Resuscitation 93 (2015) 46–52

Contents lists available at ScienceDirect

Resuscitation

journal homepage: www.elsevier.com/locate/resuscitation

EUROPEAN RESUSCITATION COUNCIL

Rapid response systems

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?☆

Stuart Jarvis[a,b], Caroline Kovacs[a], Jim Briggs[a], Paul Meredith[c], Paul E. Schmidt[c], Peter I. Featherstone[c], David R. Prytherch[a,c], Gary B. Smith[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] *Faculty of Health and Social Sciences, University of Bournemouth, Bournemouth, UK*

A R T I C L E I N F O

Article history:
Received 29 November 2014
Received in revised form 20 April 2015
Accepted 22 May 2015

Keywords:
Monitoring
Physiologic
Vital signs
Failure to rescue
Hospital rapid response team

A B S T R A C T

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 1564,153 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.

1. 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

☆ A Spanish translated version of the summary of this article appears as Appendix in the final online version at http://dx.doi.org/10.1016/j.resuscitation.2015.05.025.

* Corresponding author.
E-mail address: gbsresearch@virginmedia.com (G.B. Smith).

http://dx.doi.org/10.1016/j.resuscitation.2015.05.025
0300-9572/

occur frequently [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.

2. Method

2.1. 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.

2.2. 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.

2.3. 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 $\leq$ 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.

2.4. 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).

2.5. 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.

2.6. 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) [15]. This represents the probability that a randomly selected observation that was followed, within 24 h, by an adverse outcome had a higher score under an EWS than a randomly selected observation that was not followed, within 24 h, by an adverse outcome. 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 [16], 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. [17]. 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. [18]. 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 h 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.

2.7. Data analysis tools

All data manipulation was performed using Microsoft® Visual FoxPro 9.0. All analyses were undertaken in R version 3.02 [19].