

## **Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS)**

Dr. Tessy Badriyah, MSc., BSc., PhD Graduate, Centre for Healthcare Modelling and Informatics, University of Portsmouth, Portsmouth, UK

Dr. James S Briggs, BA, DPhil, CEng, CITP, MBCS, FHEA, Principal Lecturer & Director, Centre for Healthcare Modelling and Informatics, University of Portsmouth, Portsmouth, UK

Dr. Paul Meredith, PhD, Data Analyst, TEAMS centre, Portsmouth Hospitals NHS Trust, Portsmouth, UK

Dr. Stuart W Jarvis, PhD, Research Assistant, Centre for Healthcare Modelling and Informatics, University of Portsmouth, Portsmouth, UK

Dr. Paul E Schmidt, MRCP, B.Med.Sc, Consultant In Acute Medicine, Portsmouth Hospitals NHS Trust & Senior Lecturer, School of Health Sciences and Social Work, University of Portsmouth, Portsmouth UK

Dr. Peter I Featherstone, FRCP, Consultant In Acute Medicine, Portsmouth Hospitals NHS Trust & Senior Lecturer, School of Health Sciences and Social Work, University of Portsmouth, Portsmouth UK

Professor David R Prytherch, PhD, MIPEM, CSci, Clinical Scientist, TEAMS centre, Portsmouth Hospitals NHS Trust & Visiting Professor, Centre for Healthcare Modelling and Informatics, University of Portsmouth, Portsmouth, UK

Professor Gary B Smith, FRCA, FRCP, Visiting Professor, 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),  
The 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](mailto:gbsresearch@virginmedia.com)

Word count = 2747

Number of references = 7

Figures = 3;

Tables =2

### **Keywords**

Early warning scores; Track and trigger systems; Scoring Systems; Medical Emergency Team; Outreach; Risk prediction

## **Abstract**

**Aim of study:** To compare the performance of a human-generated, trial and error-optimised early warning score (EWS), i.e., National Early Warning Score (NEWS), with one generated entirely algorithmically using Decision Tree (DT) analysis.

**Materials and Methods:** We used DT analysis to construct a decision-tree EWS (DTEWS) from a database of 198755 vital signs observation sets collected from 35585 consecutive, completed acute medical admissions. We evaluated the ability of DTEWS to discriminate patients at risk of cardiac arrest, unanticipated intensive care unit admission or death, each within 24 hours of a given vital signs observation. We compared the performance of DTEWS and NEWS using the area under the receiver-operating characteristic (AUROC) curve.

**Results:** The structures of DTEWS and NEWS were very similar. The AUROC (95% CI) for DTEWS for cardiac arrest, unanticipated ICU admission, death, and any of the outcomes, all within 24 h, were 0.708 (0.669–0.747), 0.862 (0.852–0.872), 0.899 (0.892–0.907), and 0.877 (0.870–0.883), respectively. Values for NEWS were 0.722 (0.685–0.759) [cardiac arrest], 0.857 (0.847–0.868) [unanticipated ICU admission], 0.894 (0.887–0.902) [death], and 0.873 (0.866–0.879) [any outcome].

**Conclusions:** The decision-tree technique independently validates the composition and weightings of NEWS. The DT approach quickly provided an almost identical EWS to NEWS, although one that admittedly would benefit from fine-tuning using clinical knowledge. We believe that DT analysis could be used to quickly develop candidate models for disease-specific EWSs, which may be required in future.

## **INTRODUCTION**

In 2010, our group developed a novel early warning scoring system – ViEWS (VitalPAC Early Warning Score) – for use in the early recognition and response to patient deterioration.<sup>1</sup> ViEWS was constructed using an iterative, pragmatic, ‘trial and error’ approach, with the cut-offs for its scoring bands being deliberately adjusted to maximise its ability to predict in-hospital death within 24 hours of a vital signs observation. Members of the Royal College of Physicians of London’s National Early Warning Score Design and Implementation Group (NEWSDIG) used ViEWS as the basis for the newly announced National Early Warning Score (NEWS),<sup>2</sup> making only minor changes to the weightings described in ViEWS. We have since shown that the ability of NEWS to discriminate patients at risk of cardiac arrest, unanticipated intensive care unit admission or death within 24 hours of a NEWS value is superior to that of 33 other Early Warning Scores (EWSs).<sup>3</sup>

We wished to compare the structure and discriminative performance of the human-generated, trial and error-optimised EWS, NEWS, with an EWS generated entirely algorithmically using Decision Tree (DT) analysis, against the combined outcome of cardiac arrest, unanticipated intensive care unit admission or death within 24 hours of a given vital signs observation.

## **METHOD**

Local research ethics committee approval was obtained for this study from the Isle of Wight, Portsmouth and South East Hampshire Research Ethics Committee.

### **Vital signs database and its development**

A database of vital signs, previously used for the development of VIEWS<sup>1</sup> and NEWS<sup>3</sup> was employed for this study, thereby permitting direct comparison of the performances of DTEWS and NEWS. This database had been developed from vital signs collected from completed consecutive admissions (aged  $\geq 16$  years) to the Medical Assessment Unit (MAU) between 8 May 2006 and 30 June 2008.

Data from patients who were discharged from hospital before midnight on the day of admission were excluded. The MAU is the common entry point for all general medical emergency patients, with the exception of those transferred directly on admission to the hospital's critical care areas.

As a routine part of clinical care, MAU staff used personal digital assistants (PDA) running the VitalPAC software<sup>4</sup> to collect each patient's vital signs data. Each time a vital signs measurement was made the following data were collected in VitalPAC: date/time of observation set; pulse rate; systolic and diastolic blood pressure; breathing rate; body temperature; neurological status using either the Alert-Verbal-Painful-Unresponsive (AVPU) scale or Glasgow Coma Score (GCS); peripheral oxygen saturation ( $S_pO_2$ ); and a record of the inspired gas (i.e., air or oxygen) being breathed by the patient at the time of  $S_pO_2$  measurement. Where conscious level had been assessed using a GCS instead of AVPU, we converted the GCS value to an 'AVPU equivalent', previously described.<sup>1</sup> Vital signs observation sets were not available once the patient was transferred from MAU.

### **Development of DTEWS using DT analysis**

We used the *classregtree* function in MATLAB 2011b version 7.13 (© 1994-2012 The MathWorks Inc. Natick, Massachusetts, USA) to generate DTEWS from our vital signs database. Decision tree (DT) analysis is a data mining classification technique for building decision trees by recursive splitting or partitioning of datasets into homogenous groups. The data are initially divided into two groups such that the impurity in each group (as measured against an outcome - in our case death within 24 hours of a given vital signs measurement - and one or more covariates, using Gini's diversity index<sup>5</sup>) is minimised. Each of these groups (or child nodes) is then subdivided into a further two groups, also

determined to minimise their impurity. The size of the tree is limited by the *mergeleaves* parameter. *classregtree* computes the error risk in each node and, if the sum of error risks in two child nodes is greater to or equal to the error risk of their parent, the leaves are merged back into the parent. In addition, by default, no split is attempted on any node containing fewer than ten data points. Here, 'error risk' does not refer to risk of death within 24 hours, but is the product of the node error and the node probability (the term used in MATLAB's documentation is 'risk' but here we use 'error risk' to differentiate from the risk of death within 24 hours). The node error is a measure of the fraction of the node members that, if the tree were reapplied to the training data, would have been wrongly classified (in our case, for a node in which most data corresponded to no death within 24 hours, it would be the fraction for which death did occur within 24 hours). The node probability is the fraction of the original data that meet the criteria for that node. Our tree modelling strategy assessed the following covariates independently – pulse rate; systolic blood pressure; breathing rate; body temperature; neurological status using the Alert-Verbal-Painful-Unresponsive (AVPU) scale; and peripheral oxygen saturation ( $S_pO_2$ ); and the inspired gas (i.e., air or oxygen) being breathed by the patient at the time of  $S_pO_2$  measurement.

For each terminal leaf of the DT, the MATLAB software gives the number of cases in it and the number of outcomes, so that a risk of the outcome (i.e., in-hospital death) can be generated. An example of the decision tree and the nodes obtained for pulse rate are shown in Figure 1. Following the classic approach used by most EWS, we chose to develop DTEWS with a '0, 1, 2 and 3' weighting system. The mean incidence of death was 1.0058% so the risk bands were set as follows: where the risk generated by the DT analysis was  $<$  mean risk, a value of 0 was ascribed; if the risk was  $\geq$  mean risk and  $<$  2 times mean risk, a value of 1 was ascribed; if the risk was  $\geq$  2 times mean risk and  $<$  3 times mean risk, a value of 2 was ascribed; and if the risk was  $\geq$  3 times mean risk, a value of 3 was ascribed. This process allows the concatenation of leaves that may have the same or similar relative risks. This process was repeated for each of the seven covariates to produce DTEWS.

The decision tree and the nodes obtained for temperature produced one range (temperature  $<$  32.1°C) that showed an incidence of death = 0. By the weighting rules described above, this would mean that this node should be assigned a weighting of 0. However, using logic and on the basis that

temperature values of 35.9 - 36.0°C are weighted as 2 and those of  $\leq 35.8^\circ\text{C}$  are assigned a weighting of 3, we chose to assign temperatures of  $< 32.1^\circ\text{C}$  a weighting of 3 points.

### **Evaluation of DTEWS**

We then evaluated the ability of DTEWS to discriminate patients at risk of cardiac arrest, unanticipated intensive care unit admission or death, each within 24 hours of a given vital signs observation. Where it occurred, patient death and its timing was identified from the patient administration system (PAS); confirmed cardiac arrests and their timings were identified using the hospital cardiac arrest database; and unanticipated ICU admission and its timing was identified from the ICU admission database. Patients may have suffered more than one of the three outcomes within 24 hours. We applied precedence rules for the presence of multiple outcomes attached to each observation set. Where multiple outcomes occurred, the first outcome was the outcome used in the analysis. For example, if a patient suffered a cardiac arrest followed by unanticipated ICU admission and death, this was defined as a cardiac arrest for the purposes of analysis. Alternatively, if someone was admitted to ICU and then suffered a cardiac arrest followed by death, this was defined as an ICU admission, and so on. To provide an additional, more clinically useful outcome than death, cardiac arrest, and unanticipated ICU admission alone, we also analysed for the presence of any of these outcomes (that is death or cardiac arrest or unanticipated ICU admission) within 24 hours.

The ability of DTEWS to discriminate between those suffering and those not suffering an adverse outcome at 24 hours post vital signs observation was evaluated using the area under the receiver-operating characteristic (AUROC) curve.<sup>6</sup> AUROC analysis was performed using SPSS v20. The minimum AUROC possible is 0.5 and is the value that would be expected if the model was no better than chance at predicting mortality. Reasonable discrimination is indicated by AUROC values of 0.700 to 0.800 and good discrimination by values exceeding 0.800. In entering repeated observation sets from the same patient episode into the analysis, we have made the implicit assumption that the observation sets are independent of each other. The AUROC values for DTEWS were then compared to those for NEWS, as published previously.<sup>2</sup> Additionally, as a measure of the relative number of “triggers” that would be generated at different values of DTEWS, we also produced an “EWS efficiency curve”<sup>1</sup> for DTEWS and NEWS, plotted together, using the combined outcome of cardiac arrest, unanticipated intensive care unit admission or death within 24 hours.

### **Assessment of independence of data**

In order to test the hypothesis that observations and EWS values might not be independent of each other, we used R 3.01<sup>7</sup> to analyse the data using 10,000 sample sets of data, each set comprising one randomly chosen observation set per patient episode in a sample, and we found AUROC values similar to those presented in our manuscript.

### **Data analysis**

All data manipulation was performed using Microsoft® Visual FoxPro 9.0. All analyses were undertaken in SPSS v20.

## RESULTS

The database of 198755 vital signs datasets (104379 sets from females) was obtained from 35585 consecutive, completed acute medical admissions. The mean (median) ages of the patients were 67.7 (72.6) years (male 65.9 (69.7); female 69.4 (75.5)). Of the 198755 observation sets, 1999 were followed by death within 24 hours, irrespective of any of the other outcomes. When we applied the precedence rules as described, of the 198,755 observation sets, 199 were followed by cardiac arrest, 1161 by unanticipated ICU admission, 1789 by death and 3149 by any of the outcomes, all within 24 h. The mean ( $\pm$  SD) vital signs in the dataset, details of AVPU categories and the percentage of admissions breathing air at the time of  $S_pO_2$  measurement are shown in Table 1.

The distribution of DTEWS values, compared to those for NEWS<sup>3</sup>, and their relationships to the primary outcome, is shown in Figure 2. A comparison of the structures of DTEWS and NEWS is shown in Table 2. The AUROC (95% CI) for DTEWS for cardiac arrest, unanticipated ICU admission, death, and any of the outcomes, all within 24 hours, were 0.708 (0.669–0.747), 0.862 (0.852–0.872), 0.899 (0.892–0.907), and 0.877 (0.870–0.883), respectively. This compares with AUROC values of 0.722 (0.685–0.759), 0.857 (0.847–0.868), 0.894 (0.887–0.902), and 0.873 (0.866–0.879) for cardiac arrest, unanticipated ICU admission, death, and any of the outcomes, all within 24 h, respectively, for NEWS.<sup>3</sup>

Figure 3 shows the “EWS efficiency curve”<sup>1</sup> for DTEWS and NEWS for the outcome of cardiac arrest, unanticipated ICU admission or death within 24 hours of the observation set. This compares the number of “triggers” that are generated at different values of DTEWS and NEWS. It demonstrates that the curves for DTEWS and NEWS are almost identical in shape and position; however the individual EWS values are positioned differently along their efficiency curves. For instance, the detection of ~83% of those who will die within 24 hours of a given EWS value requires a response to only 25% of either DTEWS or EWS values. However to achieve this, the trigger point for DTEWS must be 5, whilst that for NEWS must be 4.

Finally, when using one randomly chosen observation set per episode in a sample, the AUROC values for the combined outcome of cardiac arrest, unanticipated ICU admission or death, each within

24 hours of a given vital signs measurement, for NEWS and DTEWS were 0.879 (0.871-0.887) and 0.884 (0.876-0.892), respectively.

## DISCUSSION

The VitalPAC Early Warning Score, ViEWS, was developed using an iterative, pragmatic, ‘trial and error’ approach, optimised to discriminate in-hospital death within 24 hours of a vital signs observation. The National Early Warning Score, NEWS, resulted from modifications by NEWSDIG to ViEWS. Hence, NEWS is essentially a system derived by humans without intrinsic computerised modelling. Our newly developed early warning score, DTEWS, was designed algorithmically using a computer. Remarkably, the structures of DTEWS and NEWS are very similar, despite the different processes behind their development. The two major differences are the weightings assigned to low breathing rates and high systolic blood pressure values. Although this could be interpreted as indicating that is conceivable that such values are unrelated to clinical outcomes, it is most likely that they are due to their low prevalence in the patient group studied and, consequently, our vital signs dataset. Computer-derived scores such as DTEWS can only reflect the prevalence of a given value, whereas NEWS, being human-generated using clinical insight, can reflect not only the prevalence but also the clinical importance of a given value.

Unsurprisingly, given the structural similarity of DTEWS and NEWS, the performance of the two systems is also virtually identical, when used to discriminate patients at risk of the combined outcome of cardiac arrest, unanticipated intensive care unit admission or death, each within 24 hours of a given vital signs dataset. The almost identical shape and position of the efficiency curves for the DTEWS and NEWS (Figure 3) also indicates that the “workload” that would be generated by each EWS in order to detect a given number of outcomes is similar, although the trigger EWS value may be different for the two systems. This difference in trigger level arises because DTEWS assigns 3 points when supplemental Oxygen is used (as did ViEWS) but NEWS attributes only 2 points.

Two conclusions arise from these observations. First, the decision-tree technique provides an independent method for validating the composition weightings and ranges of NEWS. Second, the trial and error approach to optimisation used to develop ViEWS and NEWS was very time consuming. However, the DT approach quickly provided an almost identical EWS, although one that admittedly would benefit from fine-tuning using clinical knowledge. The development of new disease-specific early warning scores, e.g., for chronic obstructive pulmonary disease, may be required in future. We

believe that DT analysis could be a useful tool in the rapid development of candidate EWSs for use in such specific clinical conditions.

This study uses the same underlying data as used in the original ViEWS paper and the analysis of the NEWS.<sup>1,3</sup> Therefore, it has the same strengths and weaknesses. The strengths include that all vital signs variables were collected simultaneously in a standardised manner as part of the clinical process and that each vital signs observation set contained all of the necessary vital signs variables. The weaknesses include the exclusion of medical admissions that were admitted directly to critical care areas of the hospital, and the fact that DTEWS and NEWS were evaluated using only vital signs observation sets from patients in the MAU. We also used repeated observation sets from the same patient episode in the analysis, making the assumption that the observation sets are independent of each other. Indeed, the way in which EWS systems are used clinically means that EWS assessments are generally treated as independent. For example, a given EWS escalation decision is generally binary. It might demand that a score of 3 results in no medical intervention, whereas a score of 4 might elicit such a response (irrespective of the fact that the previous EWS was 0 or 3). Therefore, the clinical assumption is that the extent of derangement of physiology at any given time determines the risk of adverse outcome. Our finding that the analysis of one randomly chosen observation set per episode in a sample generated similar AUROC values for NEWS and DTEWS to those for analysis of the whole dataset confirms that any non-independence appears to have little effect on our results. Finally, our study is a single centre comparison of DTEWS and NEWS in a specific patient group, and the findings require confirmation in different patient populations (e.g., surgical), diseases (e.g., chronic respiratory disease), clinical settings (general hospital wards) and institutions.

## **CONCLUSIONS**

The Decision Tree analysis technique provided an independent method for validating the composition, weightings and ranges of the National Early Warning Score, NEWS. DT analysis may be useful ifor the rapid development of candidate models for disease-specific EWSs.

### **Conflicts of interest**

VitalPAC is a collaborative development of The Learning Clinic Ltd (TLC) and Portsmouth Hospitals NHS Trust (PHT). PHT has a royalty agreement with TLC to pay for the use of PHT intellectual property within the VitalPAC product. Professor Prytherch and Drs Schmidt, Featherstone and Meredith are employed by PHT. Professor Smith was an employee of PHT until 31/03/2011. Dr Schmidt, and the wives of Professors Smith and Prytherch are shareholders in TLC. Professors Smith and Prytherch, and Dr Schmidt, are unpaid research advisors to TLC. Professors Smith and Prytherch have received reimbursement of travel expenses from TLC for attending symposia in the UK. Dr Briggs's research has previously received funding from TLC through a Knowledge Transfer Partnership. Professor Smith was a member of the Royal College of Physicians of London's National Early Warning Score Development and Implementation Group (NEWSDIG).

### **Acknowledgements**

The authors would like to acknowledge the co-operation of the nursing and medical staff in the study hospital.

### **Funding**

Tessy Badriyah's PhD study at the Centre for Healthcare Modelling and Informatics, University of Portsmouth, UK, of which this research forms part, was funded by the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia.

## References

1. 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.
2. 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.
3. 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 - in press.
4. 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.
5. <http://www.mathworks.co.uk/help/stats/classificationtree.fit.html> (accessed 25/10/2013)
6. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29-36.
7. <http://cran.r-project.org> (accessed 25/10/2013)

## **Legends for Figures:**

### **Figure 1**

Decision tree for pulse component of DTEWS. Filled boxes indicate terminal nodes.

### **Figure 2**

The distribution of DTEWS and NEWS values, and their relationships to the primary outcome of cardiac arrest, unanticipated intensive care unit admission or death, each within 24 hours of a given vital signs.

### **Figure 3**

EWS efficiency curves for DTEWS and NEWS. For each EWS value, this plots the percentage of the total number of observations at, or above, that EWS value against the percentage of the total number of observations for which the outcome – cardiac arrest, unanticipated intensive care unit admission or death, each within 24 hours of the observation set – was true at, or above, that EWS value.

**Table 1: The vital signs data in the study dataset.**

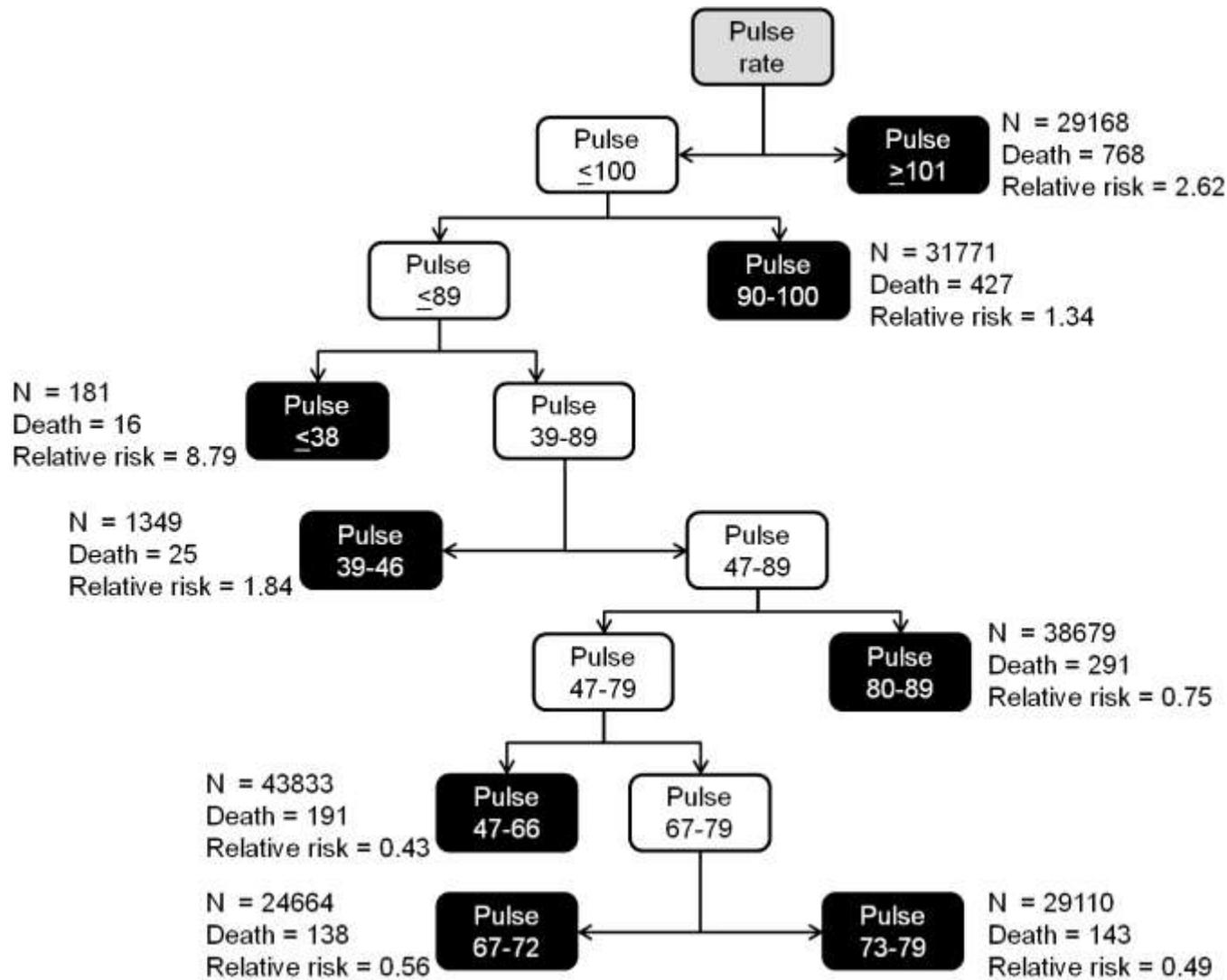
	<b>Mean</b>	<b>Standard deviation</b>
Pulse rate (beats.min <sup>-1</sup> )	81	19
Breathing rate (beats.min <sup>-1</sup> )	17	4
Systolic BP (mm Hg)	126	22
Temperature (°C)	36.7	0.4
S <sub>p</sub> O <sub>2</sub> (%)	96	3
		<b>Percentage</b>
% observations at each AVPU level		
	Alert	91.7
	Responds to Voice (V)	5.8
	Responds to Pain (P)	1.8
	Unresponsive	0.7
% breathing air at the time of S <sub>p</sub> O <sub>2</sub> measurement		77.1

1 Table 2: Comparison between the weighting ranges for DTEWS and NEWS<sup>3</sup>

		3	2	1	0	1	2	3
Respiration rate (breaths per minute)	DTEWS				≤18	19-20	21-24	≥25
	NEWS	<8		9-11	12-20		21-24	>25
S <sub>p</sub> O <sub>2</sub> (%)	DTEWS	≤89	90-92	93-94	95-99	100		
	NEWS	≤91	92-93	94-95	≥96			
Any supplemental oxygen	DTEWS				No			Yes
	NEWS				No		Yes	
Temperature (°C)	DTEWS	≤35.8	35.9-36.0	36.1-36.4	36.5-37.1	37.2-37.5	≥38.0	
	NEWS	≤35.0		35.1-36.0	36.1-38.0	38.1-39.0	>39.1	
Systolic blood pressure (mmHg)	DTEWS	≤89		90-116	117-272			≥273
	NEWS	≤90	91-100	101-110	111-219			≥220
Pulse rate (beats per minute)	DTEWS	≤38		39-46	47-89	90-100	≥101	
	NEWS	≤40		41-50	51-90	91-110	111-130	≥131
Level of consciousness	DTEWS				Alert (A)			Voice (V) Pain (P) Unresponsive (U)
	NEWS				Alert (A)			Voice (V) Pain (P) Unresponsive (U)

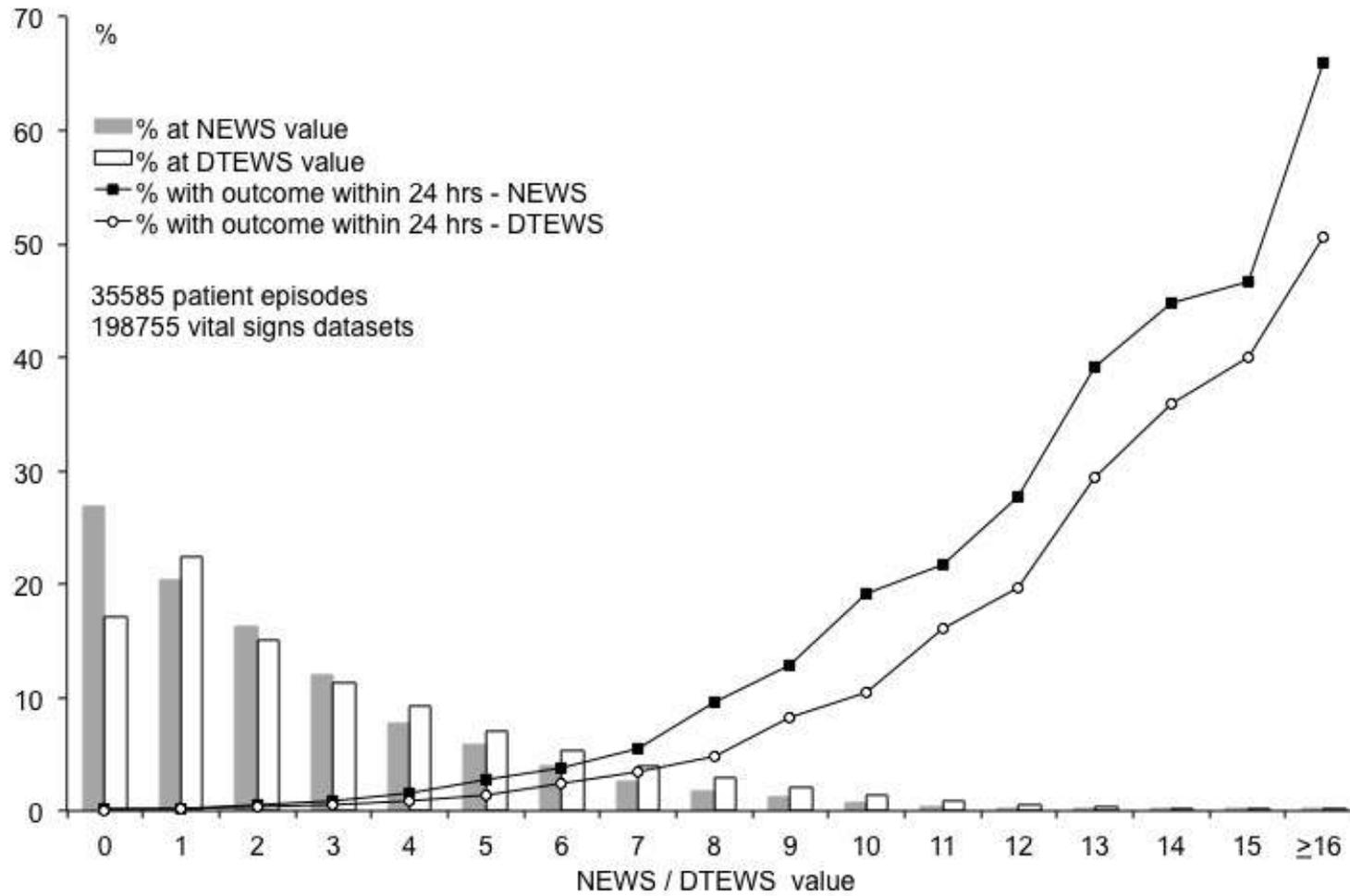
2  
3

4 Figure 1: Decision tree for pulse component of DTEWS. Filled boxes indicate terminal nodes.



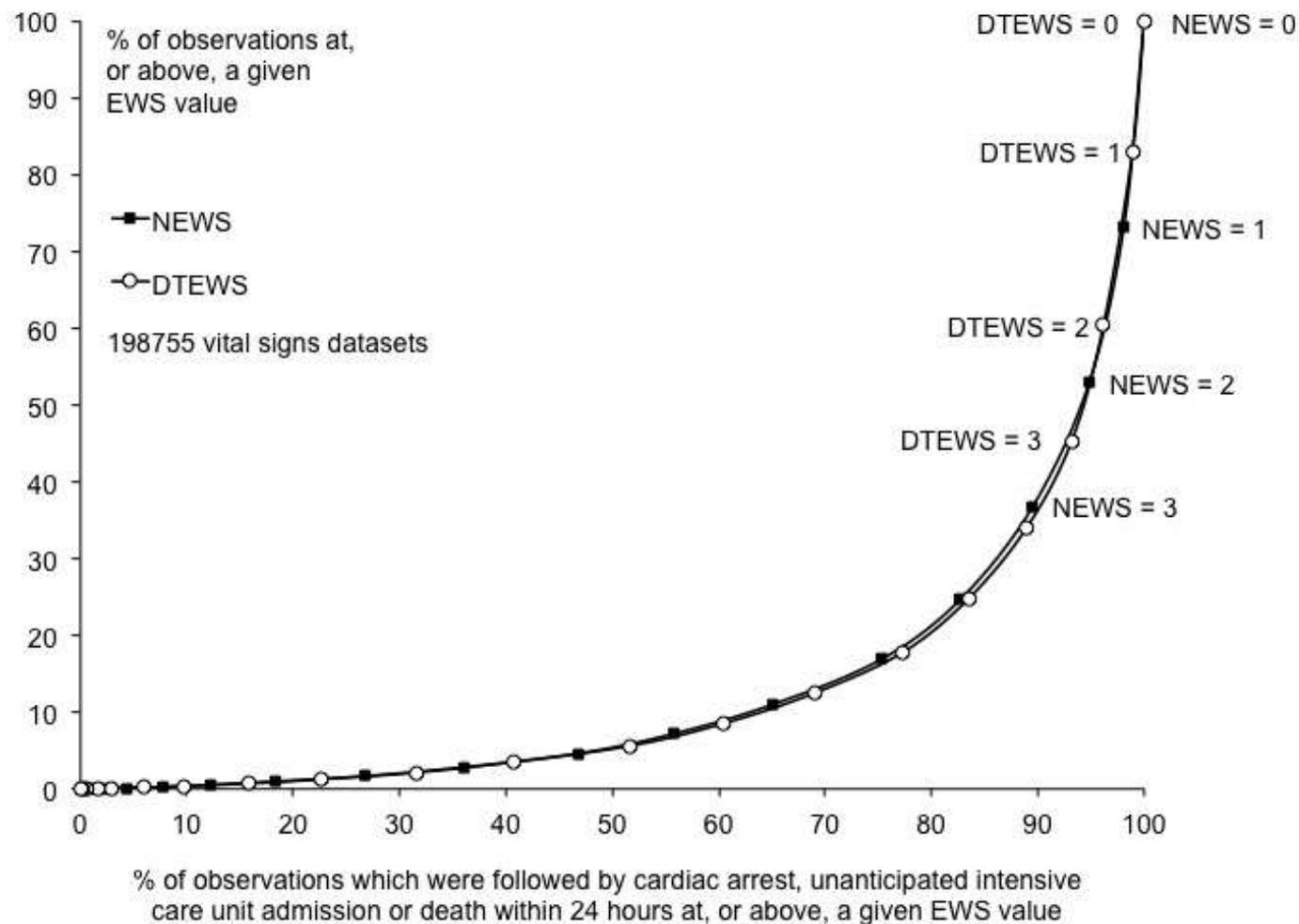
5

6 **Figure 2: The distribution of DTEWS and NEWS values, and their relationships to the combined outcome of cardiac arrest, unanticipated ICU admission**  
 7 **or death within 24 hours of a given vital signs.**



8

9 Figure 3: EWS efficiency curves for DTEWS and NEWS. For each EWS value, this plots the percentage of the total number of observations at, or above,  
 10 that EWS value against the percentage of the total number of observations for which the outcome – cardiac arrest, unanticipated ICU admission or  
 11 death, each within 24 hours of the observation set – was true at, or above, that EWS value.



12