View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Portsmouth University Research Portal (Pure)

Can binary early warning scores perform as well as standard early warning scores for discriminating a patient's risk of cardiac arrest, death or unanticipated intensive care unit admission?

Dr. Stuart Jarvis PhD.[a,b]

Mrs. Caroline Kovacs, BSc[a]

Dr. Jim Briggs, BA, DPhil[a]

Dr. Paul Meredith, PhD[c]

Dr. Paul E Schmidt, MRCP, B.Med.Sc, MBA[c]

Dr. Peter I Featherstone, FRCP[c]

Professor David R Prytherch, PhD, MIPEM, CSci[a,c]

Professor Gary B Smith, FRCA, FRCP[d]

[a]Centre for Healthcare Modelling& Informatics, University of Portsmouth, Portsmouth, UK

[b]Department of Health Sciences, University of York, York, UK

[c]Portsmouth Hospitals NHS Trust, Portsmouth, UK

[d]School of Health & Social Care, University of Bournemouth, Bournemouth, UK

Correspondence to:
Professor G B Smith, FRCA, FRCP,
Centre of Postgraduate Medical Research & Education (CoPMRE),
School of Health & Social Care, 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

Word count = 2921 Figures =3 Tables = 2; Supplementary files = 3

Keywords: Monitoring; Physiologic; Vital signs; Failure to rescue; Hospital Rapid Response Team

ABSTRACT

Introduction:

Although the weightings to be summed in an early warning score (EWS) calculation are small, calculation and other errors occur frequently, potentially impacting on hospital efficiency and patient care. Use of a simpler EWS has the potential to reduce errors.

Methods:

We truncated 36 published 'standard' EWSs so that, for each component, only two scores were possible: 0 when the standard EWS scored 0 and 1 when the standard EWS scored greater than 0. Using 1,564,153 vital signs observation sets from 68,576 patient care episodes, we compared the discrimination (measured using the area under the receiver operator characteristic curve – AUROC) of each standard EWS and its truncated 'binary' equivalent.

Results:

The binary EWSs had lower AUROCs than the standard EWSs in most cases, although for some the difference was not significant. One system, the binary form of the National Early Warning System (NEWS), had significantly better discrimination than all standard EWSs, except for NEWS. Overall, Binary NEWS at a trigger value of 3 would detect as many adverse outcomes as are detected by NEWS using a trigger of 5, but would require a 15% higher triggering rate.

Conclusions:

The performance of Binary NEWS is only exceeded by that of standard NEWS. It may be that Binary NEWS, as a simplified system, can be used with fewer errors. However, its introduction could lead to significant increases in workload for ward and rapid response team staff. The balance between fewer errors and a potentially greater workload needs further investigation.

BACKGROUND

Early warning scores (EWS) are now extensively used to identify deteriorating ward patients, either to prevent intensive care unit (ICU) admission or facilitate it early.[1,2] Additionally, EWSs provide an evaluation of the likelihood of impending cardiac arrest or death.[2] EWSs use measurements of vital signs (e.g., pulse rate, blood pressure, breathing rate) as their basis. Each vital sign component is typically awarded a weighted score in the range 0 to 3 (although the upper limit can differ), based on the derangement of patients' vital signs variables from agreed "normal" ranges. Most EWS calculations are currently undertaken manually.

Traditionally, an EWS has up to seven components. For example, the Royal College of Physicians of London (RCPL) National Early Warning System (NEWS) contains pulse rate, breathing rate, systolic blood pressure, temperature, S_pO_2, the inspired gas and the patient's conscious level.[3] Several other EWSs contain only a subset of these components and one, the Cardiac Arrest Risk Triage (CART) score,[4] uses diastolic rather than systolic blood pressure.

Typically, when the aggregate EWS exceeds pre-determined levels, clinical staff are advised to increase vital signs monitoring, involve more experienced staff or call a rapid response team (e.g. outreach or medical emergency team). Although the weightings to be summed in an EWS are small, calculation and other errors occurfrequently.[5-11] These may impact on hospital efficiency and patient care – escalating care and monitoring for patients that do not require it, or failing to escalate care for those that do. Use of a simpler EWS has the potential to reduce errors.[6] It may therefore be beneficial to develop simplified EWSs.

We hypothesised that, for the outcomes traditionally used to assess the performance of EWS, the identification of normality – and of deviation from normality – in vital signs is more important than the level of derangement. Therefore, we investigated the effectiveness of

EWS systems that have only two possible scores, 0 (normal, i.e., low risk) or 1 (abnormal, i.e., increased risk), for each vital sign. The simplified EWSs, hereinafter referred to as binary EWSs, are based on previously existing standard EWSs. To use such an EWS, staff would merely have to count the number of components in which a score of 1 was received.

METHOD

Ethical Committee Approval

The study is 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.

Study site

Portsmouth Hospitals NHS Trust (PHT) is a NHS District General Hospital on the South Coast of England, handling ~140,000 admissions per year in ~1200 inpatient beds on a single site. It has ~5500 staff and provides all acute services except burns, spinal injury, neurosurgical and cardiothoracic surgery to ~540,000 of the local population.

Vital signs test results database and its development

We constructed a database of vital signs collected from all adult patients admitted to PHT on or after 25/05/2011 and discharged on or before 31/12/2012. We excluded data from patients aged <16 years at hospital admission and patients discharged alive on the day of admission. Vital signs data were recorded in real-time at the bedside using handheld electronic equipment running VitalPAC software.[12,13] Each full set of vital signs measurements contained: pulse rate, breathing rate, systolic and diastolic blood pressure, temperature, S_pO_2, the inspired gas (e.g., oxygen or air) at the time of S_pO_2 measurement, and the patient's conscious level. Conscious level was recorded as alert (A), responds to voice (V), responds to pain (P) or unresponsive (U). For EWSs that use the Glasgow Coma Scale, the scores were converted to the AVPU system (GCS 15 = A; GCS 14 = V; GCS 13-9 = P; GCS ≤ 8 = U) as previously described.[1]Observation sets for which one or more of the vital signs measurements were absent or physiologically impossible (i.e., recorded in error) were excluded.

Outcomes

We studied the following outcomes: death, cardiac arrest and unanticipated intensive care unit (ICU) admission, each within 24 h of an observation set. Patient outcomes were identified using the hospital's patient administration system (for death), and its cardiac arrest and ICU admission databases. We used precedence rules so that, when multiple adverse outcomes occurred within 24 h of an observation set, only the first was counted (e.g. a cardiac arrest, followed by an ICU admission, followed by death – all within 24 h of an observation set – was recorded as cardiac arrest only).

Development of binary EWSs

To develop the binary EWSs, we truncated 36 published 'standard' EWS – the 34 previously compared by Smith et al,[1,2] plus CART[4] and the Centiles EWS.[14] The EWSs used are summarised in Table S1 in the supplementary information. For each component in each EWS, we assigned a score of 0 in the corresponding binary EWS if the score for that component in the standard EWS would be 0. If the score for a component in the standard EWS would be greater than 0, the score for that component in the binary EWS would be 1. As an example, NEWS and its binary equivalent ("Binary NEWS") are presented in Table 1.

Table 1:

The National Early Warning Score (NEWS) and its binary equivalent, Binary NEWS

Vital sign	NEWS								Binary NEWS		
	3	2	1	0	1	2	3		1	0	1
Respiration rate (breaths per minute)	≤8		9-11	12-20		21-24	≥25		<12	12-20	>20
S_pO_2 (%)	≤91	92-93	94-95	≥96					<96	≥96	
Any supplemental oxygen				No		Yes				No	Yes
Temperature (ºC)	≤35.0		35.1-36.0	36.1-38.0	38.1-39.0	≥39.1			<36.1	36.1-38.0	>38.0
Systolic blood pressure (mmHg)	≤90	91-100	101-110	111-219			≥220		<111	111-219	>219

Pulse rate (beats per minute)	≤40		41-50	51-90	91-110	111-130	≥131		<51	51-90	>90
Level of consciousness				A			V, P, U			A	V, P, U

Assessment of EWS performance

The ability of an EWS to discriminate a patient's risk of an adverse outcome can be measured using the area under the receiver operator characteristic curve (AUROC). This represents the probability that a randomly selected observation that was followed, within 24 hours, by an adverse outcome had a higher score under an EWS than a randomly selected observation that was not followed, within 24 hours, by an adverse outcome.[15] We calculated the AUROCs for the 36 standard EWSs and the corresponding 36 binary EWSs for the outcomes of death, cardiac arrest, unanticipated ICU admission and any of those outcomes within 24 h of the observation set. We calculated the AUROCs using (a) all observation sets in the dataset and (b) using 10,000 sample sets, each with one observation set per episode of patient care, selected at random. We took both approaches to test whether any lack of independence between observation sets for the same patient might bias the results. Previous work has shown that such effects can be important when an EWS includes age,[15] as was the case for some EWSs included in this study.

When using all observations, we calculated a 95% confidence interval for the AUROCs and assessed the significance of differences in AUROCs using the methods set out by DeLong *et al*.[16] When using 10,000 sample sets, we calculated an AUROC for each sample set and reported the mean AUROC and the 2.5 and 97.5 centiles of the AUROCs as the 95% confidence interval.

We also analysed the performance of the best performing EWS and binary EWS using the EWS efficiency curve, described by Prytherch *et al.*[17] This plots the triggering rate (i.e., workload) against sensitivity. In calculating the efficiency curve, we again used 10,000

sample sets, each with one observation set from each episode of patient care, selected at random.

Finally, we calculated some summary measures. Using 10,000 sample sets, each with one observation set per episode of patient care, we calculated sensitivity, positive predictive value, specificity and negative predictive value at triggering values that would give similar triggering rates for the best performing standard and binary EWS. Using all the observations in the dataset, we also calculated (i) the percentage of total observations that would result in escalation; (ii) the percentage of total episodes of care for which there would be at least one escalation; and (iii) the percentage of adverse outcomes for which at least one escalation would have occurred in the 24 hours before the adverse outcome (i.e., for which there would have been warning and some chance to intervene in the adverse outcome) and (iv) the mean number of patients triggering each day under each system.

Data analysis tools

All data manipulation was performed using Microsoft® Visual FoxPro 9.0. All analyses were undertaken in R version 3.02.[18]

RESULTS

In the study period, there were 68,576 discharges of 46,944 unique patients admitted on or after 25/05/2011 and discharged on or before 31/12/2012, where the patient was aged ≥16, the patient was not discharged alive on the day of admission and at least one full set of valid vital signs observations was recorded. Of these episodes, 32,720 (48%) were for male patients and the mean age at admission was 62.5 years (standard deviation: 20.5 years). Associated with these episodes of care were 1,564,143 valid, full observation sets (mean 22.8 observation sets per episode). 7,637 observation sets (0.49%) in 1697 episodes (2.47%) for 1,697 patients were followed by death within 24 h (irrespective of any other outcome occurring before). 6,380 observation sets (0.41%) in 1,441 episodes (2.10%) for 1,441 patients had death as the first outcome within 24 h; 1671 observations (0.11%) in 326 episodes (0.48%) for 325 patients had cardiac arrest as the first outcome within 24 h; 4,975 observations (0.32%) in 615 episodes (0.90%) for 606 patients had unanticipated ICU admission as the first outcome within 24 h. In total, 13,026 observation sets (0.83%) were followed by one or more adverse outcomes within 24 h and 2,301 episodes (3.36%) for 2,276 patients had an adverse outcome recorded within 24 hours of an observation set.

The discrimination of each EWS (as measured by AUROC) for the outcomes of death, cardiac arrest, unanticipated ICU admission and any of those three outcomes within 24 h of an observation set is shown in Figure 1. With the exception of Bakir's EWS[19] and (for some outcomes) CART,[4] the binary systems perform less well than the standard systems, although in some cases not significantly so. Binary NEWS performs significantly better than the standard versions of all other EWSs for all outcomes. The 95% confidence intervals for Binary NEWS and MEWS with age[20] overlap for cardiac arrest (Figure 1), but the AUROCs are significantly different (p-value < 0.001using the method of DeLong et al).[17] Analysis using one observation set per episode, chosen at random, showed similar trends (Figure 2). However, there were fewer statistically significant differences between standard and binary EWSs, and between Binary NEWS and standard versions of others.

The best performing standard and binary EWSs – NEWS and Binary NEWS – are compared for their efficiency against the study outcomes in Figure 3. Binary NEWS offers slightly lower efficiency (greater number of triggers for intervention in a given number of adverse outcomes) than NEWS. A score of 3 in Binary NEWS is closest to the standard triggering score of 5 in NEWS.

Table 2 compares the performance of NEWS using a trigger value of 5 and Binary NEWS using a trigger value of 3. NEWS would generate a trigger in 10.20% of observations and 29.01% of episodes, and would trigger in the 24 hours before an adverse outcome in 92.57% of cases. For Binary NEWS, 11.78% of observation sets would result in a trigger (35.27% of episodes) and 93.05% of episodes ending in an adverse outcome would trigger in the 24 hours preceding that outcome. Binary NEWS has a higher 'detection' rate, although not significantly so. In our hospital, NEWS would have generated triggers in, on average, 118 patients each day while Binary NEWS would have generated triggers for 145 patients. Using one observation per episode, Binary NEWS has performance not significantly different to that of NEWS as measured by sensitivity and negative predictive value. NEWS performs better as measured by positive predictive value and specificity.

Table 2:

Performance measurements for NEWS and Binary NEWS, based on triggering at a score of 5 or greater for NEWS and 3 or greater for Binary NEWS.

Data used	**Performance measure**	**NEWS**	**Binary NEWS**
All observations (multiple observations per episode)	% of observation sets that trigger (95% CI)	10.20 (10.15 - 10.25)	11.78 (11.73 - 11.83)
	% of episodes that trigger at least once (95% CI)	29.01 (28.67-29.35)	35.27 (34.91-35.62)
	% of episodes with an adverse outcome for which the EWS would trigger in the 24 hours before the outcome (95% CI)	92.57 (92.12 – 93.02)	93.05 (92.61 – 93.48)

	Mean number of unique patients triggering each day (SD)	118 (20)	145 (24)
One observation per episode, selected at random (using 10,000 sample sets)	Sensitivity: % of observations followed by an adverse outcome that trigger (95% CI)	69.69 (67.14 - 72.20)	67.71 (65.14 - 70.26)
	Positive predictive value: % of triggering observations that are followed by an adverse outcome (95% CI)	11.76 (11.16 – 12.37)	9.62 (9.11 - 10.13)
	Specificity: % of observations not followed by an adverse outcome that do not trigger (95% CI)	94.16 (94.03-94.30)	92.89 (92.74 – 93.04)
	Negative predictive value: % of non-triggering observations that are not followed by an adverse outcome (95% CI)	99.64 (99.61-99.68)	99.61 (99.58-99.65)

DISCUSSION

The results show that simplified binary EWSs can offer useful discrimination of a patient's risk of adverse outcomes. For all outcomes, except cardiac arrest, the majority of binary EWSs have AUROCs over 0.7. The best performing binary EWS, Binary NEWS, offers better discrimination than the standard versions of other EWS for all outcomes studied, except for standard NEWS. Its discrimination is lower than that of the standard NEWS, as may be expected from the binary categorization of the continuous variables.[22] However, the performance gap between them is small.

On the other hand, Binary NEWS offers slightly lower efficiency (a greater number of triggers for intervention in a given number of adverse outcomes) than NEWS. Overall, Binary NEWS at a trigger value of 3 would trigger in the 24 hours preceding an adverse outcome more often than NEWS using a trigger of 5 (although the difference is not significant), but it would require a 15% higher triggering rate (in terms of triggers per observation taken; a 22% higher triggering rate in terms of episodes having at least one trigger and a 23% higher rate in terms of unique patients triggering each day). This is clearly of concern, as it would increase the number of reviews by clinicians with competencies in the assessment of acute illness[3] and might also increase the workload of the rapid response team. Although Binary NEWS appears to at least match NEWS in terms of the number of adverse outcomes that would be preceded by a trigger, it should also be noted that there may be differences in the timing and number of triggers. Earlier detection may, within limits, be useful to give more time for interventions. A greater number of triggers before an adverse outcome may provide more chances to intervene, but may also result in later triggers being ignored if earlier ones were considered to be false alarms.

There are clear strengths to our study - it uses a large database from over 18 months of completed, hospital-wide inpatient admissions and all necessary vital signs variables were collected simultaneously in a standardised manner for all observations sets, using an

electronic data collection system.[12,13] However, there are also weaknesses. As with all modelled evaluations of EWSs, our analyses do not take account of any interventions that may have occurred in response to triggers (NEWS was used in the hospital during the study period) and that may have changed clinical outcomes. Also, we have compared only aggregate scores for the EWSs, despite the RCPL's guidance for deployment of NEWS also recommending triggering when any vital sign measurement scores 3.[3] However, scores of 3 do not exist in Binary NEWS, making such triggering and a direct comparison between NEWS and Binary NEWS impossible in this respect. Also of relevance is the finding that triggering on single extreme values of a vital sign observation, as suggested for NEWS by the RCPL, is not necessarily advantageous.[23] Finally, ours is a single centre study and similar results might not be obtained from data collected in other institutions. Therefore, an external validation exercise is necessary.

There is evidence of error with the use of early warning scoring systems,[6,8-11] and of simpler 'calling criteria' offering advantages in ease of use and increased accuracy.[6] In particular, higher scores in NEWS are associated with higher error rates.[11] However, there are no comparative data on their respective utility in clinical settings. Further investigation into the rates and implications of errors using standard EWSs, and the potential to reduce error with a binary EWS are required to determine whether binary systems might offer better performance and safety in clinical practice. In hospitals where EWSs are calculated electronically,[12,13,,24,25] a simplified EWS would offer no obvious benefit, as electronic systems remove errors due to inaccurate calculation or inaccurate assignment of score. Introducing Binary NEWS in such settings would likely merely increase staff workload. However, a simplified system, such as Binary NEWS, may have significant utility in hospitals relying on paper vital signs charts and manual early warning score calculations.

It is notable that for Bakir's EWS, which includes a weighting for age,[20] the binary version outperformed its 'parent' standard version. We hypothesise that high scores (up to 9)

awarded for age in the standard Bakir EWS did not accurately represent risk associated with age in our data, either by assigning too much weight or by assigning it at the wrong ages. The binary system reduced the impact of age on the aggregate scores and so improved performance. Binary systems, by their simplistic nature, suffer less from any over-fitting to particular data that may adversely affect their performance in other populations.

It is possible that the simple approach to defining binary systems adopted here (score 0 if original EWS scores 0; score 1 if original EWS scores > 0) is not optimal and that adjustment of the boundary between scores of 0 and 1 for each vital sign may result in a binary EWS with discriminative power greater than those considered here. However, the boundaries in NEWS have already been validated using an automated computer based system in which scores were assigned so that a 0 score in a component represented below average risk (where average risk is defined as the overall risk in the studied population) and a score of 1 or greater represented above average risk.[26] The boundaries in Binary NEWS may already represent a near optimal description of normality and abnormality for the case mix of patients studied.

SUMMARY

This study supports the hypothesis that it is the definition of normality – and therefore abnormality – in vital signs measurements that provides EWSs with most of their discriminatory power. Simplified binary EWSs based on this principle offer useful discrimination of a patient's risk of adverse outcomes. A simplified two-level version of NEWS (Binary NEWS) is outperformed only by its full ‘original’ version and is simpler to compute, possibly reducing errors in its use. However, its introduction could lead to significant increases in workload for ward and rapid response team staff. Future research is required to confirm our findings and to evaluate the operational impact, both in terms of ease of use, workload and patient safety, of using a binary EWS system.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the efforts of the medical, nursing and administrative staff at Portsmouth Hospitals NHS Trust who collected the data used in this study. Dr. Stuart Jarvis takes responsibility for the integrity and the accuracy of the data analysis.

COMPETING INTERESTS

VitalPAC, the software system used to collect the vital signs in this study, is a collaborative development of The Learning Clinic Ltd (TLC) and Portsmouth Hospitals NHS Trust (PHT). At the time of the study, PHT had a royalty agreement with TLC to pay for the use of PHT intellectual property within the VitalPAC product. PM, DP, PF and PS are employed by PHT. GS was an employee of PHT until 31/03/2011. PS, PF, and the wives of GS and DP, are minority shareholders in TLC. GS, DP, and PS are unpaid research advisors to TLC, and have received reimbursement of travel expenses from TLC for attending symposia in the UK. JB's research has previously received funding from TLC through a Knowledge Transfer Partnership.

GS was a member of the Royal College of Physicians of London's National Early Warning Score (NEWS) Development and Implementation Group. DP assisted the Royal College of Physicians of London in the analysis of data validating NEWS. SJ and CK have no conflicts of interest.

FUNDING

None.

REFERENCES

1. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted 'track and trigger' systems. Resuscitation 2008;77:170:179.
2. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation 2013;84:465-470.
3. Royal College of Physicians, London National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS. Report of a working party. London, 2012.
4. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalised patients on the wards. Chest 2013;143:1758-1765.
5. 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.
6. 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.
7. Mohammed MA, Hayton R, Clements G, Smith G, Prytherch D. Improving accuracy and efficiency of early warning scores in acute care. Br J Nurs2009;18:18-24.
8. Wilson SJ, Wong D, Clifton D et al. Track and trigger in an emergency department: an observational evaluation study. Emerg Med J 2010;30:186–191.
9. Edwards M, Van Leuvan C, Mitchell I. Modified Early Warning Scores: inaccurate summation or inaccurate assignment of score? Crit Care 2010;14:S88.
10. Smith AF, Oakey RJ. Incidence and significance of errors in a patient 'track and trigger' system during an epidemic of Legionnaires' disease: retrospective case note analysis. Anaesthesia 2006;61:222–8.

11. Kolic I, Crane S, McCartney S, Perkins Z, Taylor A. Factors affecting response to National Early Warning Score (NEWS). Resuscitation 2015 Feb 20. pii: S0300-9572(15)00076-3. doi: 10.1016/j.resuscitation.2015.02.009. [Epub ahead of print]
12. 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.
13. 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.
14. Tarassenko L, Clifton D, Pinsky M, Hravnak M, Woods J, Watkinson P. Centile-based early warning scores derived from statistical distributions of vital signs. Resuscitation 2011;82:1013-1018.
15. Jarvis SW, Kovacs C, Briggs J et al. Are observation selection methods important when comparing early warning score performance? Resuscitation 2015;90:1-6
16. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.
17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–845.
18. Prytherch D, Smith GB, Schmidt PE, Featherstone PI. ViEWS - towards a national Early Warning Score for detecting adult inpatient deterioration. Resuscitation 2010; 81:932–937.
19. 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 16 April 2014
20. Bakir A, Duckitt R, Buxton-Thomas R. A simple physiological scoring system for medical in-patients derived by modeling hospital mortality data. Poster presentation Intensive Care Society State of the Art Meeting 2005. London: Intensive Care

Society.

21. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM 2001;94:521-6.
22. Altman D, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332:1080.
23. Jarvis SW, Kovacs C, Briggs JS et al. Aggregate National Early Warning Score (NEWS) values are more important than high scores fora single vital signs parameter for discriminating the risk of adverse outcomes. Resuscitation 2015;87:75-80
24. Jones S, Mullally M, Ingleby S, Buist M, Bailey M, Eddleston JM. Bedside electronic capture of clinical observations and automated clinical alerts to improve compliance with an Early Warning Score protocol, Crit Care and Resusc 2011;13:83-88.
25. Bellomo R, Ackerman M, Bailey M et al. Vital Signs to Identify, Target, and Assess Level of Care Study (VITAL Care Study) Investigators. A controlled trial of electronic automated advisory vital signs monitoring in general hospital wards. Crit Care Med 2012;40:2349-61.
26. Badriyah T, Briggs JS, Meredith P wt al. Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS). Resuscitation 2014;85:418-423.

LEGENDS FOR FIGURES:

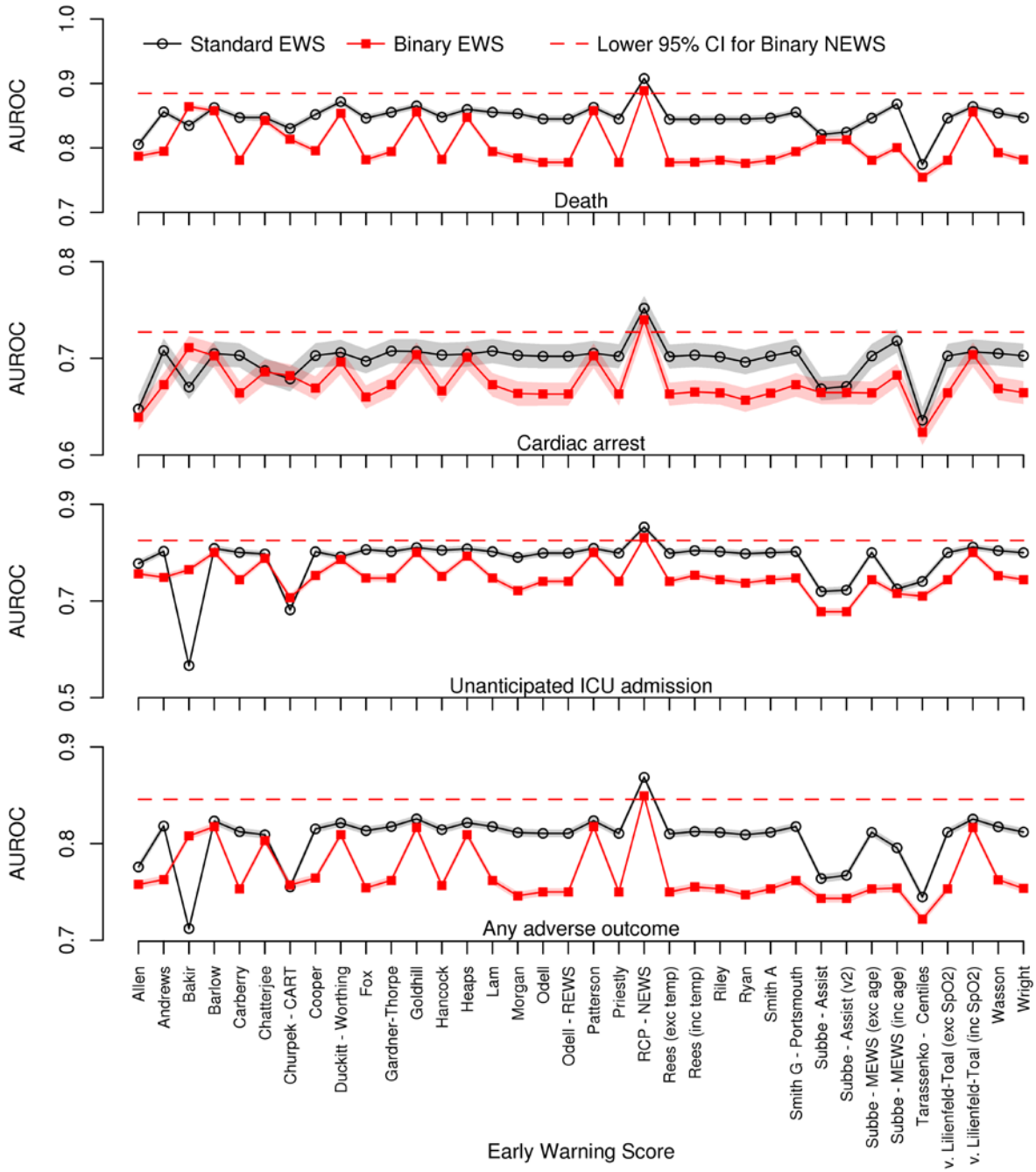

Figure 1: Performance of 36 EWSs and their binary equivalents. Discrimination of patient risk of death, cardiac arrest, unanticipated ICU admission and any of those three adverse outcomes within 24 h of an observation set is measured using the area under the receiver operator characteristic curve (AUROC). Shaded regions indicate extent of 95% confidence interval. The dashed red line shows the lower confidence interval for Binary NEWS. Full data are in Table S2 in the supplementary material.

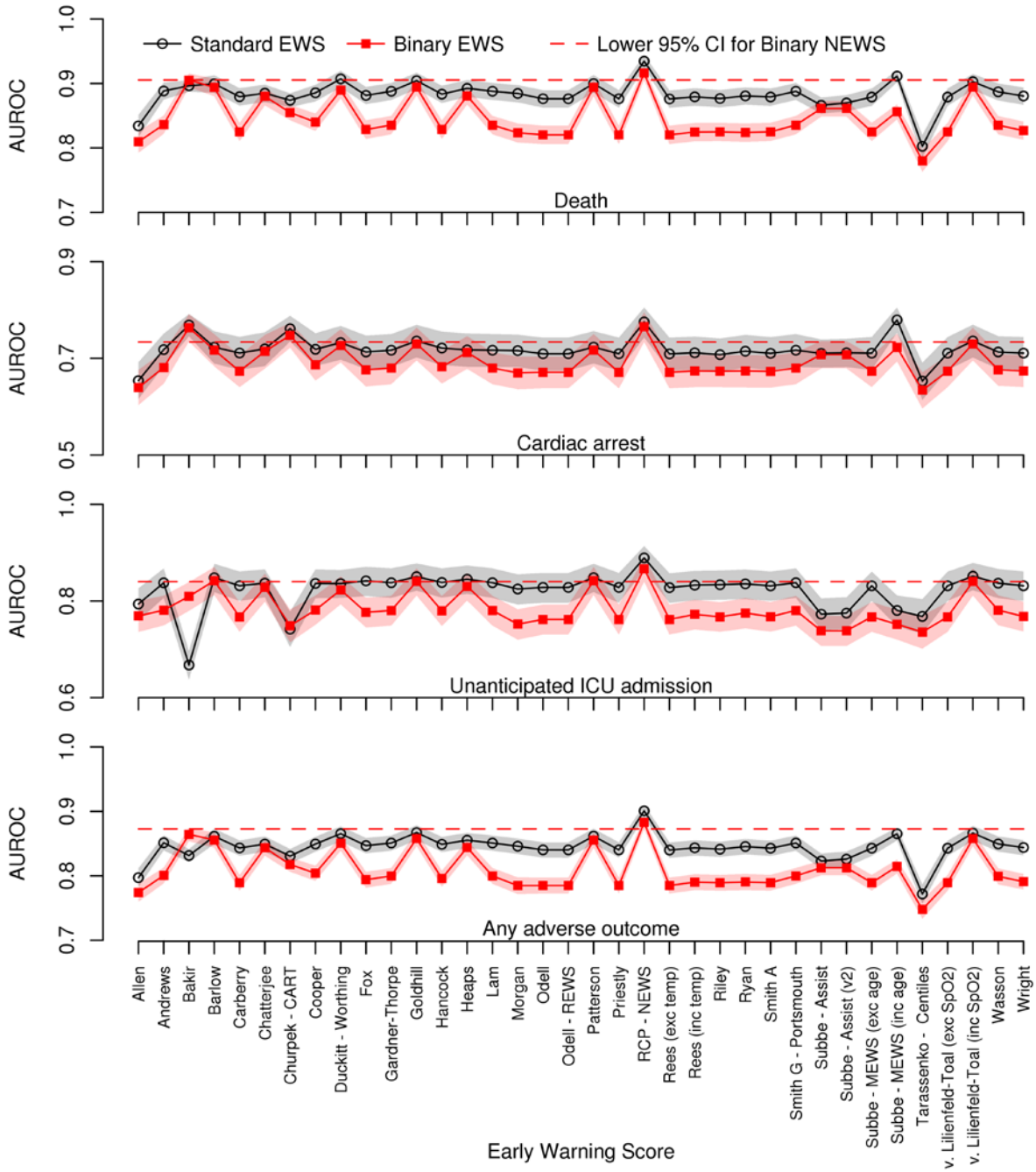

Figure 2: Performance of 36 EWSs and their binary equivalents assessed using one randomly chosen observation set from each episode. Discrimination of patient risk of death, cardiac arrest, unanticipated ICU admission and any of those three adverse outcomes within 24 h of an observation set is measured using the area under the receiver operator characteristic curve (AUROC). Shaded regions indicate extent of 95% confidence interval. The dashed red line shows the lower confidence interval for Binary NEWS. Full data are in Table S3 in the supplementary material.

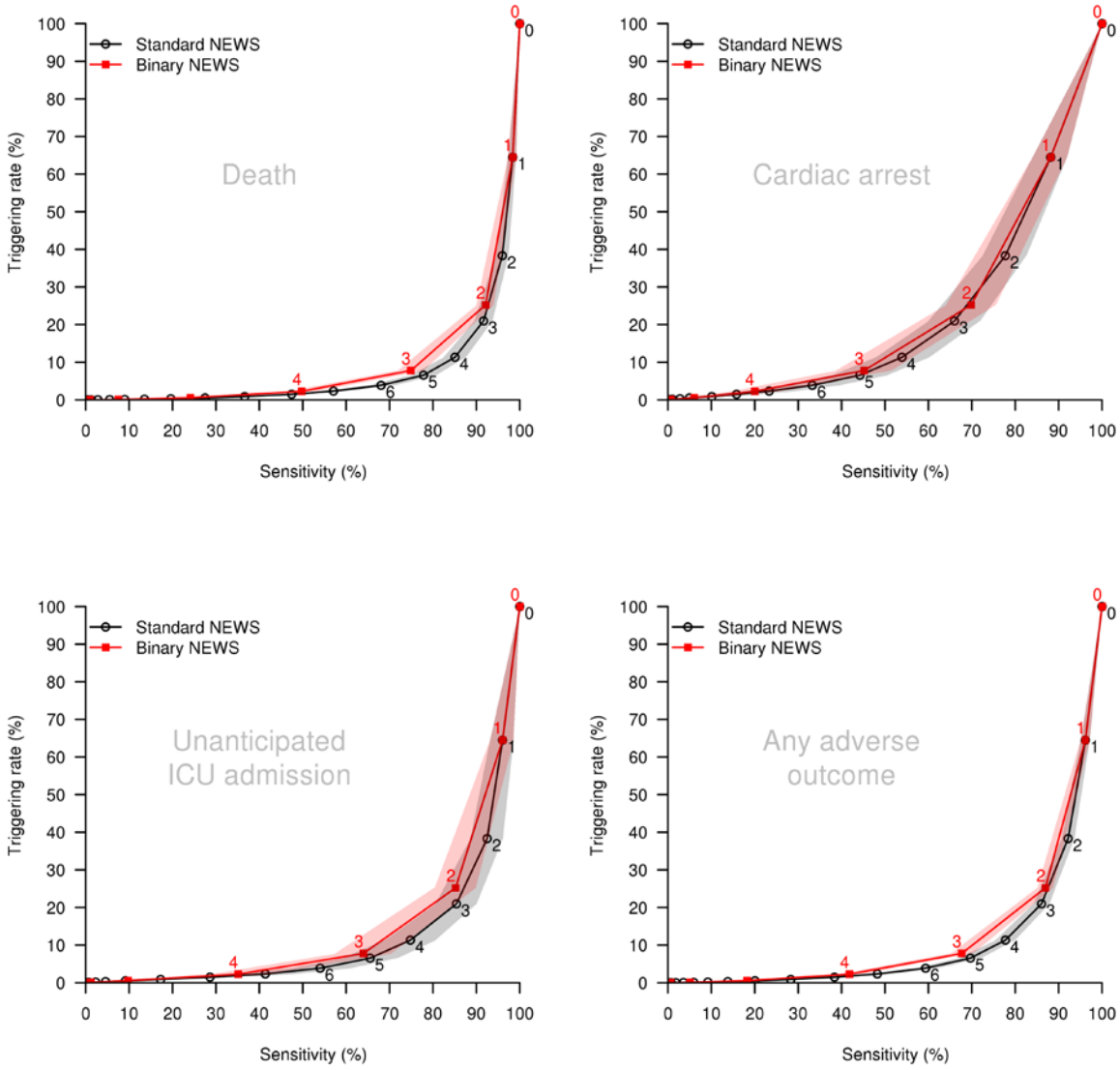

Figure 3: The workload (triggering rate) associated with the opportunity to intervene in a given number of cases experiencing any adverse outcome (sensitivity) for the best performing standard and binary EWSs (NEWS and Binary NEWS). The numbers adjacent to points indicate the associated triggering threshold and shaded areas indicate the 95% confidence interval. The curve is only defined at triggering points and lines connecting these are provided only as a visual aid.