TABLE 1. Electrolytes and Osmolarity of Blood Plasma, NaCl 0.9%, and Lactated Ringer

Composition, Properties	Plasma	NaCl 0.9%	Lactated Ringer
Na^+ (mmol/L)	140	154	131
Cl^- (mmol/L)	100	154	112
K^+ (mmol/L)	4.2	–	5.4
Ca^{++} (mmol/L)	1.2	–	1.8
Mg^{++} (mmol/L)	0.9	–	0.5
HCO_3^- (mmol/L)	25	–	28
Osmolarity (mOsm/kg)	290	308	280
Strong ion difference (mEq/L)	24	0	28

Although the findings in this article do not result from a classical prospective randomized large scale trial, which most probably never will be done due to its complexity and costs, the large number of patients analyzed allows for a quite robust conclusion: For fluid replacement and resuscitation use of solutions with a low SID as NaCl 0.9% and D5W should be abandoned. One century after the development of a balanced solution with a SID close to plasma, it is really time to change.

REFERENCES

1. Zampieri FG, Ranzani OT, Azevedo LCP, et al: Lactated Ringer Is Associated With Reduced Mortality and Less Acute Kidney Injury in Critically Ill Patients: A Retrospective Cohort Analysis. *Crit Care Med* 2016; 44:2163–2170
2. Latta T: Saline venous injection in cases of malignant cholera, performed while in the vapour bath. *Lancet* 1832; 19:208–209
3. Hamburger HJ: Koninkl. Akad. v. Wetensch. te Amsterdam, meeting of December 29th. *Amsterdam*, 1883
4. Quilley CP, Lin YS, McGiff JC: Chloride anion concentration as a determinant of renal vascular responsiveness to vasoconstrictor agents. *Br J Pharmacol* 1993; 108:106–110
5. Yunos NM, Bellomo R, Hegarty C, et al: Association between a chloride-liberal vs chloride-restrictive intravenous fluid administration strategy and kidney injury in critically ill adults. *JAMA* 2012; 308:1566–1572
6. Shaw AD, Bagshaw SM, Goldstein SL, et al: Major complications, mortality, and resource utilization after open abdominal surgery: 0.9% saline compared to Plasma-Lyte. *Ann Surg* 2012; 255:821–829
7. Shaw AD, Raghunathan K, Peyerl FW, et al: Association between intravenous chloride load during resuscitation and in-hospital mortality among patients with SIRS. *Intensive Care Med* 2014; 40:1897–1905

Moving Beyond Single-Parameter Early Warning Scores for Rapid Response System Activation*

Matthew M. Churpek, MD, MPH, PhD
Dana P. Edelson, MD, MS
Department of Medicine
University of Chicago
Chicago, IL

Early warning systems provide an assessment of a patient's likelihood of developing critical illness and thus requiring additional critical care resources. The groundwork for these systems was laid millennia ago, with the Hippocratic "Book of Prognostics." The statement attributed to Hippocrates that "it is bad if he has dyspnoea, and urine that is thin and acrid, and if sweats come out about the neck and head" includes clinical variables (respiratory rate and urine output) still used in early warning systems today (1). These systems now form the foundation for activating Rapid Response and Medical Emergency Teams.

Traditionally, early warning systems have come in two primary configurations: single-parameter criteria and

***See also p. 2171.**

Key Words: hospital rapid response team; monitoring, physiologic; quality improvement; vital signs, medical emergency team

Dr. Churpek is supported by a career development award from the National Heart, Lung, and Blood Institute (K08 HL121080) and an American Thoracic Society Foundation Recognition Award for Early Career Investigators. He has received funding (honoraria from Chest for invited speaking engagements), disclosed other support (A patent pending [ARCD. P0535US.P2] for risk stratification algorithms for hospitalized patients), and received support for article research from the National Institutes of Health. Dr. Edelson disclosed other support (A patent pending [ARCD. P0535US.P2] for risk stratification algorithms for hospitalized patients) and received support from ownership interest in QuantHC (Chicago, IL), which is developing products for risk stratification of hospitalized patients. Her institution received funding from grant from Philips Healthcare (Andover, MA) and grant from EarlySense (Tel Aviv, Israel).

DOI: 10.1097/CCM.0000000000002105

aggregated weighted scores (2). The former originated in Australia over 2 decades ago as a set of equally weighted abnormal physiologic thresholds (e.g., respiratory rate > 36), the presence of any of which would trigger the system (3). In contrast, aggregated weighted scoring systems, such as the Modified Early Warning Score (MEWS), which arose in the United Kingdom around the same time, involve summing up points from multiple parameters based on the degree of derangement (e.g., two points for a respiratory rate of 21–29 and three points for ≥ 30) (4, 5).

In this issue of *Critical Care Medicine*, Smith et al (6) provide important evidence regarding the comparative accuracy of the National Early Warning Score (NEWS), an aggregated weighted score similar to the MEWS, which was developed by the Royal College of Physicians as a uniform method of identifying clinical deterioration in patients across the National Health Service (NHS) in the United Kingdom (6). Using data from an NHS District General Hospital, the authors compared NEWS to 44 distinct single-parameter tools and found it to be superior for predicting death, cardiac arrest, and/or unanticipated ICU transfer. This study is limited by the fact that this is a single center study using a population arising from the same hospital in which the VitalPAC Early Warning Score (ViEWS), its immediate precursor, was originally derived (7). However, these concerns are largely mitigated by the fact that there is no overlap between the ViEWS derivation cohort and the current study population, and that the findings are consistent with independent studies demonstrating the superiority of aggregated weighted scoring systems over single-parameter criteria (2, 8).

From a statistical modeling perspective, the finding that an aggregate weighted scoring system is more accurate than single-parameter criteria is not surprising. Single-parameter tools are generally based on single cut-points of continuous variables, which result in the loss of valuable information. For example, respiratory rates of 18 and 30 count similarly if they are both below the activation threshold. Furthermore, these criteria will miss subtle abnormalities in multiple vital signs, which have been shown to be more important for predicting outcomes than more dramatic elevations in a single vital sign (9). Aggregate weighted scores, which include several gradations of derangement and allow high scores to occur from both individual and combinations of vital sign abnormalities, do not suffer from these limitations. The NEWS has the added benefit of being informed by the dataset used to derive the ViEWS, rather than having been developed solely on the basis of expert opinion, upon which the vast majority of single parameter and many commonly used aggregated weighted scores were, including the MEWS. This is evident in the heavier weighting of subtler respiratory rate derangements, for example, which has been shown to be the vital sign with the strongest correlation to clinical deterioration (10, 11). In fact, the use of patient data in its development is the likely rationale for the superiority of NEWS to MEWS in prior head-on comparisons.

However, the improvements in accuracy need not stop there. Additional variables like laboratory data can be added and the full range of values can be utilized with logistic regression models and other similar models (12, 13). The use of vital sign trends can also increase accuracy, although accounting for these is more complicated than initially thought (14). Furthermore, the advent of machine learning tools, such as random forests, enable even more accurate models for predicting clinical deterioration (11).

If one believes that accuracy matters, and any hospital that has ever struggled with false alarms or missed opportunities would be hard pressed to argue that it does not, each hospital system owes it to its providers and patients to implement the most accurate activation tool it can. For those hospitals still using paper charts, that should be one of the aggregated weighted scores, of which the NEWS appears to be one of the stronger contenders. However, for those hospitals that have transitioned to the computer age, it is time to start thinking beyond paper-based screening tools and make our expensive computers and electronic health records (EHRs) do the work they were designed to do. Retrofitting them with less accurate paper-based tools makes little sense.

Although results like the paper by Smith et al (6) suggest that this could and should be the beginning of the end for single-parameter tools, it is becoming clear that sometime in the future we will be saying the same thing about simple aggregated weighted scores, like the NEWS, at least in their current form. EHRs are already ubiquitous in the United States, and are becoming more common in Europe, Australia, and other parts of the world as well. The EHR can harness the promise of "big data," with countless variables and high power computing to automatically calculate complex and accurate algorithms in real-time. The future will belong to comprehensive and complex scores that are more accurate than NEWS, examples of which are already up and running in several hospitals today (13, 15). For hospitals that have already fully transitioned to using EHRs, it is time to make this future a reality. At a minimum, it is time to retire the single-parameter activation criteria once and for all.

REFERENCES

1. Hippocrates: On Regimen in Acute Diseases. Whitefish, MT, Kessinger Publishing Co Printed, 2004
2. Churpek MM, Yuen TC, Edelson DP: Risk stratification of hospitalized patients on the wards. *Chest* 2013; 143:1758–1765
3. Lee A, Bishop G, Hillman KM, et al: The medical emergency team. *Anaesth Intensive Care* 1995; 23:183–186
4. Morgan RM, Williams F, Wright MM: An early warning scoring system for detecting developing critical illness. *Clin Intensive Care* 1997; 8:100
5. Subbe CP, Kruger M, Rutherford P, et al: Validation of a modified Early Warning Score in medical admissions. *QJM* 2001; 94:521–526
6. Smith GB, Prytherch DR, Jarvis S, et al: A Comparison of the Ability of the Physiologic Components of Medical Emergency Team Criteria and the U.K. National Early Warning Score to Discriminate Patients at Risk of a Range of Adverse Clinical Outcomes. *Crit Care Med* 2016; 44:2171–2181
7. Prytherch DR, Smith GB, Schmidt PE, et al: ViEWS–towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010; 81:932–937
8. Tirkkonen J, Olkkola KT, Huhtala H, et al: Medical emergency team activation: Performance of conventional dichotomised criteria

versus national early warning score. *Acta Anaesthesiol Scand* 2014; 58:411–419
9. Jarvis S, Kovacs C, Briggs J, et al: Aggregate National Early Warning Score (NEWS) values are more important than high scores for a single vital signs parameter for discriminating the risk of adverse outcomes. *Resuscitation* 2015; 87:75–80
10. Cuthbertson BH, Boroujerdi M, McKie L, et al: Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient? *Crit Care Med* 2007; 35:402–409
11. Churpek MM, Yuen TC, Winslow C, et al: Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med* 2016; 44:368–374
12. Churpek MM, Yuen TC, Winslow C, et al: Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med* 2014; 190:649–655
13. Escobar GJ, LaGuardia JC, Turk BJ, et al: 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
14. Churpek MM, Adhikari R, Edelson DP: The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation* 2016; 102:1–5
15. Kang MA, Churpek MM, Zadravecz FJ, et al: Real-time risk prediction on the wards: A feasibility study. *Crit Care Med* 2016; 44:1468–1473

Dynamic Systemic Inflammatory Response Syndrome Monitoring: A reSIRSgence?*

R. Phillip Dellinger, MD, MCCM
Cooper Medical School of Rowan University
Cooper University Health Care
Camden, NJ

In this issue of *Critical Care Medicine*, Lindner et al (1) breathe new life into the much maligned systemic inflammatory response syndrome (SIRS) criteria. Using a sophisticated and elegant research design, these authors leverage an automated algorithm embedded in an electronic patient data management system (PDMS) to capture a dynamic minute to minute look at changes in meeting from 0-4 of the SIRS criteria and correlate this with sepsis prediction and diagnosis in a polytrauma cohort. Sepsis diagnosis and onset was defined as time of order to initiate antimicrobials. SIRS prevalence and how that changes over time was monitored using three algorithm-derived SIRS "descriptors" (see below). The use of these dynamic SIRS descriptors is shown to improve specificity of sepsis prediction and diagnosis.

The concept of leveraging an electronic PDMS in attempts to assist in the diagnosis of sepsis and infection is promising (2, 3). Success in this area will require refinement of integrative and predictive systems in order to define the time variance among variables of interest. The dynamic approach taken by Lindner et al (1) is the progress in that direction. Systems such as this may not only improve patient care but could also allow for more valid interfacility comparisons. Real-time electronic surveillance tools leveraging electronic medical records have shown improved diagnostic capability for sepsis and typically have included the SIRS criteria as part of the alert (4–6). The SIRS criteria, despite the poor specificity for diagnosis of sepsis, have been shown to correlate with outcome prediction (7). Large European databases have shown prognostic importance of the SIRS criteria in predicting severity of infection, organ failure, and outcome (8). The SIRS criteria have also been demonstrated to independently predict infection in blunt trauma patients when present at the time of admission (9).

Recognizing the problems with the conventional use of SIRS criteria for diagnosing infection (high sensitivity but lack of specificity), Lindner et al (1) tapped the PDMS to build a "dynamic" outcome probability estimation using the SIRS criteria. The traditional spot check static use of the SIRS criteria was replaced with a computerized algorithm for monitoring the number of SIRS criteria met using intuitive temporal summary measures called "SIRS descriptors." This type of system has the capability to use the PDMS to continuously monitor changes in meeting SIRS criteria over an entire 1,440 minutes of a 24-hour block of time. The three SIRS descriptors used included 1) the average number of SIRS criteria over a time period of interest (0–4), 2) the first to last minute difference in the integration of the number of SIRS criteria (–4 to +4), and 3) the frequency of change in SIRS criteria for all 1,439 minute-to-minute transitions (fluctuation range, 0–1,439). Of note, circulatory support by catecholamines was given equal weight as tachycardia for meeting SIRS criteria. Likewise, respiratory support was allowed to fulfill the tachypnea/hyperventilation SIRS criteria.

Over the first 24 hours of admission, there was a minute-by-minute increase in the mean number of SIRS criteria across the population as a whole. For sepsis prediction, sepsis cases diagnosed at any point in time were compared with controls who had remained sepsis free throughout the ICU stay. Of the 256 polytrauma patients, 86 (33.2%) developed sepsis in the ICU with a median time to diagnosis of 7 days.

This innovative study design analyzed SIRS criteria minute to minute using SIRS descriptors, producing a shift from a

*See also p. 2199.

Key Words: electronic medical records; sepsis; sepsis prediction; systemic inflammatory response syndrome

Dr. Dellinger's institution received consultant fees from Spectral Medical and received grant support from Ferring and Spectral Medical. Dr. Dellinger received consultant fees from PeraHealth and Biotest. He provided expert testimony for Medicolegal, received royalties from Elsevier (textbook), and received support for travel from Alina Health System (Grand Rounds), ACP, and the Society of Critical Care Medicine.

DOI: 10.1097/CCM.0000000000002112

A Comparison of the Ability of the Physiologic Components of Medical Emergency Team Criteria and the U.K. National Early Warning Score to Discriminate Patients at Risk of a Range of Adverse Clinical Outcomes*

Gary B. Smith, FRCA, FRCP[1]; David R. Prytherch, PhD, MIPEM, CSci[2,3]; Stuart Jarvis, PhD[3,4]; Caroline Kovacs, BSc[3]; Paul Meredith, PhD[2]; Paul E. Schmidt, MRCP, BMedSc, MBA[2]; Jim Briggs, BA, DPhil[3]

***See also p. 2283.**

[1]Centre of Postgraduate Medical Research and Education, Faculty of Health and Social Sciences, University of Bournemouth, Bournemouth, UK.

[2]Department of Research & Development, Portsmouth Hospitals NHS Trust, Portsmouth, UK.

[3]Centre for Healthcare Modelling & Informatics, School of Computing, University of Portsmouth, Portsmouth, United Kingdom.

[4]Department of Health Sciences, University of York, York, United Kingdom.

This work was performed at Portsmouth Hospitals NHS Trust & University of Portsmouth.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccmjournal).

Dr. Smith consulted for Laerdal Medical, United Kingdom (Independent contractor to Laerdal Medical, Orpington, Kent on continuous monitoring of vital signs); received grant support from the National Institute for Health Research (NIHR) (grant relating to research into nurse staffing and vital signs observations commenced in April 2015); lectured for the Swedish Medical Association (Honorarium from the Swedish Medical Association for lecture in 2014) and for Huddersfield University, United Kingdom (Honorarium from Huddersfield University re Masterclass on National Early Warning Scores [NEWS]); and received support for travel from The Learning Clinic (TLC). (He was an unpaid research advisor to TLC until May 2016 and has received reimbursement of travel expenses for attending symposia in the UK). He disclosed other relationships: His wife was a shareholder in TLC (manufacturer of VitalPAC, the system used to collect data used in the current research) until October 2015; he acts as an expert advisor to the National Institute for Health and Clinical Excellence during the development of the NICE clinical guideline (50: "Acutely ill patients in hospital: recognition of and response to acute illness in adults in hospital" development group). He was a member of the National Patient Safety Agency committee that wrote the two reports on "Recognizing and responding appropriately to early signs of deterioration in hospitalized patients" and "Safer care for the acutely ill patient: learning from serious incidents." He is a member of the Royal College of Physicians of London's NEWS Development and Implementation Group. His institution received royalties and other support from TLC. (At the time of the research, his previous employers, Portsmouth Hospitals NHS Trust [PHT], had a royalty agreement with TLC to pay for the use of PHT intellectual property within TLC VitalPAC product [VitalPAC is the system used to collect data used in the current research]. He worked for PHT from 1985 to 2011 [until May 31, 2011].) Dr. Prytherch is employed by PHT, received support for travel (he was an unpaid research advisor to TLC until May 2016. In the past they have reimbursed his travel expenses for attending meetings in the UK), and disclosed other relationships (His wife was a minor shareholder in TLC until October 2015 and he assisted Royal College of Physicians [London] in the analysis of data to validate the NEWS.) His institution received other support (at the time of the research, PHT had a royalty agreement with TLC. TLC paid royalties to PHT for IP within their VitalPAC product. VitalPAC was used to collect the vital signs data used in this research), grant support from the Wellcome Trust (grant to start August 2015) and the NIHR (grant started April 2015), and royalties (at the time of his research, PHT had a royalty agreement with TLC. TLC paid royalties to PHT for IP within their VitalPAC product). Dr. Meredith is employed by PHT and received royalties from TLC (at the time of his research, my employer PHT had a royalty agreement with TLC. TLC paid royalties to PHT for IP within their VitalPAC product. VitalPAC was used to collect the vital signs data used in this research). His institution received other support from TLC until October 2015. (At the time of the research, PHT had a royalty agreement TLC. TLC paid royalties to PHT for IP within their VitalPAC product. VitalPAC was used to collect the vital signs data used in this research), grant support from Wellcome Trust (grant to start August 2015) and NIHR (grant started April 2015). Dr. Schmidt received royalties from TLC (at the time of his research, PHT had a royalty agreement with TLC. TLC paid royalties to PHT for IP within their VitalPAC product. VitalPAC was used to collect the vital signs data used in this research). His institution received royalties from TLC (related to development of VitalPAC, the information system used to collect data for this study). Dr. Briggs received support for article research from the Technology Strategy Board. His institution received grant support from TLC and Technology Strategy Board (some of the preliminary work was funded by a Knowledge Transfer Partnership award). The remaining authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: gbsresearch@virginmedia.com

DOI: 10.1097/CCM.0000000000002000

Objective: To compare the ability of medical emergency team criteria and the National Early Warning Score to discriminate cardiac arrest, unanticipated ICU admission and death within 24 hours of a vital signs measurement, and to quantify the associated workload.

Design: Retrospective cohort study.

Setting: A large U.K. National Health Service District General Hospital.

Patients: Adults hospitalized from May 25, 2011, to December 31, 2013.

Interventions: None.

Measurements and Main Results: We applied the National Early Warning Score and 44 sets of medical emergency team criteria to a database of 2,245,778 vital signs sets (103,998 admissions). The National Early Warning Score's performance was assessed using the area under the receiver-operating characteristic curve and compared with sensitivity/specificity for different medical emergency team criteria. Area under the receiver-operating characteristic curve (95% CI) for the National Early Warning Score for the combined outcome (i.e., death, cardiac arrest, or unanticipated ICU admission) was 0.88 (0.88–0.88). A National Early Warning Score value of 7 had sensitivity/specificity values of 44.5% and 97.4%, respectively. For the 44 sets of medical emergency team criteria studied, sensitivity ranged from 19.6% to 71.2% and specificity from 71.5% to 98.5%. For all outcomes, the position of the National Early Warning Score receiver-operating characteristic curve was above and to the left of all medical emergency team criteria points, indicating better discrimination. Similarly, the positions of all medical emergency team criteria points were above and to the left of the National Early Warning Score efficiency curve, indicating higher workloads (trigger rates).

Conclusions: When medical emergency team systems are compared to a National Early Warning Score value of greater than or equal to 7, some medical emergency team systems have a higher sensitivity than National Early Warning Score values of greater than or equal to 7. However, all of these medical emergency team systems have a lower specificity and would generate greater workloads. (*Crit Care Med* 2016; 44:2171–2181)

Key Words: hospital rapid response team; monitoring, physiologic; quality improvement; vital signs, medical emergency team

Staff failures in recognizing and responding to patient deterioration have led hospitals to use Early Warning Scoring Systems (EWSS) (1) or medical emergency team (MET) calling criteria (2) to improve vital signs monitoring and facilitate the calling of expert help to a patient's bedside.

EWSS allocate points in a weighted manner, based on the derangement of a patient's measured vital signs from arbitrarily agreed "normal" ranges—the sum of these is termed the Early Warning Score (EWS). The EWS is used to direct subsequent care, for example, changes to vital signs monitoring frequency; involvement of more experienced ward staff; or calling a rapid response team (RRT). Many EWSS are in use, with marked variation in measured physiologic variables, assigned weightings, and outcome discrimination (3–8). In 2012, the Royal College of Physicians of London (RCPL) recommended the use of a standardized EWSS in the National Health Service (NHS)—the National EWS (NEWS) (**Supplementary Digital Content 1**, http://links.lww.com/CCM/B981) (9). To produce NEWS, the RCPL used clinical opinion to make minor adjustments to the VitalPAC EWS (ViEWS) (5). The RCPL recommends that NEWS values of greater than or equal to 7 should prompt assessment by an RRT (9). NEWS demonstrates better ability than other published EWSS to discriminate patients at risk of a range of clinical outcomes (6) and has been validated outside its development site (10–13).

Some hospitals, especially in the United States and Australia, use MET calling criteria in preference to EWSS. Most MET criteria are based on extreme values of specific objective physiologic variables (e.g., pulse rate, < 40 or > 120 beats/min) (2) (**Supplementary Digital Content 2**, http://links.lww.com/CCM/B982). When one or more objective MET criteria occur, or staff are "worried" about a patient, a MET or other RRT is called to provide expert assistance (14). As with EWSS, a wide range of MET criteria is in use, with varied abilities to discriminate patients at risk of adverse events (3, 15–17).

Ideally, hospitals should use an RRT triggering system that provides the highest discrimination of patient outcome and the lowest trigger rate, thereby minimizing both the risk of missing serious outcomes and of excessive staff workload. A recent study comparing the performances of NEWS and one set of MET criteria suggests that NEWS is a better (and earlier) detector of patient deterioration (13). Therefore, we used a large database of vital sign measurements to a) compare the abilities of NEWS and 44 different MET criteria to discriminate patients at risk of four outcomes (i.e., cardiac arrest, unanticipated [i.e., emergency] ICU admission, death, or a combined outcome of any of these three) within 24 hours of a vital signs dataset and b) measure the associated trigger rates.

MATERIALS AND METHODS

The study was covered by local research ethics committee approval ref 08/02/1394, granted by the Isle of Wight, Portsmouth, and South East Hampshire Research Ethics Committee.

Setting and Study Population

Portsmouth Hospitals NHS Trust is a single site NHS District General Hospital with ~ 1,000 inpatient beds and ~ 5,500 staff. It provides all acute services except burns, spinal injury, neurosurgic, and cardiothoracic surgery to a local population of ~ 540,000. Routinely, staff use hand-held devices and commercially available software (VitalPAC; The Learning Clinic, London, United Kingdom) (18, 19) to record all vital signs at the bedside in all adult in-patient areas of the hospital, except high care areas such as critical care units. For this study, vital signs were collected during routine clinical care from adult patients (≥ 16 yr) admitted on or after May 25, 2011, and discharged on or before December 31, 2013. Data from patients discharged alive from the hospital before midnight on the day of admission were excluded. Data were not captured from patients transferred directly on admission to critical care areas of the hospital.

Vital Signs Database and its Development

For each vital signs measurement, the following data were recorded using VitalPAC software: date/time of observation set; pulse rate; systolic and diastolic blood pressures; breathing rate; temperature; neurologic status using the Alert-Verbal-Painful-Unresponsive (AVPU) scale; peripheral oxygen saturation (SpO_2); and the inspired gas (i.e., air or supplemental oxygen). Where oxygen was used, VitalPAC estimated its fractional inspired concentration (FIO_2) using the mask type ± flow rate (or in the case of a Venturi mask, the concentration), which were recorded during each vital signs collection. Vital signs sets for which data were absent or physiologically impossible were excluded.

Evaluation of NEWS and MET Criteria

The vital signs database was used to evaluate the performance of NEWS and 44 different MET criteria (identified from two previous publications [16, 17]; **Supplementary Digital Content 3**, http://links.lww.com/CCM/B983). As the subjective component of MET criteria—staff concern (14)—is also used to escalate care when using NEWS, we made an a priori decision to evaluate only the following physiologic components of the MET criteria: high/low pulse rate, high/low breathing rate, high/low systolic blood pressure, high/low temperature, SpO_2, and reduced consciousness. For the same reason, we did not evaluate criteria such as threatened airway or repeated/prolonged seizures. For MET criteria, "reduced consciousness" was considered to be equivalent to a score of P or U on the AVPU scale. Two sets of MET criteria (20, 21) require knowledge of the FIO_2 when using SpO_2 (i.e., Ball [20] triggers when SpO_2 < 90% and FIO_2 ≥ 0.35 simultaneously; Hickey [21] when SpO_2 < 90% and FIO_2 ≥ 0.24 simultaneously). The remainder require only an SpO_2 value or simply whether supplemental oxygen was being administered. In the majority of observation sets where supplemental oxygen was administered, there was sufficient information on mask type and oxygen concentration/flow for an estimate of FIO_2 to be made. We removed hospital episodes where FIO_2 could not be estimated.

Outcomes

Deaths, cardiac arrests, and unanticipated ICU admission data were identified from the hospital's patient administration system (PAS), cardiac arrest database, and ICU admission database, respectively. We limited the analysis to the first of any of three outcomes (death, unanticipated ICU admission, or cardiac arrest) within 24 hours of a given observation set, within any episode of care. These outcomes were combined to produce a fourth—the combined outcome of any death, unanticipated ICU admission, or cardiac arrest within 24 hours of a given observation set. We excluded episodes of care where 1) the episode had a first outcome before the first observation set and 2) the episode did not have an observation set within the last 24 hours before the outcome.

Statistical Analysis

All data manipulation was performed using Microsoft` Visual FoxPro 9.0 (Microsoft Corporation, Redmond, WA). We used IBM SPSS Statistics v22 (IBM Corp., Armonk, NY) and R v3.02 (22) to calculate the area under the receiver-operating characteristic (ROC) curve (AUROC); R was also used to generate the figures.

We used the AUROC (23) to evaluate the ability of NEWS to discriminate between patients experiencing/not experiencing an adverse outcome at 24 hours post vital signs observation. An ROC curve plots sensitivity against 1 – specificity, and each point on it represents a sensitivity/specificity pairing corresponding to a particular decision threshold for NEWS. We plotted ROC curves for all four outcomes for NEWS. For each set of MET criteria, and for each outcome, we calculated the sensitivity and specificity. Any individual point represents a sensitivity/specificity pairing (there can only be one point per set of MET criteria). To compare the performance of NEWS and the different MET criteria, we superimposed the sensitivity/specificity points for the 44 sets of MET criteria on the NEWS ROC, for each outcome. The closer the NEWS ROC curve or any individual MET sensitivity/1 – specificity point is to the upper left corner, the higher the discrimination of the test.

We also plotted an efficiency curve (5) for NEWS for each outcome. These plot sensitivity against trigger rate (i.e., percentage of observations at, or above, a given NEWS value). To compare the efficiency of NEWS and the MET criteria, we superimposed the sensitivity/trigger rate points for the 44 sets of MET criteria on the NEWS efficiency curves. The closer the NEWS efficiency curve, or any individual MET sensitivity/trigger rate point, is to the lower right corner, the higher the efficiency of the test (i.e., more outcomes are detected for a lower trigger rate).

Additional Analyses

We have previously shown that the use of multiple observation sets from a single episode does not bias the ranking of EWSs when assessing the performance of these systems (24). This has not previously been done for sets of MET criteria. Therefore, we repeated the above analyses using 10,000 samples of observation sets, each sample being constructed by selecting one observation set at random from every care episode (i.e., so each observation set in an episode had an equal chance of being selected in each sample).

RESULTS

A total of 2,606,050 vital signs datasets were obtained from 111,389 hospital episodes. All sets were complete, valid, and contained sufficient data to permit the calculation of a NEWS value. Following exclusions (**Fig. 1**), the final dataset consisted of 2,245,778 vital signs sets from 103,998 episodes. There were 20,053 observation sets (0.89%) from 5,809 episodes where FIO_2 could not be estimated. For some of these 5,809 episodes, there were other observation sets where FIO_2 could be estimated, so the episode itself remained in the analysis (with fewer observation sets). Only 34 episodes were completely removed from the analysis because none of their observation

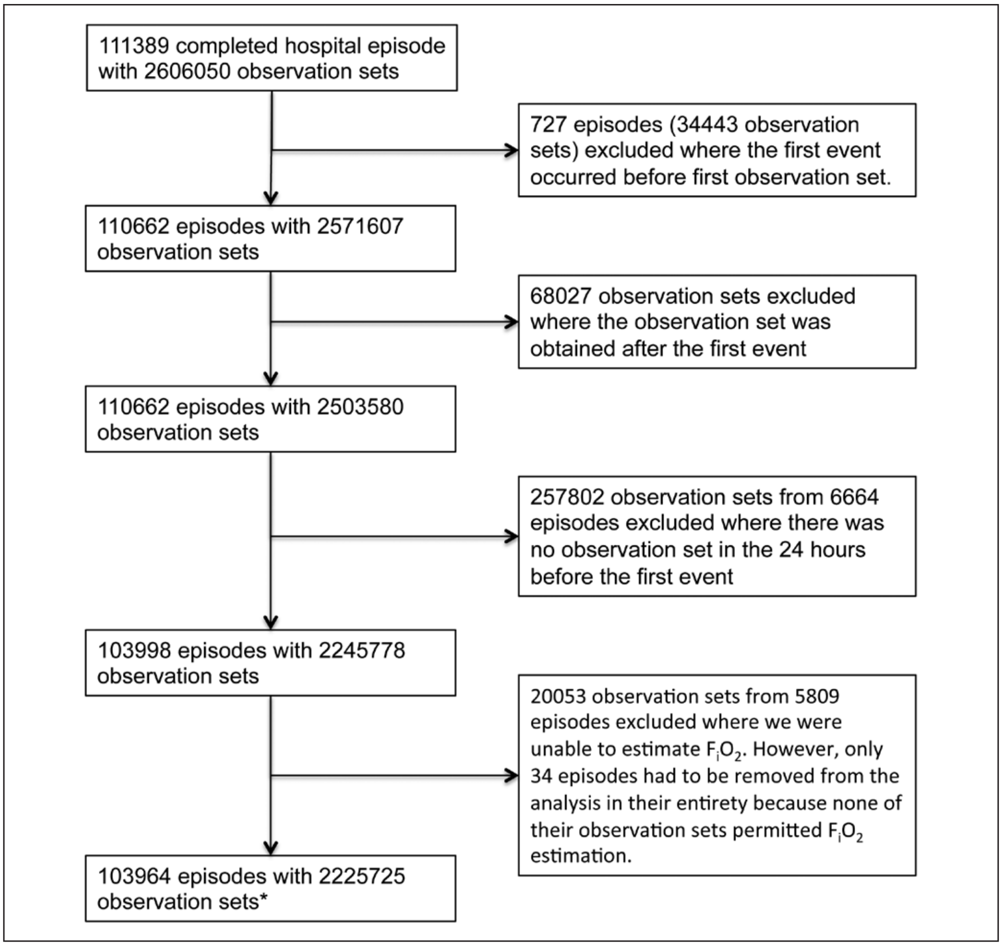

Figure 1. Consort diagram, outlining exclusions. *For two medical emergency team systems only (see main text of article).

sets permitted FiO_2 estimation. For the two sets of MET criteria requiring FiO_2 (20, 21), these observation sets were removed and the analysis was performed on 2,225,725 observation sets from 103,964 episodes.

Table 1 shows the patient demographics, number and value of the vital signs measurements, and observations followed by an adverse outcome. The study data were collected from 66,712 unique patients admitted to medicine (34,204), surgery (33,808), and other specialties (6,441). Patients may have more than one admission and may belong to different groups during different admissions—hence the sum of admissions to medicine, surgery, and other specialties is 74,453, and not 66,712. The 66,712 unique patients had 103,998 hospital episodes during the study period.

The AUROCs (95% CI) for NEWS for cardiac arrest, unanticipated ICU admission, death, and the combined outcome, each within 24 hours, were as follows: 0.78 (0.76–0.78), 0.86 (0.85–0.86), 0.91 (0.91–0.92), and 0.88 (0.88–0.88), respectively.

Table 2 shows that the sensitivity and specificity for the MET criteria varied considerably for the different outcomes. These findings were similar when using the 10,000 random sample sets (**Supplementary Digital Content 4**, http://links.lww.com/CCM/B984).

Figure 2, *A–D* and **Supplementary Digital Content 5** (http://links.lww.com/CCM/B985; **legend**, **Supplemental Digital Content 11**, http://links.lww.com/CCM/B991) show the sensitivity and specificity (plotted as 1 – specificity) points for NEWS (i.e., the NEWS ROC curve) and the MET criteria for the outcomes studied. For all outcomes, the NEWS ROC curve lies above and to the left of the MET criteria, indicating better discrimination. Although some MET systems have a higher sensitivity than NEWS values of greater than or equal to 7, all have a lower specificity. The relative positions of MET criteria and NEWS were essentially unchanged when using the 10,000 random sample sets (**Supplementary Digital Content 6**, http://links.lww.com/CCM/B986 [legend, Supplemental Digital Content 11, http://links.lww.com/CCM/B991]; and **Supplementary Digital Content 7**, http://links.lww.com/CCM/B987 [legend, Supplemental Digital Content 11, http://links.lww.com/CCM/B991]).

Figure 3, *A–D* and **Supplementary Digital Content 8** (http://links.lww.com/CCM/B988 [legend, Supplemental Digital Content 11, http://links.lww.com/CCM/B991] show the efficiency curve for NEWS and the sensitivity/trigger rate points for each set of MET criteria for the outcomes studied. For all four outcomes, the MET criteria points lie above and to the left of the NEWS efficiency curve, indicating a higher workload (trigger rate) is required to detect a given percentage of the considered outcome. Although some MET systems have a higher sensitivity than NEWS values of greater than or equal to 7, all would generate greater workloads (i.e., higher trigger rates). The relative positions of the MET criteria and NEWS were essentially unchanged when using the 10,000 random sample sets (**Supplementary Digital Content 9**, http://links.lww.com/CCM/B989 [legend, Supplemental Digital Content 11, http://links.lww.com/CCM/B991]; and **Supplementary Digital Content 10**, http://links.lww.com/CCM/B990 [legend, Supplemental Digital Content 11, http://links.lww.com/CCM/B991).

DISCUSSION

The selection of an RRT triggering system can be based upon several criteria, including the balance between its sensitivity and the workload it generates. To minimize missed outcomes and

TABLE 1. Patient Demographics, Number of Vital Signs Observations, Vital Sign Values, and Observations Followed by an Adverse Outcome for the Study Group (After Exclusions)

	Admissions to Medicine	Admissions to Surgery	Admissions to Other Specialties	All Admissions
Admissions				
No. of episodes, *n* (%)	53,466 (51.4)	42,641 (41.0)	7,891 (7.6)	103,998 (100)
Age at admission, mean (SD)	67.5 (18.6)	57.4 (20.6)	52.6 (20.3)	62.2 (20.4)
Male, *n* (%)	26,910 (50.3)	20,916 (49.1)	1,584 (20.1)	49,410 (47.5)
First adverse event in episode, *n* (%)				
None	50,574 (94.6)	41,786 (98.0)	7,768 (98.4)	100,128 (96.3)
Death	1,966 (3.7)	304 (0.7)	81 (1.0)	2,351 (2.3)
Cardiac arrest	397 (0.7)	71 (0.2)	11 (0.1)	479 (0.5)
Unanticipated ICU admission	529 (1.0)	480 (1.1)	31 (0.4)	1,040 (1.0)
Any	2,892 (5.4)	855 (2.0)	123 (1.6)	3,870 (3.7)
Observations				
No. of observations, *n* (%)	1,264,024 (56.3)	864,590 (38.5)	117,164 (5.2)	2,245,778 (100)
Vital signs values				
Pulse rate (beats/min), mean (SD)	80.5 (16.5)	78.3 (14.7)	79.0 (14.8)	79.6 (15.8)
Respiration rate (beats/min), mean (SD)	17.3 (3.1)	15.9 (2.6)	16.0 (2.7)	16.7 (3.0)
Temperature (°C), mean (SD)	36.7 (0.5)	36.8 (0.5)	36.7 (0.4)	36.7 (0.5)
BP Systolic (mm Hg), mean (SD)	126.6 (22.2)	125.4 (20.9)	122.6 (20.7)	126.0 (21.6)
SpO_2 (%), mean (SD)	96.0 (2.5)	96.5 (2.0)	96.5 (2.1)	96.2 (2.4)
SpO_2 recorded on supplemental O_2, *n* (%)	232,591 (18.4)	167,536 (19.4)	18,965 (16.2)	419,092 (18.7)
Conscious level, *n* (%)				
Alert (A)	1,238,015 (97.9)	860,651 (99.5)	116,487 (99.4)	2,215,153 (98.6)
Responds to Voice (V)	16,469 (1.3)	3,110 (0.4)	517 (0.4)	20,096 (0.9)
Responds to Pain (P)	8,227 (0.7)	583 (0.1)	90 (0.1)	8,900 (0.4)
Unresponsive (U)	1,313 (0.1)	246 (< 0.1)	70 (0.1)	1,629 (0.1)
Observations followed by outcomes (first event within 24 hr), *n* (%)				
Death	8,686 (0.7)	1,693 (0.2)	388 (0.3)	10,767 (0.5)
Cardiac arrest	1,756 (0.1)	402 (< 0.1)	73 (0.1)	2,231 (0.1)
Unanticipated ICU admission	3,720 (0.3)	3,738 (0.4)	249 (0.2)	7,707 (0.3)
Any outcome	14,162 (1.1)	5,833 (0.7)	710 (0.6)	20,705 (0.9)

SpO_2 = peripheral oxygen saturation.

excessive staff workload, hospitals should choose a system that provides the highest discrimination of patient outcome and the lowest trigger rate. Depending upon their specific criteria, all sets of MET calling criteria have fixed relationships between sensitivity and workload, and the resulting clinical response can only ever be of an "all or none" nature. In contrast, the multinodal nature of NEWS provides the opportunity to titrate the RRT trigger value to available resources. When MET systems were compared to a NEWS value of greater than or equal to 7, some MET systems have a higher sensitivity than NEWS. However, all of these MET systems had a lower specificity and would generate greater workloads (i.e., higher trigger rates).

Our results complement those of Tirkkonen et al (13) who showed that high NEWS values were associated with serious adverse events in hospital, but MET criteria were not. NEWS was also independently associated with higher mortality, whereas

TABLE 2. The Sensitivity and Specificity of 44 Medical Emergency Team Calling Criteria, and National Early Warning Score Values of 3–7, for Cardiac Arrest, Unanticipated ICU Admission, Death, and Any of These Outcomes, Each Within 24 Hours of a Vital Signs Dataset

	Outcome Within 24 Hr of a Vital Signs Dataset							
	Death		Cardiac Arrest		Unanticipated ICU Admission		Any of Death; Cardiac Arrest; Unanticipated ICU Admission	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Medical emergency team calling criteria								
Bell (standard)	49.9	95.2	21.5	95.0	38.0	95.1	42.4	95.3
Bell (extended)	61.8	86.8	35.6	86.6	52.9	86.7	55.7	87.0
Bell (restricted)	41.7	97.3	14.7	97.1	25.3	97.2	32.7	97.4
Ball	47.7	93.6	23.0	93.4	45.5	93.5	44.2	93.7
Parissopoulos	58.3	93.6	27.4	93.4	47.1	93.5	50.8	93.8
Hickey	50.3	94.9	21.8	94.7	40.3	94.8	43.5	95.0
Salamonson	41.0	96.1	14.7	95.9	29.3	96.0	33.8	96.2
Buist	49.9	95.2	21.3	95.0	37.8	95.1	42.3	95.4
Bellomo	49.9	95.2	21.5	95.0	38.0	95.1	42.4	95.3
Jones	51.0	94.7	22.8	94.5	38.5	94.6	43.3	94.8
Green	52.3	94.5	23.8	94.3	42.4	94.5	45.5	94.7
Harrison (early)	77.5	71.4	54.6	71.2	67.2	71.3	71.2	71.5
Harrison (late)	40.0	96.9	14.6	96.8	22.7	96.8	30.8	97.0
Smith	73.0	81.5	47.1	81.3	69.9	81.4	69.0	81.7
Lee	57.2	87.8	32.4	87.6	50.4	87.7	52.0	87.9
Buist	47.4	94.6	20.0	94.4	37.0	94.5	40.6	94.8
McGloin	69.7	78.7	44.7	78.5	59.4	78.6	63.2	78.9
de Pennington	46.4	95.1	19.4	94.9	40.4	95.0	41.3	95.2
McArthur-Rouse	42.7	95.3	17.2	95.2	36.1	95.3	37.5	95.5
Foraida	34.9	97.7	10.0	97.5	21.0	97.6	27.1	97.7
Cioffi	41.1	95.9	14.7	95.7	29.6	95.8	34.0	96.0
Holder	58.3	93.6	27.4	93.4	47.1	93.5	50.8	93.8
Buist	51.3	93.8	23.4	93.6	38.9	93.7	43.7	94.0
McQuillan	43.4	95.4	16.5	95.3	34.0	95.3	37.0	95.5
Hourihan	37.9	96.3	13.0	96.1	28.0	96.2	31.5	96.4
Bristow	40.4	95.6	13.0	96.1	30.4	95.5	33.9	95.7
Jones	41.4	97.3	14.7	95.4	24.8	97.2	32.3	97.4
Subbe	27.0	98.4	14.6	97.2	13.5	98.3	19.6	98.5
Jacques (late)	40.0	96.9	14.6	97.2	22.7	96.8	30.8	97.0
Offner	53.9	94.5	5.2	98.3	44.5	94.4	47.1	94.7

(*Continued*)

TABLE 2. (*Continued*). The Sensitivity and Specificity of 44 Medical Emergency Team Calling Criteria, and National Early Warning Score Values of 3–7, for Cardiac Arrest, Unanticipated ICU Admission, Death, and Any of These Outcomes, Each Within 24 Hours of a Vital Signs Dataset

	Outcome Within 24 Hr of a Vital Signs Dataset							
	Death		Cardiac Arrest		Unanticipated ICU Admission		Any of Death; Cardiac Arrest; Unanticipated ICU Admission	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Cretikos (original)	37.8	96.3	14.6	96.8	27.8	96.2	31.4	96.4
Cretikos set 1	57.0	93.3	22.9	94.3	49.4	93.2	50.9	93.5
Cretikos set 2	55.7	93.7	12.7	96.2	47.2	93.6	49.3	93.9
Cretikos set 3	54.3	94.7	26.7	93.1	45.8	94.7	47.8	94.9
Cretikos set 4	52.9	95.2	25.0	93.5	43.5	95.1	46.1	95.4
Cretikos set 5	52.0	95.5	23.9	94.5	41.9	95.4	44.9	95.6
Cretikos set 6	51.2	95.7	22.3	95.0	40.1	95.6	43.7	95.8
Cretikos set 7	50.1	95.9	21.0	95.2	38.8	95.8	42.6	96.0
Cretikos set 8	46.6	96.3	19.8	95.5	35.5	96.2	39.2	96.5
Cretikos set 9	44.1	97.0	19.0	95.7	31.5	96.9	36.3	97.1
Cretikos set 10	39.1	97.5	16.9	96.2	27.2	97.4	31.7	97.6
Jones	49.9	95.2	21.3	95.0	37.8	95.1	42.3	95.4
Jones	45.5	95.6	18.1	95.4	31.7	95.5	37.4	95.7
Parr	39.0	96.2	13.4	96.1	29.0	96.1	32.5	96.3
National Early Warning Scores								
3	90.5	73.6	68.0	73.3	82.8	73.5	85.2	73.8
4	82.9	84.9	56.0	84.6	72.3	84.7	76.1	85.1
5	75.2	91.2	43.1	90.9	61.9	91.0	66.8	91.4
6	65.4	95.0	31.5	94.8	49.9	94.9	56.0	95.2
7	54.2	97.2	22.2	97.0	37.4	97.1	44.5	97.4

MET criteria were not. Churpek et al (7) compared the performance of several EWSs other than NEWS (5, 8, 25–27) with two sets of MET criteria (15, 28) and showed that EWSs generally had higher predictive accuracy. That EWSs, such as NEWS, are better discriminators of outcomes than MET criteria is perhaps not surprising. The activation of an RRT, when based on objective physiologic criteria, depends upon the presence of one or more extreme vital sign value (2, 15–17, 20, 21, 28). Cretikos et al (16) studied the impact of varying MET criteria and showed that all tested modifications provided positive predictive values of less than 16% (i.e., ~ 84% of resultant calls would be to patients who would not experience an adverse event). Consequently, workload and the proportion of false positive calls would be high, and a substantial number of at risk patients might remain unidentified (16). In contrast, EWSS provide an aggregate score based upon weightings for the, sometimes subtle, physiologic disturbance of several vital signs. This may better reflect changes that occur in many disease states. This is supported by the observation that aggregate NEWS values are more important for discriminating adverse outcomes than high scores for a single vital signs variable (i.e., extreme values of a given vital sign) (29). Taking this body of evidence together with our own findings, it seems reasonable to conclude that EWSs, such as NEWS, provide better detection of adverse outcomes at a lower trigger rate (i.e., workload) than MET systems.

Advocates of MET criteria often argue that an advantage over EWSS is their inclusion of trigger criteria other than vital signs, for example, “staff concern,” threatened airway, seizures. Staff concern may account for a large proportion of RRT activations (14, 30, 31), but EWSS can also trigger in response

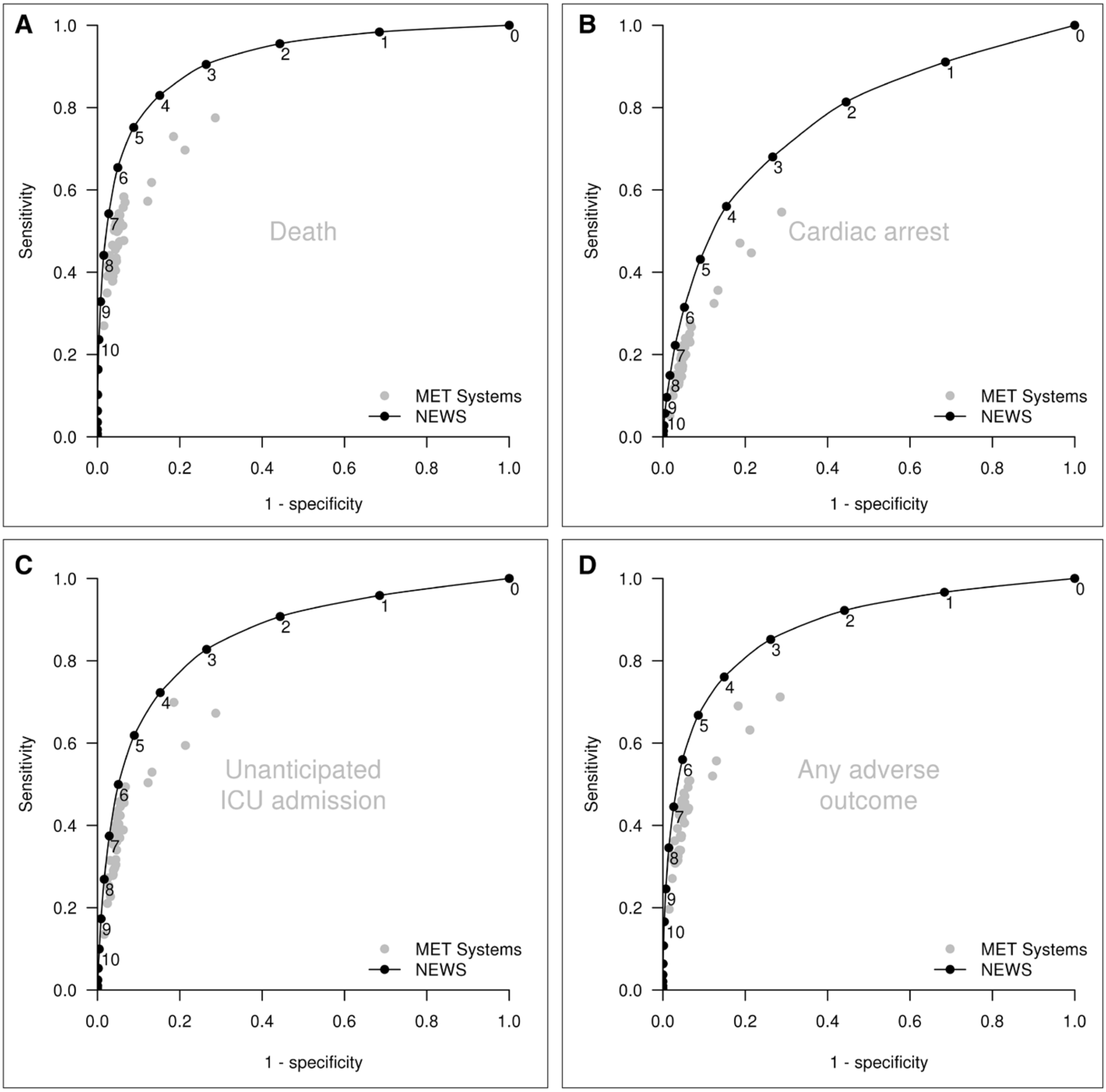

Figure 2. Comparative sensitivity and specificity (plotted as 1 − specificity) of National Early Warning Score (NEWS) (i.e., NEWS receiver-operating characteristic [ROC] curve) and 44 sets of medical emergency team (MET) criteria to discriminate **A**, death occurring within 24 hr of a vital signs dataset; **B**, cardiac arrest occurring within 24 hr of a vital signs dataset; **C**, unanticipated ICU admission occurring within 24 hr of a vital signs dataset; and **D**, the combined outcome of cardiac arrest, unanticipated ICU admission or death, occurring within 24 hr of a vital signs dataset. Each point on the NEWS ROC curve represents a NEWS value from 0 to 20.

to "staff concern" (current advice is that concern about a patient's condition should always override the NEWS value) (9). Threatened airway and seizures are not included in NEWS but would generate an escalation (or call) exactly as they do when using MET criteria. Data show that threatened airway was the trigger for a MET call in only 2% of calls averaged over a 10-year period, with the figure for seizures being only 0.6% (31). Therefore, the impact of the omission of threatened airway and seizure from our analysis can reasonably be expected to be small, given their infrequent occurrence as MET triggers. Other than a fall in Glasgow Coma Scale (GCS) greater than or equal to 3 points, none of the MET criteria analyzed include changes in vital signs as MET calling criteria. In our analysis, we considered "reduced consciousness" to be equivalent to P or U on the AVPU scale. We consider this to be an appropriate approach, as the chance of a patient having a fall in GCS value of greater than or equal to 3, but continuing to score A or V on the AVPU scale is negligible.

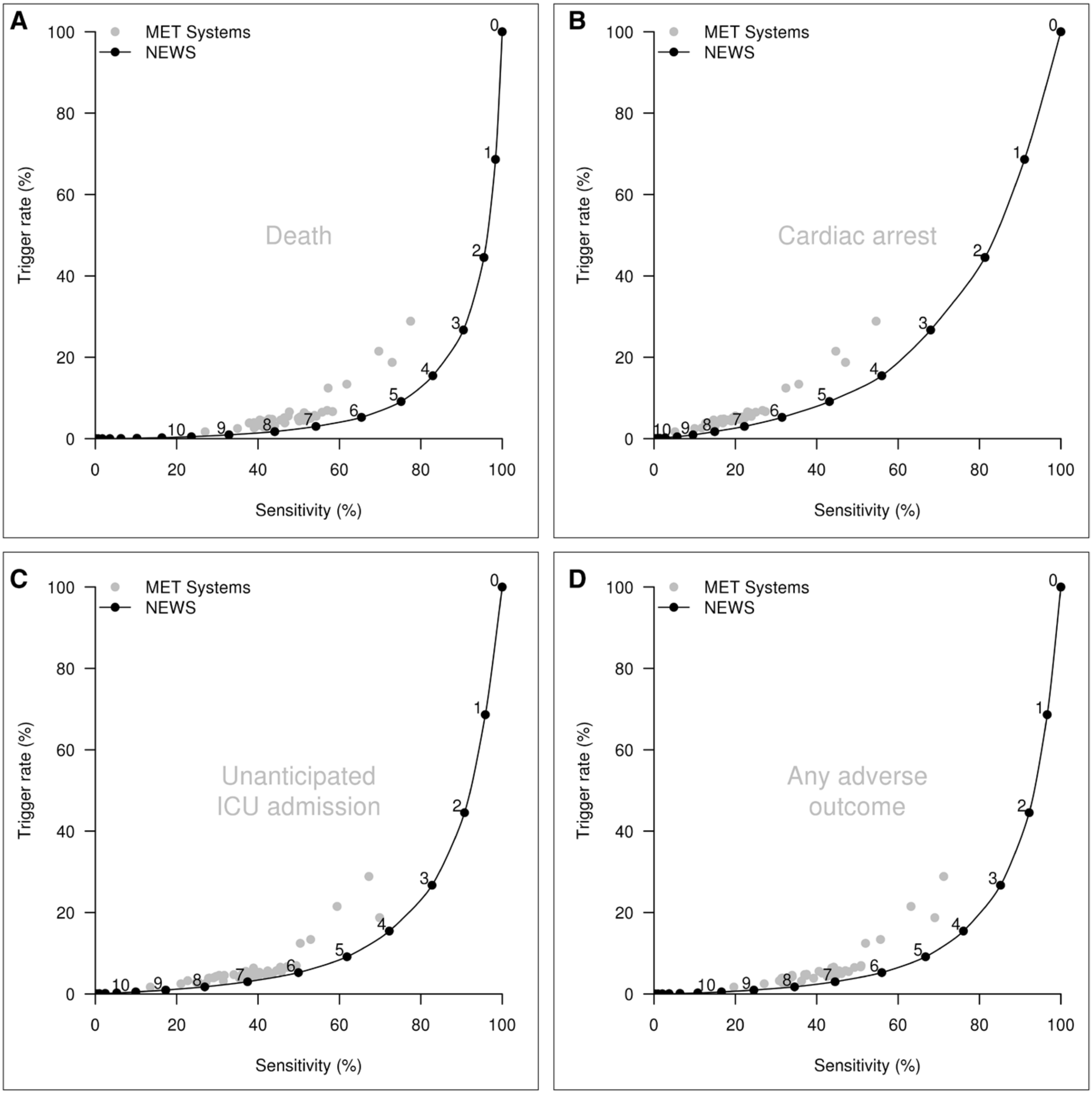

Figure 3. Efficiency curve for National Early Warning Score (NEWS), with comparable data for 44 sets of medical emergency team (MET) criteria superimposed, for **A**, death occurring within 24 hr of a vital signs dataset; **B**, cardiac arrest occurring within 24 hr of a vital signs dataset; **C**, unanticipated ICU admission occurring within 24 hr of a vital signs dataset; and **D**, the combined outcome of cardiac arrest, unanticipated ICU admission, or death, occurring within 24 hr of a vital signs dataset. This plots workload (trigger rate) against the sensitivity of the Early Warning Score or set of MET criteria in question. Each point on the NEWS efficiency curve represents a NEWS value from 0 to 20, starting with NEWS equal to 0 at the top right. Trigger rate = ([true positive + false positive]/[true positive + false positive + true negative + false negative]).

This large study has several strengths. It considers all completed admissions over 31 months. All necessary vital sign variables were collected simultaneously in a standardized manner as part of clinical care using an electronic data collecting system (18). However, there are also weaknesses. Patients admitted directly to critical care areas were excluded and patients with Do Not Attempt Cardiopulmonary Resuscitation (DNACPR) decisions were included. We have partially mitigated the latter by excluding patients with no vital signs observations within the last 24 hours of their stay. This should exclude most patients on a recognized end-of-life pathway, but will mean that patients with DNACPR decisions who are not on one are included. Patients in the latter continue to receive normal care, including the measurement of vital signs, and EWSs still have utility in identifying their early deterioration. We have also assumed that the treatment of all study patients was both optimal and equitable. Additionally, we obtained the date/time of death (or discharge) from the hospital's PAS and these data are likely to be systematically late. Therefore, the number of observations followed by death within 24 hours may have been underestimated.

A further weakness is that the study was conducted in a single site, where the precursor of NEWS—ViEWS (5)—was developed. Prediction models are almost always more accurate in the population in which they were derived. However, the current study differs markedly from the NEWS development work, using a larger database, a different study period, medical and surgical patients (compared to only acute medicine), and vital signs from the whole patient admission rather than merely from the patient's stay in the Medical Assessment Unit. Nevertheless, as with all studies on models that are tested in the site of their development, our results require external validation. Finally, our study is a statistical evaluation of NEWS and MET criteria. There is no guarantee that similar results would be generated operationally when human factors may have an influence (32–37).

CONCLUSIONS

When MET systems are compared to a NEWS value of greater than or equal to 7 (i.e., the recommended trigger points for RRT intervention for each system), some MET systems have a higher sensitivity than NEWS. However, all of these MET systems have a lower specificity and would generate greater workloads. NEWS also provides the opportunity to titrate the trigger value against available resources and permits a graduated, multitiered clinical response, whereas the clinical response resulting from a MET call can only ever be of an "all or none" nature.

ACKNOWLEDGMENTS

We would like to acknowledge the cooperation of the nursing and medical staff in the study hospital.

REFERENCES

1. Morgan R, Williams F, Wright M: An Early Warning Scoring System for detecting developing critical illness. *Clin Intensive Care* 1997; 8:100
2. Hourihan F, Bishop G, Hillman KM, et al: The medical emergency team: A new strategy to identify and intervene in highrisk surgical patients. *Clin Intensive Care* 1995; 6:269–272
3. 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; 33:667–679
4. Smith GB, Prytherch DR, Schmidt PE, et al: Review and performance evaluation of aggregate weighted 'track and trigger' systems. *Resuscitation* 2008; 77:170–179
5. Prytherch DR, Smith GB, Schmidt PE, et al: ViEWS–Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010; 81:932–937
6. Smith GB, Prytherch DR, Meredith P, et al: 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; 84:465–470
7. Churpek MM, Yuen TC, Edelson DP: Risk stratification of hospitalized patients on the wards. *Chest* 2013; 143:1758–1765
8. Tarassenko L, Clifton DA, Pinsky MR, et al: Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation* 2011; 82:1013–1018
9. National Early Warning Score (NEWS): Standardising the Assessment of Acute-Illness Severity in the NHS. Report of a Working Party. London, Royal College of Physicians, 2012
10. Abbott TE, Vaid N, Ip D, et al: A single-centre observational cohort study of admission National Early Warning Score (NEWS). *Resuscitation* 2015; 92:89–93
11. Alam N, Hobbelink EL, van Tienhoven AJ, et al: Exploring the performance of the National Early Warning Score (NEWS) in a European emergency department. *Resuscitation* 2015; 90:111–115
12. Silcock DJ, Corfield AR, Gowens PA, et al: Validation of the National Early Warning Score in the prehospital setting. *Resuscitation* 2015; 89:31–35
13. Tirkkonen J, Olkkola KT, Huhtala H, et al: Medical emergency team activation: Performance of conventional dichotomised criteria versus national early warning score. *Acta Anaesthesiol Scand* 2014; 58:411–419
14. Santiano N, Young L, Hillman K, et al: Analysis of medical emergency team calls comparing subjective to "objective" call criteria. *Resuscitation* 2009; 80:44–49
15. Bell MB, Konrad D, Granath F, et al: Prevalence and sensitivity of MET-criteria in a Scandinavian University Hospital. *Resuscitation* 2006; 70:66–73
16. Cretikos M, Chen J, Hillman K, et al; MERIT study investigators: The objective medical emergency team activation criteria: A case-control study. *Resuscitation* 2007; 73:62–72
17. Smith GB, Prytherch DR, Schmidt PE, et al: A review, and performance evaluation, of single-parameter "track and trigger" systems. *Resuscitation* 2008; 79:11–21
18. Smith GB, Prytherch DR, Schmidt P, et al: Hospital-wide physiological surveillance-a new approach to the early identification and management of the sick patient. *Resuscitation* 2006; 71:19–28
19. Schmidt PE, Meredith P, Prytherch DR, et al: Impact of introducing an electronic physiological surveillance system on hospital mortality. *BMJ Qual Saf* 2015; 24:10–20
20. Ball C: Critical care outreach services–do they make a difference? *Intensive Crit Care Nurs* 2002; 18:257–260
21. Hickey C, Allen M: A critical care liaison service. *Br J Anaesthesia* 1998; 81:650
22. R Core Team: R: A Language and Environment for Statistical Computing, 2013. Vienna, Austria, R Foundation for Statistical Computing. Available at: http://www.R-project.org/. Accessed June 20, 2015
23. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29–36
24. Jarvis SW, Kovacs C, Briggs JS, et al: Are vital signs sets independent? Effects of observation selection methods on early warning score performance. *Resuscitation* 2015; 90:1–6
25. Subbe CP, Kruger M, Rutherford P, et al: Validation of a modified Early Warning Score in medical admissions. *QJM* 2001; 94:521–526
26. Paterson R, MacLeod DC, Thetford D, et al: Prediction of in-hospital mortality and length of stay using an Early Warning Scoring System: Clinical audit. *Clin Med (Lond)* 2006; 6:281–284
27. Churpek MM, Yuen TC, Park SY, et al: Derivation of a cardiac arrest prediction model using ward vital signs*. *Crit Care Med* 2012; 40:2102–2108
28. Hillman K, Chen J, Cretikos M, et al; MERIT study investigators: Introduction of the medical emergency team (MET) system: A cluster-randomised controlled trial. *Lancet* 2005; 365:2091–2097
29. Jarvis SW, Kovacs C, Briggs JS, et al: Aggregate scores are more important than high scores in a single component of the National Early

Warning Score (NEWS) in discriminating risk of adverse outcomes. *Resuscitation* 2015; 87:75–80

30. Chen J, Bellomo R, Hillman K, et al; MERIT Study Investigators for the Simpson Centre and the ANZICS Clinical Trials Group: Triggers for emergency team activation: A multicenter assessment. *J Crit Care* 2010; 25:359.e1–359.e7
31. Herod R, Frost SA, Parr M, et al: Long term trends in medical emergency team activations and outcomes. *Resuscitation* 2014; 85:1083–1087
32. Edwards M, McKay H, Van Leuvan C, et al: Modified Early Warning Scores: Inaccurate summation or inaccurate summation of score? *Crit Care* 2010; 14:S88
33. Smith AF, Oakey RJ: Incidence and significance of errors in a patient 'track and trigger' system during an epidemic of Legionnaires' disease: Retrospective casenote analysis. *Anaesthesia* 2006; 61:222–228
34. Prytherch DR, Smith GB, Schmidt P, et al: Calculating early warning scores–a classroom comparison of pen and paper and hand-held computer methods. *Resuscitation* 2006; 70:173–178
35. Mohammed M, Hayton R, Clements G, et al: Improving accuracy and efficiency of early warning scores in acute care. *Br J Nurs* 2009; 18:18–24
36. Kolic I, Crane S, McCartney S, et al: Factors affecting response to national early warning score (NEWS). *Resuscitation* 2015; 90:85–90
37. Subbe CP, Gao H, Harrison DA: Reproducibility of physiological track-and-trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med* 2007; 33:619–624