Resuscitation 90 (2015) 1–6

Contents lists available at ScienceDirect

Resuscitation

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

Rapid Response Systems

Are observation selection methods important when comparing early warning score performance?☆

Stuart W. Jarvis[a,b], Caroline Kovacs[a], Jim Briggs[a], Paul Meredith[c], Paul E. Schmidt[d], Peter I. Featherstone[d], David R. Prytherch[a,c], Gary B. Smith[e,*]

[a] Centre for Healthcare Modelling and Informatics, University of Portsmouth, Portsmouth, UK
[b] Department of Health Sciences, University of York, York, UK
[c] TEAMS Centre, Portsmouth Hospitals NHS Trust, Portsmouth, UK
[d] Acute Medicine, Portsmouth Hospitals NHS Trust, Portsmouth, UK
[e] School of Health & Social Care, University of Bournemouth, Bournemouth, UK

ARTICLE INFO

Article history:
Received 16 August 2014
Received in revised form 19 January 2015
Accepted 25 January 2015

Keywords:
Early warning score
Illness severity score
Vital signs
Monitoring
Risk

ABSTRACT

Introduction: Sicker patients generally have more vital sign assessments, particularly immediately before an adverse outcome, and especially if the vital sign monitoring schedule is driven by an early warning score (EWS) value. This lack of independence could influence the measured discriminatory performance of an EWS.

Methods: We used a population of 1564,143 consecutive vital signs observation sets collected as a routine part of patients' care. We compared 35 published EWSs for their discrimination of the risk of death within 24 h of an observation set using (1) all observations in our dataset, (2) one observation per patient care episode, chosen at random and (3) one observation per patient care episode, chosen as the closest to a randomly selected point in time in each episode. We compared the area under the ROC curve (AUROC) as a measure of discrimination for each of the 35 EWSs under each observation selection method and looked for changes in their rank order.

Results: There were no significant changes in rank order of the EWSs based on AUROC between the different observation selection methods, except for one EWS that included age among its components. Whichever method of observation selection was used, the National Early Warning Score (NEWS) showed the highest discrimination of risk of death within 24 h. AUROCs were higher when only one observation set was used per episode of care (significantly higher for many EWSs, including NEWS).

Conclusions: Vital sign measurements can be treated as if they are independent – multiple observations can be used from each episode of care – when comparing the performance and ranking of EWSs, provided no EWS includes age.

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Several prior publications by our group have assessed the performance of the early warning scores (EWS) used to identify patients' severity of illness.[1–3] EWS systems allocate points in a weighted manner, based on the derangement of a predetermined set of patient vital signs variables (e.g., blood pressure, heart rate, breathing rate, temperature) from an arbitrarily agreed "normal" range. The points for each variable are summed and the total is used to inform a change in the patient's vital sign monitoring schedule and/or trigger a call for expert help at the bedside.

Our performance evaluations of EWS systems have often used all the observations sets from a sample of patient episodes and, therefore, contain multiple vital sign observation sets from the same patient episode in the analysis.[2,3] Multiple observations may be within 24 h of death (or another adverse outcome). We have considered an EWS to be better than another if it has a significantly ($p < 0.05$) higher area under the ROC curve (AUROC,[4] a measure of discrimination). Sicker patients generally have more vital sign assessments, particularly immediately before an adverse outcome, and especially if the vital sign monitoring schedule is driven by an EWS value. A previous review of our manuscripts have suggested

☆ 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.2015.01.033.

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

http://dx.doi.org/10.1016/j.resuscitation.2015.01.033
0300-9572/© 2015 Elsevier Ireland Ltd. All rights reserved.

that this lack of independence of the data points in the sample data sets may influence the measured discriminatory performance of an EWS. By extension, it is possible that an EWS that appears significantly better than another when all observations are used may appear significantly worse if only one observation was used from each episode.

EWS systems are implemented clinically as if vital sign measurements and derived EWS values are independent. EWS escalation decisions are generally binary. For example, an EWS value of 4 might result in no clinical intervention, whereas a value of 5 might require both a change in vital signs frequency and an assessment by a doctor (irrespective of the fact that the previous EWS was 0 or 4). Consequently, it is the extent of derangement of physiology at any given time, and not the degree of abnormality of any previous measurements, that determines actions taken based on the EWS score.

One study by our group[5] has suggested that treating vital signs and derived EWS values as independent may be reasonable, as an alternative technique of using one randomly chosen observation set per episode did not significantly affect discrimination of the combined outcome of cardiac arrest, unanticipated ICU admission or death within 24 h. In this study,[5] as with others,[1–3] the ability of the EWSs to discriminate the risk of a range of adverse outcome has been compared using the AUROC.[4] The use of multiple observation sets per episode has the potential to bias the AUROC as episodes with more observations may disproportionately influence the AUROC compared to those with fewer observations.

The aim of this study was to determine whether a lack of independence between data points when sampling patient observations might significantly change the ranking of EWS systems by their AUROC (i.e., lead to one EWS having significantly higher AUROC than another under one method of choosing observations, but significantly lower AUROC than the other under another method). We compared the performance of EWSs using three methods of observation selection: (1) all observations, (2) one randomly chosen observation set per episode, and (3) one observation set per episode based on choosing a random point in time within each episode.

2. Method

This research falls within local research ethics committee approval (08/02/1394) from the Isle of Wight, Portsmouth and South East Hampshire Research Ethics Committee.

2.1. Vital signs test results database and its development

We constructed a database of vital signs collected from all adult (≥16 years old) patients admitted to Portsmouth Hospitals NHS Trust on or after 25/05/2011 and discharged on or before 31/12/2012. We excluded data from patients discharged alive on the day of admission. We also excluded data from episodes in which no observations were recorded during the final 24 h (this was to reduce the numbers of patients in the dataset that might be on a recognised end of life pathway, where routine observations are halted). Vital signs data were recorded in real-time at the bedside using handheld electronic equipment running the VitalPAC software.[6] Each vital signs measurement set 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).

2.2. Within-episode dependence of vital signs observations

We tested for the level of within-episode dependence of the eight recorded vital signs and age at admission by calculating the intraclass correlation (ICC) for each using the episode identifier as the grouping variable.

2.3. Outcome

The outcome studied was death within hospital within 24 h of a vital signs dataset. Where it occurred, patient death was identified from the patient administration system (PAS). Therefore, multiple observations within a single episode of care might be followed by death within 24 h.

2.4. Observation selection methods

We used three different methods to select observations. First, we used all the observations in our dataset (i.e., for each patient care episode, all observations taken during that episode were used). Second, we made 10,000 samples of observation sets, each sample being constructed by selecting one observation set at random from every patient care episode (i.e., so each observation set in an episode had an equal chance of being selected in each sample). Third, we made 10,000 samples containing one observation set per episode. The observations were chosen by first randomly selecting a time during every episode in every sample, and then selecting the observation set closest to it.

In the case of the time-based selection method, we encountered problems due to inaccurate recording of patient admission and discharge times (some episodes had observations recorded either before admission or after discharge). Observations are time-stamped automatically as they are entered at the patient's bedside, whereas admissions and discharges are recorded independently on PAS and are less likely to be accurate.[2] Therefore we used the first and last observation dates and times, as the basis of determining the time period from which to choose observations. To avoid biasing against selection of the first and last observation sets, we added to the beginning and end of the selection period a length of time equal to half of the mean time between observation sets for that patient episode.

2.5. Selection of early warning scores and assessment of performance

We compared 35 published EWSs—33 previously compared by Smith et al.,[1,2] plus the cardiac arrest risk triage (CART) model[7] and the centiles EWS.[8] The full list of EWSs evaluated is listed in Table S1 of the Supplementary information. EWS performance – the ability to discriminate risk of death within 24 h of an observation set – was assessed by calculating the area under the ROC curve for each of the 35 EWSs under each of the three observation selection methods. When using all observations, we calculated a 95% confidence interval for the AUROCs using the methods set out by DeLong et al.[9] 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.

2.6. Effects of age

For those EWSs that include age, our analyses used age at admission in the EWS calculation. This is therefore constant throughout an episode and clearly not independent between observations. To study whether the inclusion of age in an EWS changed its ranking under different observation selection methods, we repeated

Download English Version:

https://daneshyari.com/en/article/3007927

Download Persian Version:

https://daneshyari.com/article/3007927

Daneshyari.com