Article:

Smith, Gary B., Prytherch, David R, Jarvis, Stuart William orcid.org/0000-0001-8447-0306 et al. (4 more authors) (2016) 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. Critical Care Medicine. 2171–2181. ISSN 0090-3493

https://doi.org/10.1097/CCM.0000000000002000

A comparison of the ability of the physiological components of Medical Emergency Team criteria and the UK National Early Warning Score (NEWS) to discriminate patients at risk of a range of adverse clinical outcomes

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

[1] Faculty of Health and Social Sciences, University of Bournemouth, Bournemouth, UK
[2] Portsmouth Hospitals NHS Trust, Portsmouth, UK
[3] Centre for Healthcare Modelling & Informatics, University of Portsmouth, Portsmouth, UK
[4] Department of Health Sciences, University of York, York, UK

Work performed at Portsmouth Hospitals NHS Trust & University of Portsmouth
No reprints will be available

Correspondence from:
Professor G B Smith, FRCA, FRCP,
Centre of Postgraduate Medical Research & Education (CoPMRE),
Faculty of Health and Social Sciences,
Bournemouth University, Royal London House,
Christchurch Road, Bournemouth,
Dorset BH1 3LT, United Kingdom

Tel: +44 (0) 1202 962782; Fax: +44 (0) 1202 962218
Email: gbsresearch@virginmedia.com

Funding: Nil

Key words: Hospital rapid response team; Monitoring, Physiologic; Quality improvement; Vital Signs, Medical Emergency Team.

Word count = 2998. Abstract word count = 248

Number of references = 37; Figures = 3; Tables = 2; Supplementary Digital Content = 10

ABSTRACT

Objective: To compare the ability of Medical Emergency Team (MET) criteria and the National Early Warning Score (NEWS) to discriminate cardiac arrest, unanticipated ICU admission and death within 24 h of a vital signs measurement, and to quantify the associated workload.

Design: Retrospective cohort study.

Setting: A large UK NHS District General Hospital.

Patients: Adults hospitalized from 25/05/2011 to 31/12/2013.

Interventions: None

Measurements and Main Results: We applied NEWS and 44 sets of MET criteria to a database of 2,245,778 vital signs sets (103,998 admissions). NEWS' performance was assessed using the area under the receiver-operating characteristic (ROC) curve (AUROC) and compared with sensitivity/specificity for the different MET criteria. AUROC (95% CI) for NEWS for the combined outcome (i.e., death, cardiac arrest or unanticipated ICU admission) was 0.88 (0.88 - 0.88). A NEWS value of 7 had sensitivity/specificity values of 44.5%/97.4%. For the 44 sets of MET 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 NEWS ROC curve was above and to the left of all MET criteria points, indicating better discrimination. Similarly, the positions of all MET criteria points were above and to the left of the NEWS efficiency curve, indicating higher workloads (trigger rates).

Conclusions: When MET systems are compared to a NEWS value of ≥7, some MET systems have a higher sensitivity than NEWS values of ≥7. However, all of these MET systems have a lower specificity and would generate greater workloads.

CONFLICT OF INTERESTS STATEMENT

VitalPAC is a collaborative development of The Learning Clinic Ltd (TLC) and Portsmouth Hospitals NHS Trust (PHT). At the time of the research, PHT had a royalty agreement with TLC to pay for the use of PHT intellectual property within the VitalPAC product. Professor Prytherch, Dr Schmidt, and Dr Meredith are employed by PHT. Professor Smith was an employee of PHT until 31/03/2011. Professors Smith and Prytherch, and Dr Schmidt, are unpaid research advisors to TLC and have received reimbursement of travel expenses from TLC for attending symposia in the UK. Dr Briggs' research has previously received funding from TLC through a Knowledge Transfer Partnership. Professor Smith acted as 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'. He was also a member of the National Patient Safety Agency committee that wrote the two reports: 'Recognising and responding appropriately to early signs of deterioration in hospitalised patients' and ' Safer care for the acutely ill patient: learning from serious incidents'. He was a member of the Royal College of Physicians of London’s National Early Warning Score Development and Implementation Group (NEWSDIG). Professor Prytherch assisted the Royal College of Physicians of London in the analysis of data validating NEWS. Dr Jarvis and Mrs Kovacs have no conflicts of interest.

INTRODUCTION

Staff failures in recognising 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, e.g. 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 physiological variables, assigned weightings and outcome discrimination (3-8). In 2012, the Royal College of Physicians of London (RCPL) recommended the use of a standardised EWSS in the National Health Service (NHS) - the National EWS (NEWS) (Supplementary Digital Content 1) (9). To produce NEWS, the RCPL used clinical opinion to make minor adjustments to the VitalPAC Early Warning Score (ViEWS) (5). The RCPL recommends that NEWS values of $\geq$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 USA and Australia, use MET calling criteria in preference to EWSS. Most MET criteria are based on extreme values of specific objective physiological parameters (e.g., pulse rate <40 or >120 beats.min^{-1}) (2) (Supplementary Digital Content 2). When one or more objective MET criteria occurs, 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 minimising 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 h 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 (PHT) is a single site NHS District General Hospital with ~1000 inpatient beds and ~5500 staff. It provides all acute services except burns, spinal injury, neurosurgical and cardiothoracic surgery to a local population of ~540,000. Routinely, staff use hand-held devices and commercially available software (VitalPAC, The Learning Clinic Ltd, London, UK) (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 ($\geq$16 years) admitted on or after 25/05/2011 and discharged on or before 31/12/2013. 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; neurological status using the Alert-Verbal-Painful-Unresponsive (AVPU) scale; peripheral oxygen saturation (S_pO_2); and the inspired gas (i.e., air or supplemental oxygen). Where oxygen was used, VitalPAC estimated its fractional inspired concentration (F_iO_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) - see Supplementary Digital Content 3). 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 physiological components of the MET criteria:

high/low pulse rate, high/low breathing rate, high/low systolic blood pressure, high/low temperature, S_pO_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 F_iO_2 when using S_pO_2 (i.e., Ball (20) triggers when $S_pO_2 < 90\%$ and $F_iO_2 \geq 0.35$ simultaneously; Hickey (21) when $S_pO_2 < 90\%$ and $F_iO_2 \geq 0.24$ simultaneously). The remainder require only an S_pO_2 value or simply whether supplemental O_2 was being administered. In the majority of observation sets where supplemental O_2 was administered there was sufficient information on mask type and O_2 concentration/flow for an estimate of F_iO_2 to be made. We removed hospital episodes where F_iO_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 h 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 h of a given observation set. We excluded episodes of care where (i) the episode had a first outcome before the first observation set and (ii) the episode did not have an observation set within the last 24 h before the outcome.

Statistical analysis

All data manipulation was performed using Microsoft® Visual FoxPro 9.0. We used IBM SPSS Statistics v22 and R v3.02. (22) to calculate the AUROC; R was also used to generate the figures.

We used the area under the receiver-operating characteristic (ROC) curve (AUROC) (23) to evaluate the ability of NEWS to discriminate between patients experiencing/not experiencing an adverse outcome at 24 h 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 2606050 vital signs datasets were obtained from 111389 hospital episodes. All sets were complete, valid and contained sufficient data to permit the calculation of a NEWS value. Following exclusions (see Figure 1), the final dataset consisted of 2245778 vital signs sets from 103998 episodes. There were 20053 observation sets (0.89%) from 5809 episodes where F_iO_2 could not be estimated. For some of these 5809 episodes there were other observation sets where F_iO_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 sets permitted F_iO_2 estimation. For the two sets of MET criteria requiring F_iO_2 (20, 21), these observation sets were removed and the analysis was performed on 2225725 observation sets from 103964 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 66712 unique patients admitted to medicine (34204), surgery (33808) and other specialties (6441). 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 74453, and not 66712. The 66712 unique patients had 103998 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 h, were: 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 (see Supplementary Digital Content 4).

Figures 2a-d and Supplementary Digital Content 5 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 $\geq$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 (see Supplementary Digital Content 6 and 7).

Figures 3a-d and Supplementary Digital Content 8 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 ≥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 (see Supplementary Digital Content 9 and 10).

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 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 multi-nodal 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 ≥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. who showed that high NEWS values were associated with serious adverse events in hospital, but MET criteria were not (13). NEWS was also independently associated with higher mortality, whereas MET criteria were not. Churpek at 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 a RRT, when based on objective physiological criteria, depends upon the presence of one or more extreme vital sign value (2, 15-17, 20, 21, 28). Cretikos et al. studied the impact of varying MET criteria and showed that all tested modifications provided positive predictive values of <16% (i.e., ~84% of resultant calls would be to patients who would not experience an adverse event) (16). 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, physiological 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 parameter (i.e., extreme values of a given vital sign) (29). Taking this body of evidence together with our own findings, it seem 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, e.g., ‘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 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 GCS ≥3 points, none of the MET criteria analysed 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 ≥3, but continuing to score A or V on the AVPU scale is negligible.

This large study has several strengths. It considers all completed admissions over 31 months. All necessary vital sign variables were collected simultaneously in a standardised 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 h of their stay. This should exclude most patients on a recognised end-of-life (EoL) 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 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 h 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 $\geq$7 (i.e., the recommended triggers 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, multi-tiered clinical response, whereas the clinical response resulting from a MET call can only ever be of an 'all or none' nature.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the co-operation 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 high-risk surgical patients. *Clin Intensive Care* 1995; 6:269-72.
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 P, et al. Review and performance evaluation of aggregate weighted 'track and trigger' systems**.** *Resuscitation* 2008; 77:170-179.
5. Prytherch D, 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 D, Pinsky M, 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. Royal College of Physicians, London, 2012.
10. Abbott TE, Vaid N, Ip D, Cron N, Wells M, Torrance HD, Emmanuel J. . 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, van de Ven PM, Jansma EP, Nanayakkara PW. Exploring the performance of the National Early Warning Score (NEWS) in a European emergency department. *Resuscitation* 2015; 90:111-5.
12. Silcock DJ, Corfield AR, Gowens PA, Rooney KD. Validation of the National Early Warning Score in the prehospital setting. *Resuscitation* 2015; 89:31-5.

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; 58:411-9.
14. Santiano N, Young L, Hillman K, et al. Analysis of Medical Emergency Team calls comparing subjective to “objective” call criteria. *Resuscitation* 2009; 80:44-9
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. A. The objective medical emergency team activation criteria: a case-control study. *Resuscitation* 2007; 73:62-72.
17. Smith GB, Prytherch DR, Schmidt P, 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; 71:19-29.
19. Schmidt PE, Meredith P, Prytherch DR, *et al.* Impact of introducing an electronic physiological surveillance system on hospital mortality. *BMJ Qual Saf* 2014;**0**:1**-**11. doi:10.1136/bmjqs-2014-003073
20. Ball C. Critical care outreach services—do they make a difference? *Intensive Crit Care Nurs* 2002; 18:257-60
21. Hickey C, Allen M. A critical care liaison service. *Br J Anaesthesia* 1998; 81:650.
22. R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ (accessed 20/06/2015))
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* 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. A 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, Meredith P, Schmidt PE, Featherstone PI, Prytherch DR, Smith GB. 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. Triggers for emergency team activation: A multicenter assessment. *J Crit Care* 2010;25: 359.e1-7.

31. Herod R, Frost SA, Parr M, Hillman K, Aneman A. Long term trends in Medical Emergency Team activations and outcomes. *Resuscitation* 2014; 85:1083-7.

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 D. Smith G, 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, Perkins Z, Taylor A. 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; 33:619–24.

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				
Number of episodes, n (%)	53466 (51.4%)	42641 (41.0%)	7891 (7.6%)	103998 (100%)
Age at admission, mean (SD)	67.5 (18.6)	57.4 (20.6)	52.6 (20.3)	62.2 (20.4)
Male, n (%)	26910 (50.3%)	20916 (49.1%)	1584 (20.1%)	49410 (47.5%)
First adverse event in episode:				
None, n (%)	50574 (94.6%)	41786 (98.0%)	7768 (98.4%)	100128 (96.3%)
Death n (%)	1966 (3.7%)	304 (0.7%)	81 (1.0%)	2351 (2.3%)
Cardiac arrest n (%)	397 (0.7%)	71 (0.2%)	11 (0.1%)	479 (0.5%)
Unanticipated ICU admission n (%)	529 (1.0%)	480 (1.1%)	31 (0.4%)	1040 (1.0%)
Any n (%)	2892 (5.4%)	855 (2.0%)	123 (1.6%)	3870 (3.7%)
OBSERVATIONS				
Number of observations, n (%)	1264024 (56.3%)	864590 (38.5%)	117164 (5.2%)	2245778 (100%)
Vital signs values				
Pulse rate (bpm), mean (SD)	80.5 (16.5)	78.3 (14.7)	79.0 (14.8)	79.6 (15.8)
Respiration rate (bpm, 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 (%)	232591 (18.4%)	167536 (19.4%)	18965 (16.2%)	419092 (18.7%)
Conscious level				
Alert (A), n (%)	1238015 (97.9%)	860651 (99.5%)	116487 (99.4%)	2215153 (98.6%)
Responds to Voice (V), n (%)	16469 (1.3%)	3110 (0.4%)	517 (0.4%)	20096 (0.9%)
Responds to Pain (P), n (%)	8227 (0.7%)	583 (0.1%)	90 (0.1%)	8900 (0.4%)
Unresponsive (U), n (%)	1313 (0.1%)	246 (<0.1%)	70 (0.1%)	1629 (0.1%)

OBSERVATIONS FOLLOWED BY OUTCOMES (first event within 24 h)				
Death, n (%)	8686 (0.7%)	1693 (0.2%)	388 (0.3%)	10767 (0.5%)
Cardiac arrest, n (%)	1756 (0.1%)	402 (<0.1%)	73 (0.1%)	2231 (0.1%)
Unanticipated ICU admission, n (%)	3720 (0.3%)	3738 (0.4%)	249 (0.2%)	7707 (0.3%)
Any outcome, n (%)	14162 (1.1%)	5833 (0.7%)	710 (0.6%)	20705 (0.9%)

Table 2: The sensitivity and specificity of 44 MET calling criteria, and NEWS values of 3-7, for cardiac arrest, unanticipated ICU admission, death, and any of these outcomes, each within 24 h of a vital signs dataset.

	Outcome within 24 h 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 (%)
MET 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
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
NEWS											
NEWS = 3	90.5%	73.6%		68.0	73.3		82.8	73.5		85.2	73.8
NEWS = 4	82.9%	84.9%		56.0	84.6		72.3	84.7		76.1	85.1
NEWS = 5	75.2%	91.2%		43.1	90.9		61.9	91.0		66.8	91.4

NEWS = 6	65.4%	95.0%		31.5	94.8		49.9	94.9		56.0	95.2
NEWS = 7	54.2%	97.2%		22.2	97.0		37.4	97.1		44.5	97.4

FIGURES:

Figure 1

Consort diagram, outlining exclusions

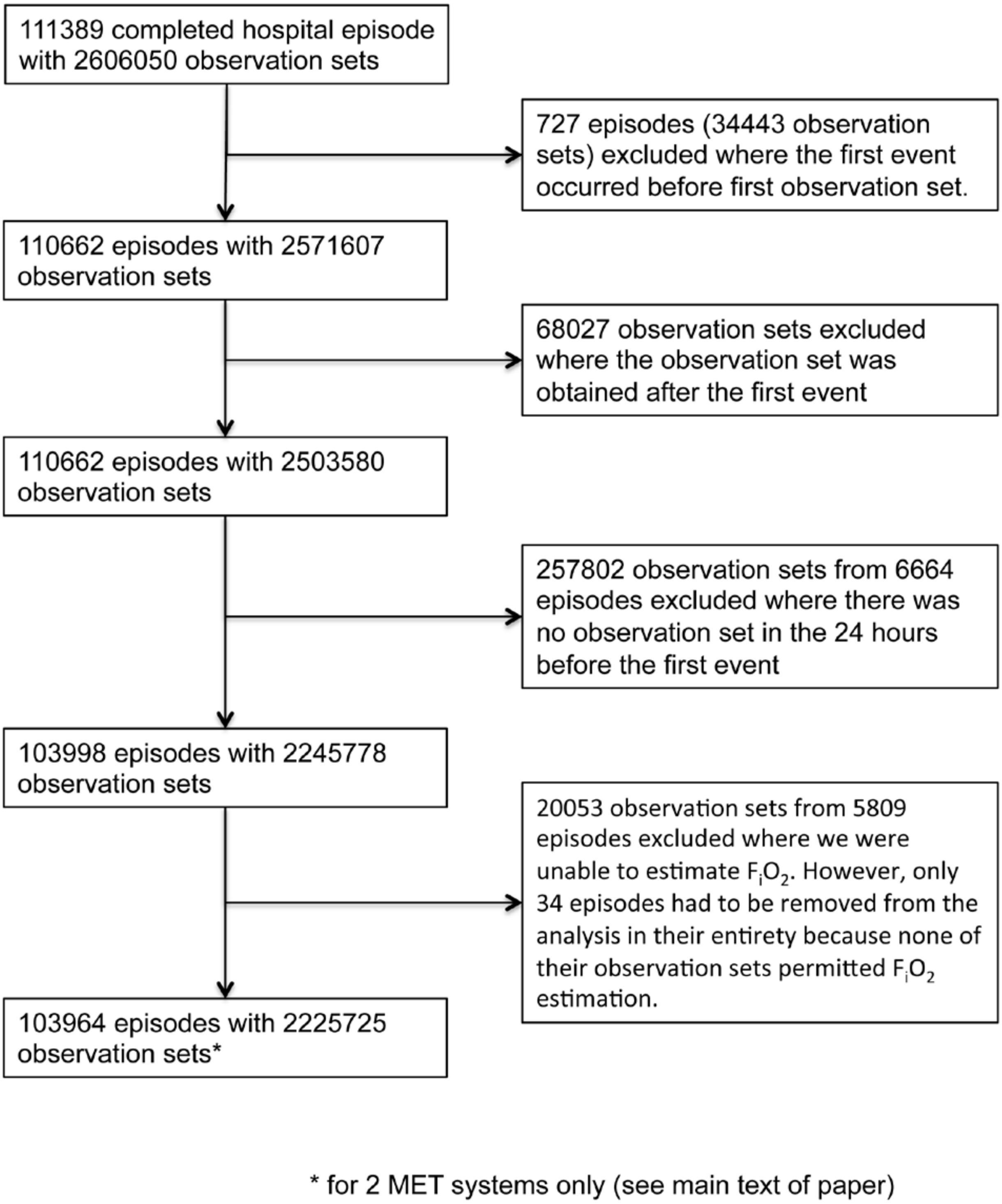


Figure 2

Comparative sensitivity and specificity (plotted as 1-specificity) of NEWS (i.e., NEWS ROC curve), and 44 sets of MET criteria to discriminate (a) death occurring within 24 h of a vital signs dataset; (b) cardiac arrest occurring within 24 h of a vital signs dataset; (c) unanticipated intensive care unit (ICU) admission occurring within 24 h of a vital signs dataset; and (d) the combined outcome of cardiac arrest, unanticipated intensive care unit (ICU) admission or death, occurring within 24 h of a vital signs dataset. Each point on the NEWS ROC curve represents a NEWS value from 0-10.

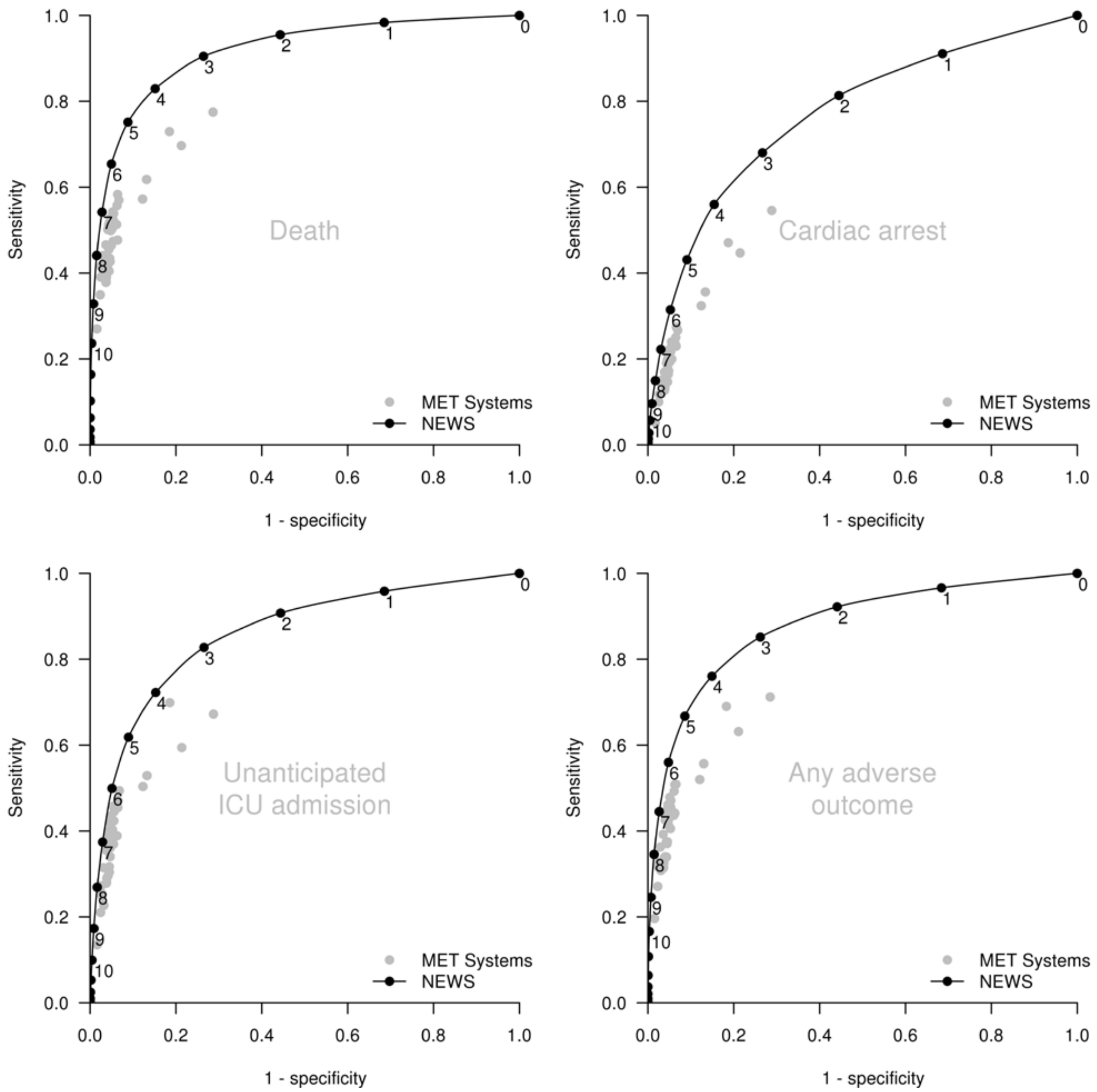
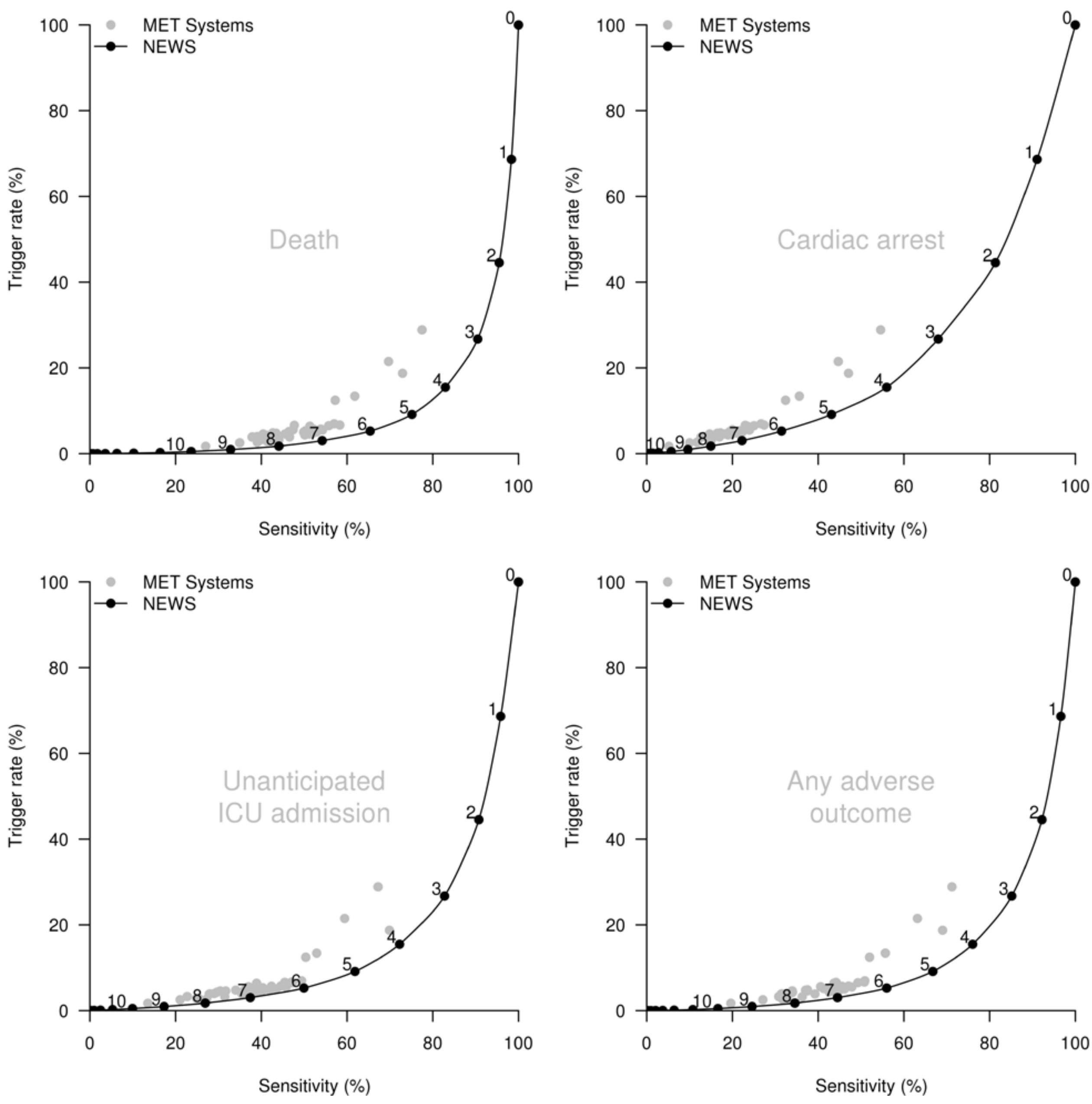


Figure 3

Efficiency curve for NEWS, with comparable data for 44 sets of MET criteria superimposed, for (a) death occurring within 24 h of a vital signs dataset; (b) cardiac arrest occurring within 24 h of a vital signs dataset; (c) unanticipated intensive care unit (ICU) admission occurring within 24 h of a vital signs dataset; and (d) the combined outcome of cardiac arrest, unanticipated intensive care unit (ICU) admission or death, occurring within 24 h of a vital signs dataset. This plots workload (trigger rate) against the sensitivity of the EWS or set of MET criteria in question. Each point on the NEWS efficiency curve represents a NEWS value from 0-10, starting with NEWS = 0 at the top right. Trigger rate = ((true positive + false positive)/(true positive + false positive + true negative + false negative)).



Supplementary digital content 1: The National Early Warning Score (NEWS)

Physiological parameters	**3**	**2**	**1**	**0**	**1**	**2**	**3**
Respiration Rate (breaths per minute)	$\leq$8		9 - 11	12 - 20		21 - 24	$\geq$25
S_pO_2 %	$\leq$91	92 - 93	94 - 95	$\geq$96			
Any supplemental oxygen?		Yes		No			
Temperature (^oC)	$\leq$35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	$\geq$39.1	
Systolic BP (mm Hg)	$\leq$90	91 - 100	101 - 110	111 - 219			$\geq$220
Heart/pulse rate (beats per minute)	$\leq$40		41 - 50	51 - 90	91 - 110	111 - 130	$\geq$131
Level of consciousness using the AVPU system				A			V, P or U

Level of consciousness: A = Alert; V = Responds to voice; P = Responds to pain; U = Unresponsive

Modified from National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS. Report of a working party. Royal College of Physicians, London, 2012.[12]

Supplementary digital content 2: A typical set of Medical Emergency Team (MET) calling criteria

Airway	Threatened
Breathing	All respiratory arrests Respiratory rate <5 per min. Respiratory rate >30 per min.
Circulation	All cardiac arrests Pulse rate <40 per min. Pulse rate >140 per min. Blood pressure < 90 mm Hg
Neurology	Sudden fall in level of consciousness (e.g., fall in GCS of >2 points) Repeated or prolonged seizures
Other	Any patient that you are seriously worried about that does not fit the above criteria

Supplementary Digital Content 3: The Medical Emergency Team Calling Criteria systems evaluated in the study and their trigger criteria.

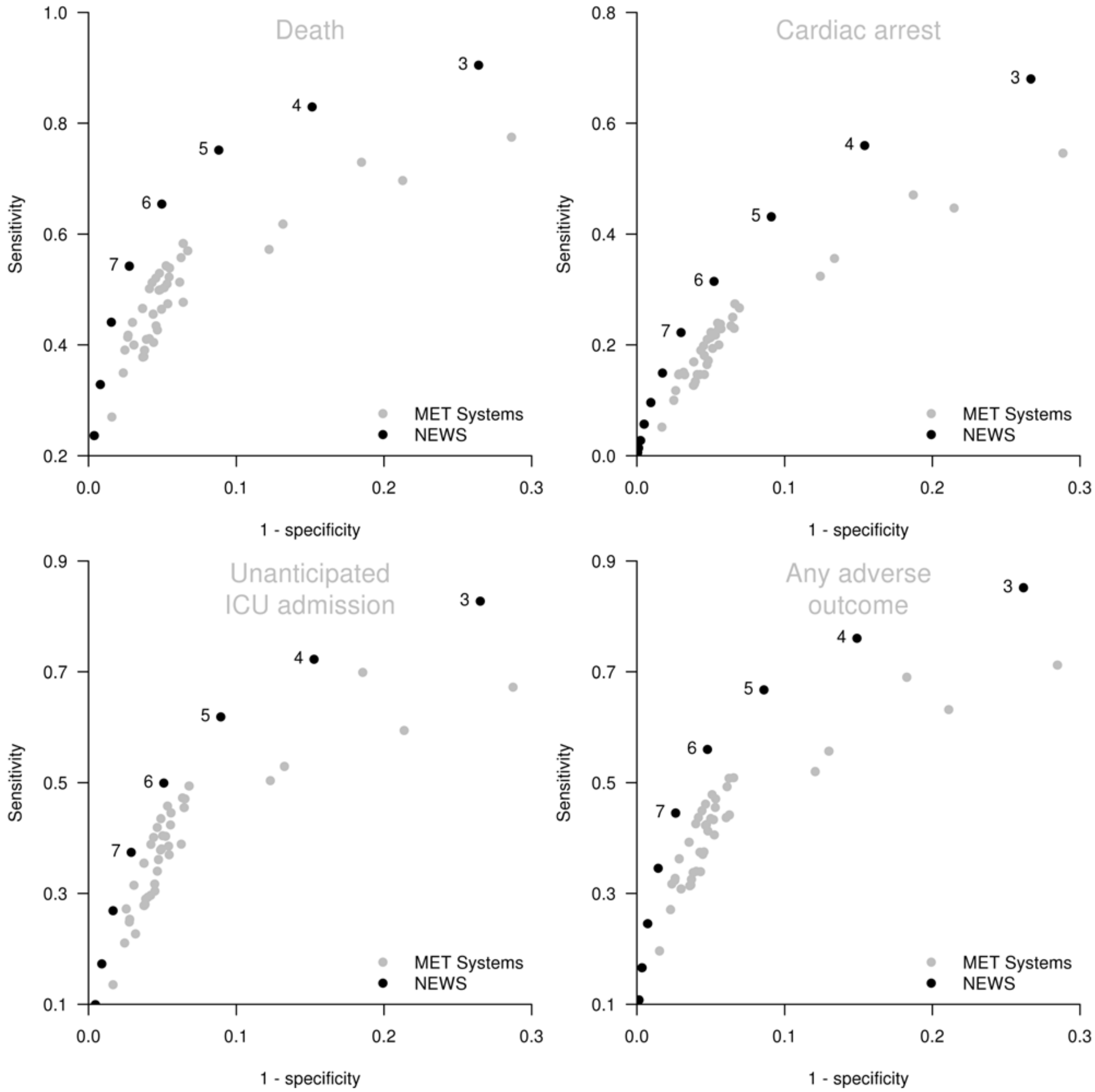
	Pulse rate beats.m^{-1}		Breathing rate breaths.m^{-1}		Systolic BP mmHg		Temperature °C		Reduced consciousness	S_pO_2 (%)	F_iO_2	Reference
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper				
Bell (MET criteria)	< 40	> 130	< 8	> 30	< 90				*	<90	>0.21	Bell MB, Konrad D, Granath F, Ekbom A, Martling C. Prevalence and sensitivity of MET-criteria in a Scandinavian University Hospital. Resuscitation 2006;70:66-73.
Bell (Extended)	< 50	> 120	≤ 10	> 28	< 100				*	<90	>0.21	
Bell (Restricted)	< 35	> 140	≤ 6	> 32	< 80				*	<90	>0.21	
Ball	< 50	> 125	< 8	> 25	< 90	> 200		> 38.0		<90	>0.35	Ball C. Critical care outreach services--do they make a difference? Intensive Crit Care Nurs 2002;18:257-60.
Parissopoulos	< 45	> 125	< 8	> 25	< 90	> 200			●	<90	>0.21	Parissopoulos S, Kotzabassaki S. Critical care outreach and the use of early warning scoring systems; a literature review. ICUs Nurs Web J 2005;21: 1-11.
Hickey	< 45	> 125	< 8	> 30	< 90				●	<90	≥0.24	Hickey C, Allen M. A critical care liaison service. British Journal of Anaesthesia 1998;81:650.
Salamonson	< 40	> 140	< 6	> 36	< 90				●	<85	≥0.21	Salamonson Y, Kariyawasam A, van Heere B, O'Connor C. The evolutionary process of Medical Emergency Team (MET) implementatio reduction in unanticipated ICU transfers. Resuscitation 2001;49:135-41.
Buist		> 130	< 6	> 30	< 90				●	<90	>0.21	Buist MD, Moore GE, Bernard SA, Waxman BP, Anderson JN, Nguyen Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: preliminary study. BMJ 2002;324:387-90.
Bellomo	< 40	> 130	< 8	> 30	< 90				●	<90	>0.21	Bellomo R, Goldsmith D, Uchino S et al. A prospective before-and-after trial of a medical emergency team. Med J Aust 2003;179:283-7.
Jones	< 40	> 130	< 8	> 30	< 90				●	<90	≥0.21	Jones D, Bates S, Warrillow S et al. Circadian pattern of activation of the medical emergency team in a teaching hospital. Crit Care 2005;9:R303-306.
Green	< 40	> 120	< 5	> 30	< 90				●	<90	>0.21	Green AL, Williams A. An evaluation of an early warning clinical marker referral tool. Intensive Crit Care Nurs 2006;22:274-82.
Harrison (Early)	< 50	> 120	< 10	> 30	< 100	> 180			●	<95	≥0.21	Harrison GA, Jacques TC, Kilborn G, McLaws M. The prevalence of recordings of the signs of critical conditions and emergency responses i hospital wards--the SOCCER study. Resuscitation 2005;65:149-57.
Harrison (Late)	< 40	> 140	< 5	> 40	< 80	> 240			●	<90	≥0.21	
Smith		≥ 100		≥ 25	≤ 95	≥ 200	≤ 35.0	≥ 38.0	●			Smith AF, Wood J. Can some in-hospital cardio-respiratory arrests be prevented? A prospective survey. Resuscitation 1998;37:133-7.
Lee	< 40	> 120	< 10	> 30	< 100	> 200	< 35.5	> 39.5	●			Lee A, Bishop G, Hillman KM, Daffurn K. The Medical Emergency Team Anaesth Intensive Care 1995;23:183-6.
Buist	< 50	> 130		> 30	< 90	> 200	< 35.0	> 40.0	●	< 85	≥0.21	Buist MD, Jarmolowski E, Burton PR, Bernard SA, Waxman BP, Anders J.Recognising clinical instability in hospital patients before cardiac arres or unplanned admission to intensive care. A pilot study in a tertiary-care hospital. Med J Aust 1999;171:22-5.
McGloin	< 60	> 120	<10	> 25	< 100	> 200	< 35.5	> 38.5	●	<90	≥0.21	McGloin H, Adam SK, Singer M. Unexpected deaths and referrals to intensive care of patients on general wards. Are some cases potentially avoidable? J R Coll Physicians Lond 1999;33:255-9.

Supplementary Digital Content 4: The sensitivity and specificity of 44 MET calling criteria, and NEWS 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 using one randomly chosen observation set from each episode.

	Outcome within 24 hours 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 (%)
MET calling criteria								
Bell (Standard)	55.8	96.5	18.6	96.2	37.9	96.3	44.0	96.8
Bell (Extended)	66.7	88.5	32.7	88.1	52.0	88.3	56.5	88.8
Bell (Restricted)	46.9	98.2	12.6	97.9	26.3	98.0	34.5	98.4
Ball	51.4	94.8	22.8	94.5	43.8	94.6	44.5	95.0
Parissopoulos	63.4	95.3	25.5	94.9	45.6	95.1	51.5	95.6
Hickey	55.8	96.2	20.4	95.8	39.0	96.0	44.6	96.4
Salamonson	46.9	97.2	11.3	96.8	29.0	97.0	35.3	97.3
Buist	55.7	96.6	18.2	96.2	37.8	96.4	43.8	96.8
Bellomo	55.8	96.5	18.6	96.2	37.9	96.3	44.0	96.8
Jones	57.1	96.2	20.1	95.8	38.3	96.0	44.9	96.4
Green	58.1	96.0	21.4	95.7	42.1	95.8	47.0	96.3
Harrison (Early)	80.6	75.2	51.4	74.8	66.1	75.0	71.1	75.5
Harrison (Late)	45.4	97.9	12.5	97.6	23.1	97.7	32.6	98.1
Smith	76.2	85.2	44.5	84.8	67.9	85.0	68.7	85.6
Lee	62.9	89.5	30.1	89.1	50.1	89.3	53.5	89.8
Buist	54.6	95.7	19.2	95.3	36.7	95.5	43.0	95.9
McGloin	74.0	78.9	43.4	78.5	58.7	78.7	64.0	79.1
de Pennington	52.4	96.4	18.2	96.0	40.1	96.2	43.0	96.6
McArthur-Rouse	48.7	96.6	15.2	96.2	35.2	96.4	39.0	96.8
Foraida	40.6	98.4	8.5	98.1	21.6	98.2	29.1	98.5
Cioffi	47.2	97.0	11.3	96.7	29.3	96.8	35.5	97.2
Holder	63.4	95.3	25.5	94.9	45.6	95.1	51.5	95.6
Buist	57.5	95.1	21.6	94.7	38.6	94.9	45.5	95.3
McQuillan	48.9	96.6	14.6	96.3	33.7	96.4	38.5	96.8
Hourihan	43.6	97.3	10.3	97.0	27.8	97.1	33.1	97.4
Bristow	45.9	96.7	12.5	96.4	30.0	96.5	35.4	96.9

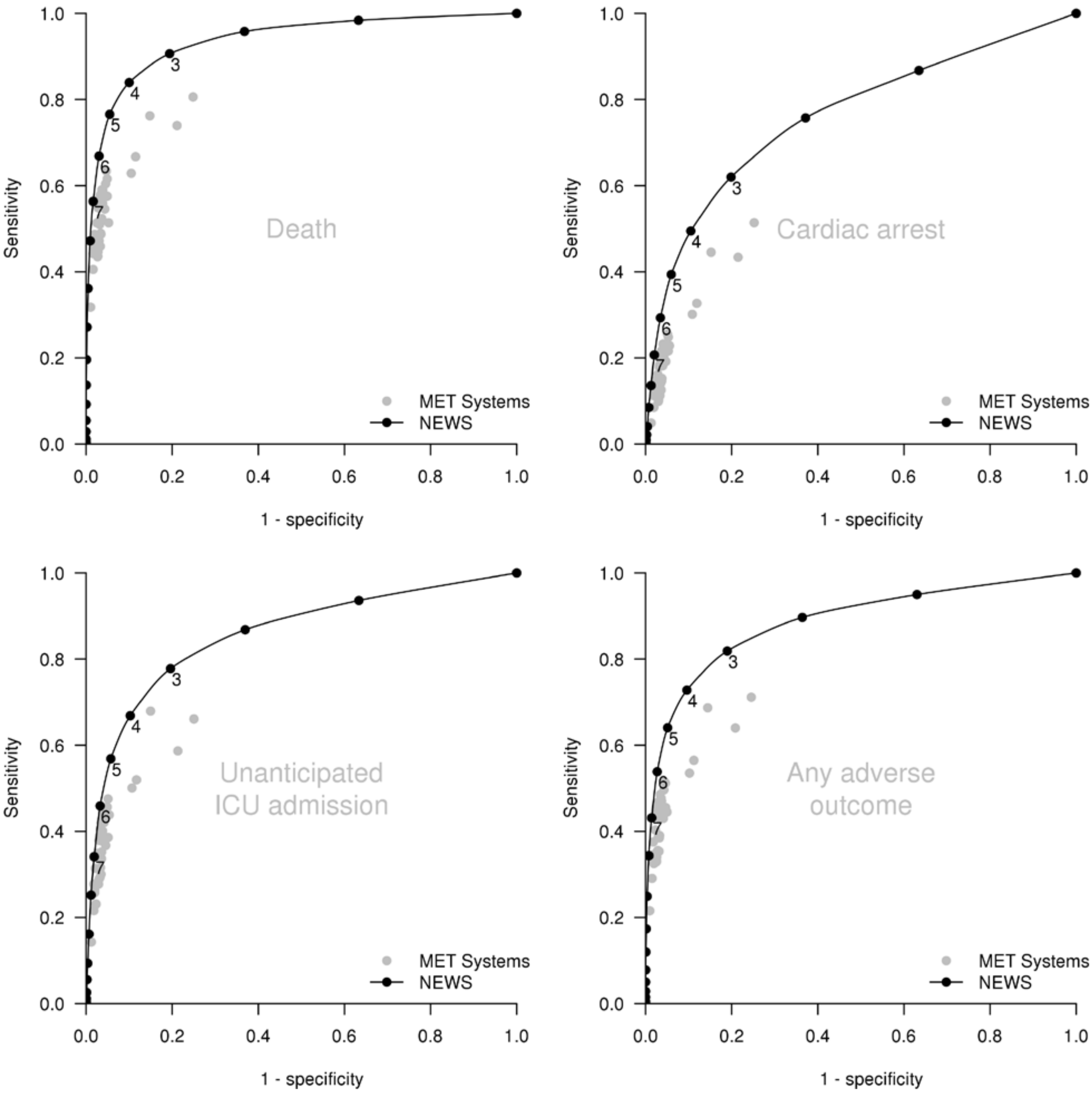
Supplementary Digital Content 5:

Close up of comparative sensitivity and specificity (plotted as 1-specificity) of NEWS, and 44 sets of MET criteria to discriminate (a) death occurring within 24 h of a vital signs dataset; (b) cardiac arrest occurring within 24 h of a vital signs dataset; (c) unanticipated intensive care unit (ICU) admission occurring within 24 h of a vital signs dataset; and (d) the combined outcome of cardiac arrest, unanticipated intensive care unit (ICU) admission or death, occurring within 24 h of a vital signs dataset. Each point on the NEWS ROC curve represents a NEWS value from 0-10.



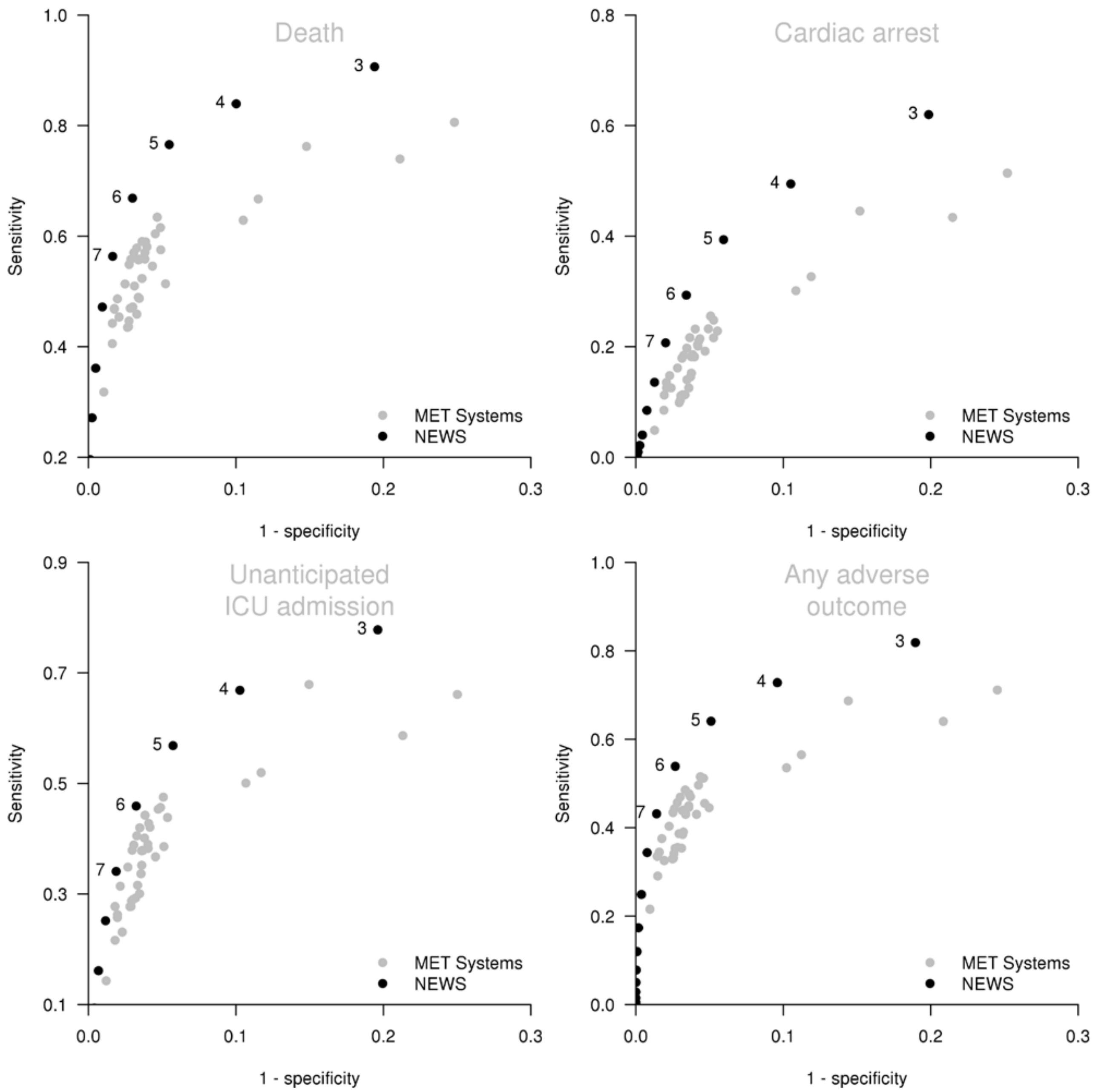
Supplementary Digital Content 6:

Comparative sensitivity and specificity (plotted as 1-specificity) of NEWS, and 44 MET calling criteria to discriminate (a) death occurring within 24 hours of a vital signs dataset; (b) cardiac arrest occurring within 24 hours of a vital signs dataset; (c) unanticipated ICU admission occurring within 24 hours of a vital signs dataset; and (d) the combined outcome of cardiac arrest, unanticipated ICU admission or death, occurring within 24 hours of a vital signs dataset when using 10,000 random sample sets (one set from each episode).



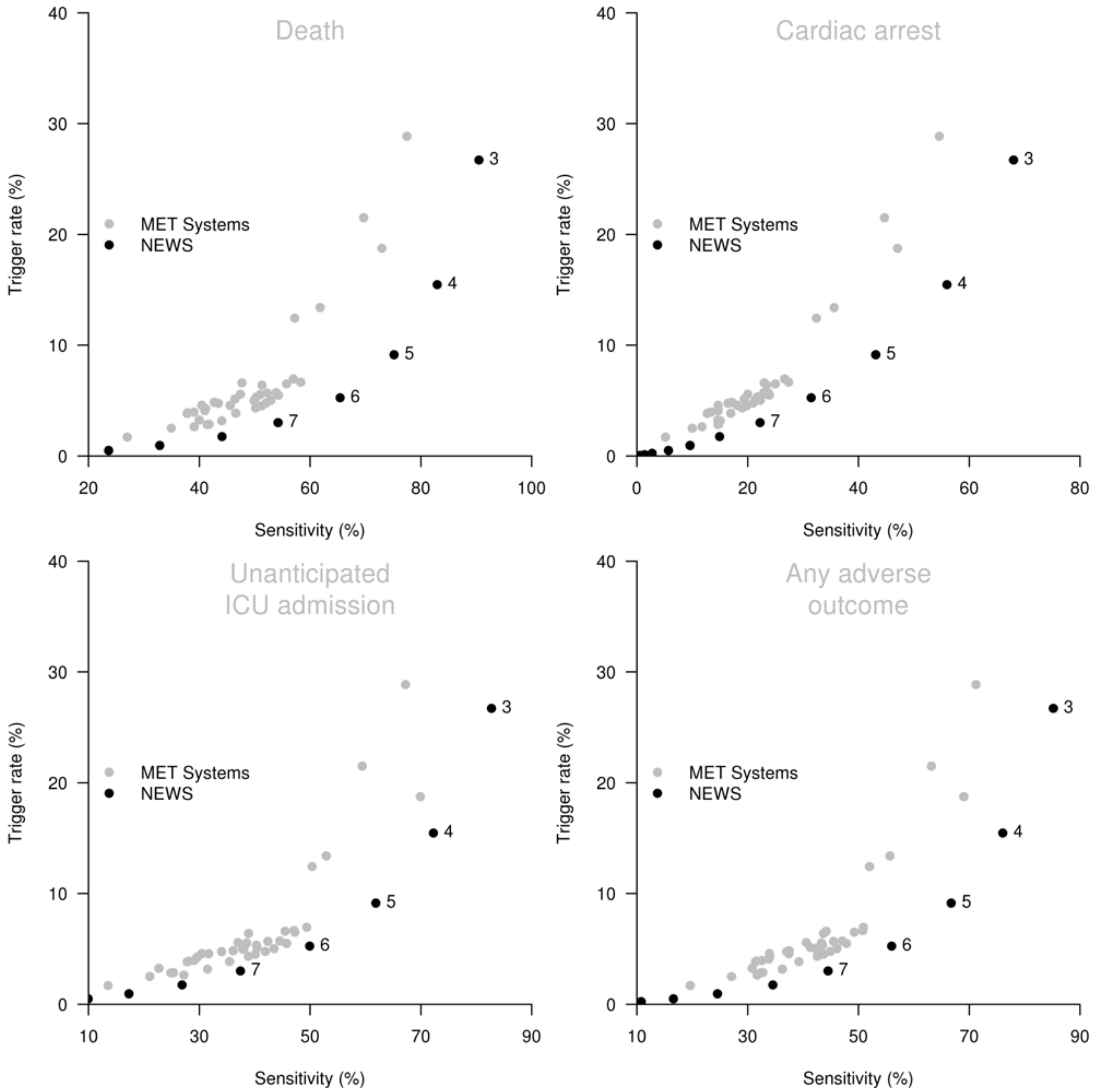
Supplementary Digital Content 7:

Close up of comparative sensitivity and specificity (plotted as 1-specificity) of NEWS, and 44 MET calling criteria to discriminate (a) death occurring within 24 hours of a vital signs dataset; (b) cardiac arrest occurring within 24 hours of a vital signs dataset; (c) unanticipated ICU admission occurring within 24 hours of a vital signs dataset; and (d) the combined outcome of cardiac arrest, unanticipated ICU admission or death, occurring within 24 hours of a vital signs dataset when using 10,000 random sample sets (one set from each episode).



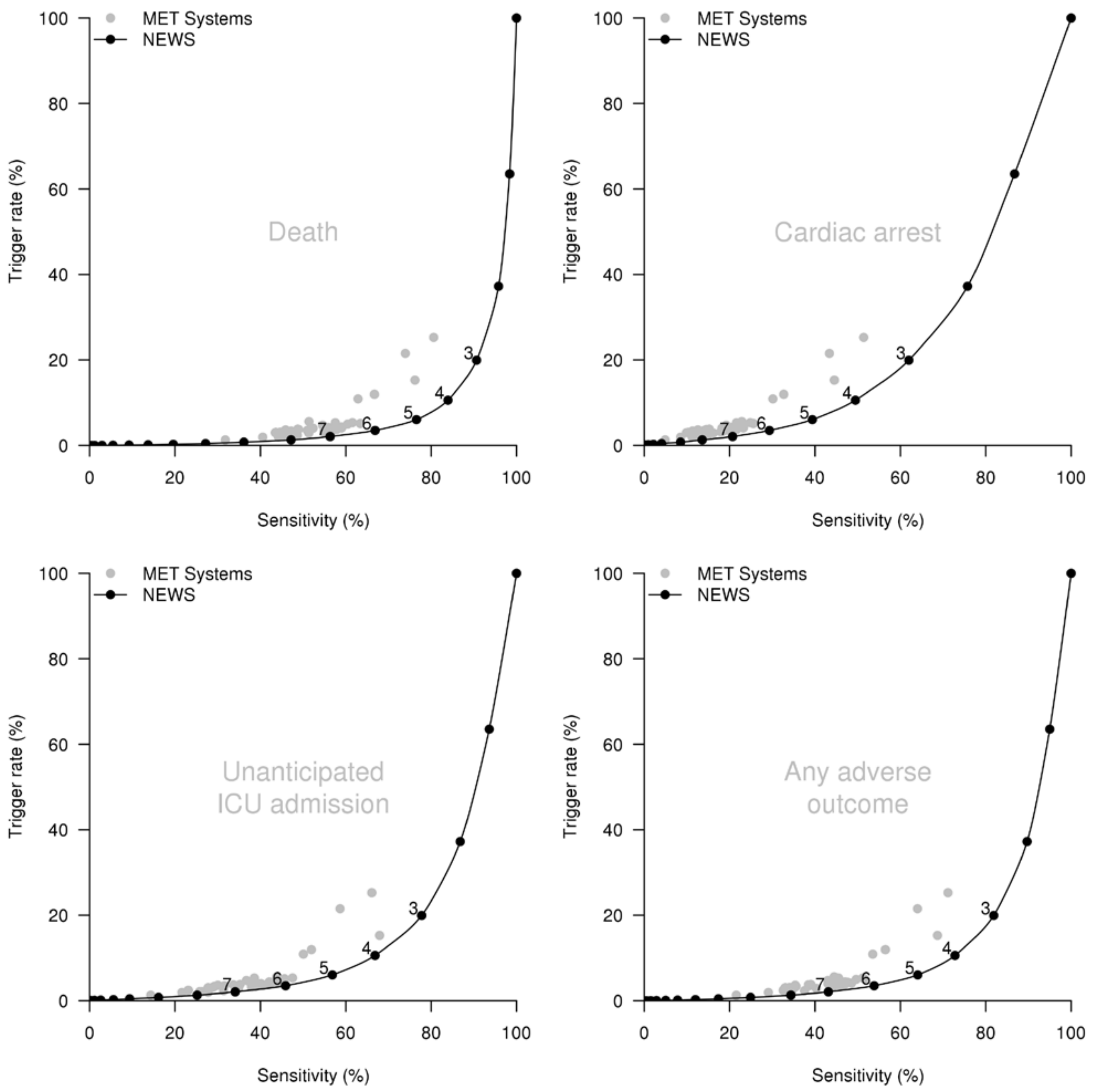
Supplementary Digital Content 8:

Close up of efficiency curve for NEWS, with comparable data for 44 sets of MET criteria superimposed, for (a) death occurring within 24 hours of a vital signs dataset; (b) cardiac arrest occurring within 24 hours of a vital signs dataset; (c) unanticipated ICU admission occurring within 24 hours of a vital signs dataset; and (d) the combined outcome of cardiac arrest, unanticipated ICU admission or death, occurring within 24 hours of a vital signs dataset.



Supplementary Digital Content 9:

Efficiency curve for NEWS, with comparable data for 44 sets of MET criteria superimposed, for (a) death occurring within 24 hours of a vital signs dataset; (b) cardiac arrest occurring within 24 hours of a vital signs dataset; (c) unanticipated ICU admission occurring within 24 hours of a vital signs dataset; and (d) the combined outcome of cardiac arrest, unanticipated ICU admission or death, occurring within 24 hours of a vital signs dataset. This plots workload against the sensitivity of the EWS or set of MET criteria when using 10,000 random sample sets (one set from each episode).



Supplementary Digital Content 10:

Close up of efficiency curve for NEWS, with comparable data for 44 sets of MET criteria superimposed, for (a) death occurring within 24 hours of a vital signs dataset; (b) cardiac arrest occurring within 24 hours of a vital signs dataset; (c) unanticipated ICU admission occurring within 24 hours of a vital signs dataset; and (d) the combined outcome of cardiac arrest, unanticipated ICU admission or death, occurring within 24 hours of a vital signs dataset. This plots workload against the sensitivity of the EWS or EWS or set of MET criteria when using 10,000 random sample sets (one set from each episode).