primary composite outcome of death (n = 1) or need for higher care (n = 76). Of these 76 patients, 69 were admitted from the ED to a high level of care, and 7 were initially admitted to a lower level of care and required transfer to a higher level of care within 24 hours. Thirty-seven patients requiring a higher level of care were admitted to an ICU (ICU = 31; BMT, CCU, and burn unit with 2 patients each), 9 to intermediate care, and 30 to an acute care bed.

Demographics and presenting characteristics from the study participants can be seen in Table 2. The mean age of
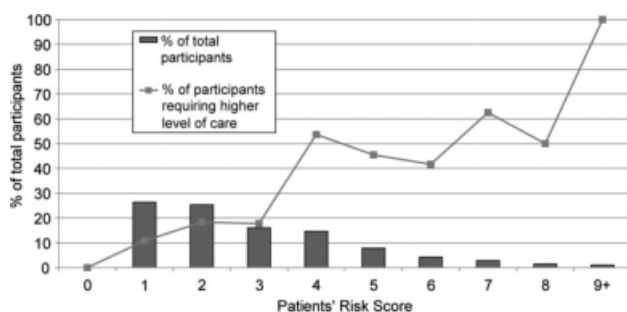
TABLE 2. Characteristics of Study Participants

Patient Characteristics	Composite Endpoint Not Met (n = 204)*	Composite Endpoint Met (n = 76)*	P Value
Demographics			
Age (years)[†]	56 (42, 73)	55 (41, 71)	0.66
Female sex (%)	51	43	0.28
White race (%)	65	63	0.91
Arrival via ambulance (%)	43	63	0.004
Length of stay (hours)[†]	5.8 (4.6, 7.2)	5.5 (4.3, 6.9)	0.15
Presenting characteristics			
Systolic BP (mmHg)[†]	132 (117, 148)	135 (118, 159)	0.26
Heart rate (beats/minute)[†]	87 (74, 100)	96 (82, 111)	0.003
Respiratory rate (breaths/minute)[†]	20 (18, 22)	20 (18, 24)	0.26
Temperature (degrees F)[†]	97.9 (97.1, 98.8)	97.8 (96.8, 99.6)	0.78
Glasgow coma scale[†]	15 (15, 15)	15 (14, 15)	<0.001
Antibiotic therapy			
On antibiotics at arrival (%)	9	9	1.00
IV antibiotics in the ED (%)	31	34	0.67

NOTE: Fisher's exact test used for categorical comparisons; Wilcoxon rank-sum test used for continuous variable comparisons.

Abbreviations: BP, blood pressure; ED, Emergency Department; IV, intravenous.

*n reflects number of participants per group. Values for each variable shown for nonmissing data; all categories have <4% missing data elements.
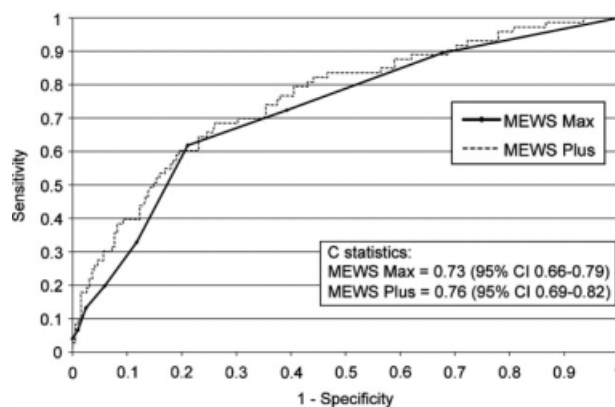
[†]Data presented as median (first quartile, third quartile).

FIGURE 1. MEWS Max distribution and outcome rates.

FIGURE 2. ROC curves for MEWS Max and MEWS Plus.

TABLE 3. Performance Characteristics at Each Scoring System Cutoff

MEWS Max Cutoff	Number at or Above the Cutoff Needing a Higher Level of Care	Sensitivity % (95% CI)	Specificity % (95% CI)	Positive Predictive Value (%)	Negative Predictive Value (%)
≥1	76	100 (95-100)	0 (0-2)	27	NA
≥2	68	89 (80-95)	32 (26-39)	33	89
≥3	55	72 (61-82)	61 (54-68)	41	86
≥4	47	62 (50-73)	79 (73-84)	52	85
≥5	25	33 (23-45)	88 (83-92)	51	78
≥6	15	20 (11-30)	94 (90-97)	56	76
≥7	10	13 (6-23)	98 (94-99)	67	75
≥8	5	7 (2-15)	99 (97-100)	71	74
≥9	3	4 (1-11)	100 (98-100)	100	74

Abbreviations: CI, confidence interval, MEWS, Modified Early Warning Score, NA, not available.

participants was 56 years and was similar for the 2 groups. Approximately one-half of the study participants were female (49%) and there was no statistical association between experiencing the composite outcome and gender (P = 0.28). The majority (64%) of participants were Caucasian, followed by African American (33%) and Hispanic or "other" (2%). Similar distributions were seen when stratified by outcome. Vital signs of the participants in total and stratified by outcome fell within normal parameters. ED length of stay was similar among those meeting and not meeting the composite outcome (5.5 hours vs. 5.8 hours, P = 0.15). Patients who met the composite outcome were more likely to have arrived by ambulance (63% vs. 43%, P = 0.004).

The distribution of scores and the proportion of participants with each score that met the composite outcome are shown in Figure 1. The MEWS Max was significantly associated with the primary composite outcome (P < 0.001, Cochran-Armitage trend test). The scoring system demonstrates an increase in the proportion of participants meeting the composite endpoint as the score increases, and all participants with a MEWS Max score ≥9 met the composite outcome.

ROC are shown in Figure 2. The optimum threshold for MEWS Max based on the sum of sensitivity and specificity is ≥4, associated with a sensitivity of 62% and a specificity of 79% (Table 3) The predictive ability of the MEWS Max was moderate (C statistic MEWS Max 0.73; 95% CI, 0.66-0.79), with each 1-point increase in the MEWS Max score associated with a 60% increase in the odds of meeting the composite endpoint (odds ratio [OR], 1.6; 95% CI, 1.3-1.8).

Table 4 shows calibration of the model using different subgroups of the patient population. Grouping patients by age or gender did not reveal a higher event rate in any particular group. Using the Hosmer and Lemeshow[18] goodness-of-fit test to stratify by risk category, no evidence for lack of fit was found (P = 0.06).

In the exploratory analysis, 267 subjects had complete data for all candidate variables. Simple logistic regression revealed that the most predictive MEWS measurement was the MEWS Max (C statistic MEWS Max 0.725, MEWS Initial 0.668, MEWS Admit 0.653). Stepwise selection, forward selection, and backward elimination produced the same model containing method of arrival ($P = 0.03$), MEWS Max ($P < 0.001$), IV antibiotics in the ED ($P = 0.17$), length of stay ($P = 0.05$), and gender ($P = 0.12$). In the subset of subjects with these complete data elements (n = 268), the inclusion of the additional measures increased the C statistic to 0.76 (95% CI, 0.69-0.82), a 0.04 increase over the model that only included MEWS Max in the same subset of subjects.

MEWS Max resulted in no patients being classified as low-risk, with the majority (81.7%) classified as intermediate-risk, 15.7% classified as high-risk, and 2.6% classified as very high risk (Table 5). In all categories the actual event rate fell within the predicted event rate interval. The addition of variables included in MEWS Plus resulted in 14.6% of patients being classified as low-risk, 64.0% as intermediate risk, 17.2% with high-risk, and 4.1% as very-high-risk. In 58 cases (21.7%), using MEWS Plus would have placed patients in a more appropriate risk category than that assigned by MEWS Max; ie, a lower risk category for those who did not have events, and a higher risk category for those experiencing events. The majority of this correct reclassification was seen in the intermediate risk group by MEWS Max, where 17.6% were appropriately reclassified. Alternatively, 5.6% of cases would have resulted in inappropriate reclassification. Again, the actual event rate fell within the boundaries of predicted risk in all cases.

TABLE 4. Performance of the MEWS Max by Patient Subgroup

Characteristic	Total Participants	Observed Events	Expected Events	Observed/ Expected
Age				
>45 years	87	24	23.8	1.01
45-70 years	118	33	32.8	1.01
>70 years	75	19	19.4	0.98
Sex*				
Male	142	43	37.4	1.15
Female	137	33	38.5	0.86
Sextile of risk with MEWS Max[†]				
First	27	15	17.8	0.84
Second	22	10	10.1	0.99
Third	41	22	14.4	1.53
Fourth	45	8	11.6	0.69
Fifth	71	13	12.9	1.01
Sixth	74	8	9.2	0.87

Abbreviation: MEWS, Modified Early Warning Score.
*Sex unknown for 1 subject.
[†]Hosmer and Lemeshow goodness of fit; P value = 0.06.

TABLE 5. Change in Risk Categorization by Using the MEWS Plus Compared to the MEWS Max

MEWS Max	MEWS Plus*					Events	
	0-10[†]	>10-40	>40-70	>70	Row Totals (%)	Correctly Reclassified (%)	Incorrectly Reclassified (%)
Total (%)	39 (14.6)	171 (64.0)	46 (17.2)	11 (4.1)	267	58 (21.7)	15 (5.6)
Events[‡] (% of total)	2 (5.1)	39 (22.8)	24 (52.2)	8 (72.7)	73 (27.3)		
Nonevents[§] (% of total)	37 (94.9)	132 (77.2)	22 (47.8)	3 (27.3)	194 (72.7)		
0-10	0	0	0	0	0	0	0
Events	0	0	0	0	0		
Nonevents	0	0	0	0	0		
>10-40	39	162	17	0	218 (81.7)	47 (17.6)	9 (3.4)
Events	2	36	10	0	48 (22.0)		
Nonevents	37	126	7	0	170 (78.0)		
>40-70	0	9	27	6	42 (15.7)	10 (3.7)	5 (1.9)
Events	0	3	13	4	20 (47.6)		
Nonevents	0	6	14	2	22 (52.4)		
>70	0	0	2	5	7 (2.6)	1 (0.4)	1 (0.4)
Events	0	0	1	4	5 (71.4)		
Nonevents	0	0	1	1	2 (28.6)		

NOTE: The total number (n = 267) includes subjects with complete candidate variable data for model construction.
Abbreviations: ED, Emergency Department; MEWS, Modified Early Warning Score.
*Variables included in MEWS Plus are age ≥60, race, gender, ED length of stay, method of arrival, and antibiotics given prior to or during ED visit.
[†]Percentage risk of deterioration.
[‡]Events: met composite outcome.
[§]Nonevents: did not meet composite outcome.

Discussion

Matching the initial level of care to the patient's severity of illness can be expected to improve the efficiency of health care delivery. The MEWS is a simple prediction instrument that can be calculated at the bedside and would be ideal for this purpose. The MEWS has good predictive ability among patients on the wards or awaiting admission,[9,10] and in this investigation a variation of MEWS appears to have potential to discriminate among high-risk and low-risk ED patients.

Examination of the ROC curve for the MEWS Max score demonstrates a fair performance (C statistic = 0.73). In this analysis, we created low-risk, intermediate-risk, high-risk, and very-high-risk groups. The strength of the MEWS Max rests in its ability to classify patients as high-risk or very-high-risk. Approximately 16% of patients are classified by MEWS Max as high-risk, and 3% as very-high-risk, making the practitioner more confident in the decision to admit to a high level of care. However, MEWS Max classifies no patients as low risk and approximately 80% of patients are classified as intermediate-risk. The majority of patients being classified into this "gray zone" and the inability to classify patients as low-risk significantly limits the utility of MEWS Max.

In exploratory analysis, these data propose a model using additional readily available parameters that when added to the MEWS Max can improve patient classification. Of particular interest is the ability of the MEWS Plus model to more accurately identify patients at low risk of requiring a higher level of care. When compared to MEWS Max, approximately 22% of patients were correctly reclassified by MEWS Plus, with only 5% incorrectly reclassified. Importantly, MEWS Plus is able to reduce the size of the intermediate-risk group, predominantly by reclassifying patients as low risk. Forty-seven (17.6%) of the patients previously categorized as intermediate risk with MEWS Max were reclassified, with 39 of them becoming low risk, 2 (5.1%) of whom had events. However, the major limitation of the MEWS Plus is that it is currently not able to be calculated at the bedside as many of the included variables are time dependent. More analysis is needed to validate precisely which variables are most important, determine how they add to the calculation, and understand when or how often during the ED visit risk should be calculated. Further exploration and validation of this model is necessary.

The results of this investigation add in important ways to a previous study of the MEWS in ED patient triage.[10] Subbe et al.[10] examined the ability of the MEWS to improve admission decisions beyond those recommended by the Manchester Triage System. Their investigation was conducted among 153 ED patients who belonged to 1 of 3 cohorts being admitted from the ED in the United Kingdom. They concluded that the MEWS was unable to significantly improve admission level of care decisions over the Manchester Triage System. Our investigation differs from that reported by Subbe et al.[10] in several important ways. Methodologically, we chose to include a broad population of ED patients rather than selecting 3 cohorts for comparison, and excluded trauma and cardiology patients due to suspected differences in admission patterns in these patients. Further, we conducted our analysis using the maximum MEWS score obtained during a patient's encounter. We felt that using the maximum MEWS score takes full of advantage of all clinical data obtained during the patient's ED visit rather than relying on their severity of illness when the patient first arrives. Additionally, we selected an outcome measure that was determined at 24 hours because we feel events occurring within 24 hours of admission are more likely to reflect a progression of a disease process present at the time of the ED evaluation. Subbe et al.[10] analyzed ICU admissions after any duration of hospitalization on the wards. However, ICU admission after several days of ward care may neither be avoidable, nor predictable, while the patient is in the ED.

Subbe et al.[10] concluded that the MEWS score did not significantly add to triage decisions aided by the Manchester Triage System. However, in their results, a MEWS score >2 would have classified 7 additional patients as high risk out of 50 who required a transfer to a higher level of care when compared to the Manchester Triage System. Our findings explore the discriminatory value of the maximum MEWS score for a patient throughout the ED visit. This approach, combined with our methodologic differences, have led to more encouraging findings about the utility of the MEWS Max score, especially when combined with a few simple and reliably abstracted variables, to predict the required level of care within 24 hours.

Limitations to our results mainly relate to the study design. We chose a nonconcurrent cohort design using an explicit chart review. Chart reviews have inherent limitations that can include inaccuracy of abstracted data elements, missing data, systematic bias imposed by the abstraction process, and unmeasured confounding. To minimize avoidable biases and maintain accuracy while conducting this chart review, we followed well-described methods.[19] However, because we were relying on retrospective data, some data elements were incomplete. For instance, not all participants had multiple sets of vital signs recorded, which could have affected the predictive accuracy of the risk scores. Anticipating this difficulty, we had algorithms established to handle missing data, which we feel minimized this effect. However, despite this effort, 13 patients had to be excluded due to incomplete data. During review, it was noted that some patients admitted due to a traumatic mechanism were included in the final data analysis despite our intent to exclude them. We expect that this was a very small number, and should have had a minimal effect on risk score calculation. In addition, we modified the original MEWS model in that the GCS was used in substitution for the AVPU score. The conversion of the AVPU score to the GCS is well-described and is unlikely to have affected the accuracy of the MEWS. We did not adjust for ED length of stay in our primary MEWS model. It is possible that more severely ill patients were in the ED longer and therefore had more

opportunity to have abnormal vital signs recorded. ED length of stay was incorporated into the MEWS Plus model. Another limitation relates to our reference standard. We chose a composite of the need for a higher level of care or death within 24 hours. The need for higher care is a subjective endpoint. However, we felt this reflection of actual decision making is more informing than comparisons to other objective, unvalidated scoring systems. As more robust scoring systems are developed, researchers will need to consider developing a reference standard employing blinded adjudicators. Pediatric, cardiology, and trauma patients were excluded from this analysis and therefore our results cannot be extrapolated to these populations. MEWS model calibration was performed using the same data set as that on which the model was tested. This may have resulted in overfit of the model to the data, possibly leading to an overstatement of the model's predictive ability. Additionally, the MEWS Plus model requires validation in another study population. A final limitation is that in performing the study at 1 institution, the results may not be generalizable to other settings.

Building on previous work in defining and testing risk scores to predict poor outcomes, we have shown that the MEWS Max is a potentially useful tool to categorize patients as high-risk or very-high-risk for requiring a higher level of care. MEWS Max suffers from the creation of a large intermediate risk group and the inability to classify patients as low-risk. Adding further variables to MEWS Max creates a model with improved performance (MEWS Plus). This model may allow for 15% of admissions to be classified as low-risk and shows promise as a tool to be used in ED triage of patients who are being admitted. Further work should attempt to further refine and validate the MEWS Plus model and examine the effect of implementation of these models on admission decision making and clinical outcomes.

Address for correspondence and reprint requests:
Chadwick D. Miller, Assistant Professor, Department of Emergency Medicine, Wake Forest University School of Medicine,
Winston-Salem, NC 27157; Telephone: 336-716-1740; Fax: 336-716-1705; E-mail: cmiller@wfubmc.edu   Received 23 November 2008; revision received 27 April 2009; accepted 2 May 2009.

References

1. Goldhill DR, McNarry AF, Hadjianastassiou VG, Tekkis PP. The longer patients are in hospital before intensive care admission the higher their mortality. *Intensive Care Med.* 2004;30(10):1908–1913.
2. Goldhill DR, White SA, Sumner A. Physiological values and procedures in the 24 h before ICU admission from the ward. *Anaesthesia.* 1999;54(6): 529–534.
3. Smith AF, Wood J. Can some in-hospital cardio-respiratory arrests be prevented? *A prospective survey. Resuscitation.* 1998;37(3):133–137.
4. Goldhill DR, McNarry AF. Physiological abnormalities in early warning scores are related to mortality in adult inpatients. *Br J Anaesth.* 2004; 92(6):882–884.
5. 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.
6. Foraida MI, DeVita MA, Braithwaite RS, Stuart SA, Brooks MM, Simmons RL. Improving the utilization of medical crisis teams (Condition C) at an urban tertiary care hospital. *J Crit Care.* 2003;18(2):87–94.
7. Lee A, Bishop G, Hillman KM, Daffurn K. The medical emergency team. *Anaesth Intensive Care.* 1995;23(2):183–186.
8. Gao H, McDonnell A, Harrison DA, 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(4): 667–679.
9. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified early warning score in medical admissions. *QJM.* 2001;94(10):521–526.
10. Subbe CP, Slater A, Menon D, Gemmell L. Validation of physiological scoring systems in the accident and emergency department. *Emerg Med J.* 2006;23(11):841–845.
11. Antman EM, Cohen M, Bernink PJLM, et al. The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. *JAMA.* 2000;284(7):835–842.
12. Senkowski CK, McKenney MG. Trauma scoring systems: a review. *J Am Coll Surg.* 1999;189(5):491–503.
13. 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(2): 108–113.
14. McNarry AF, Goldhill DR. Simple bedside assessment of level of consciousness: comparison of two simple assessment scales with the Glasgow Coma scale. *Anaesthesia.* 2004;59(1):34–37.
15. Agresti A, Mehta CR, Patel NR. Exact inference for contingency tables with ordered categories. *J Am Stat Assoc.* 1990(85):453–458.
16. Margolin BH. Tests for trends in proportions. In: Klotz S, Johnson NH, eds. Encyclopedia of Statistical Sciences. Vol 9. New York: John Wiley and Sons, Inc.; 1988:334–336.
17. SAS Institute I. Cochran Armitage Test for Trend. SAS OnlineDoc 9.1.3.Cary, NC: SAS Institute, Inc.; 2007.
18. Hosmer DWJ, Lemeshow S. Applied Logistic Regression. 2nd ed. New York: John Wiley and Sons; 2000.
19. Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: Where are the methods? *Ann Emerg Med.* 1996;27(3):305–308.