Resuscitation 122 (2018) 99–105

Contents lists available at ScienceDirect

Resuscitation

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

Rapid response systems

Early Deterioration Indicator: Data-driven approach to detecting deterioration in general ward☆

Erina Ghosh*, Larry Eshelman, Lin Yang, Eric Carlson, Bill Lord

Philips Research North America, United States

ARTICLE INFO

Article history:
Received 22 June 2017
Received in revised form 13 October 2017
Accepted 29 October 2017

Keywords:
Early warning systems
Deterioration
Patient monitoring
Logistic regression

ABSTRACT

Introduction: Early detection of deterioration could facilitate more timely interventions which are instrumental in reducing transfer to higher levels of care such as Intensive Care Unit (ICU) and mortality [1,2].

Methods and results: We developed the Early Deterioration Indicator (EDI) which uses log likelihood risk of vital signs to calculate continuous risk scores. EDI was developed using data from 11,864 general ward admissions. To validate EDI, we calculated EDI scores on an additional 2418 general ward stays and compared it to the Modified Early Warning Score (MEWS) and National Early Warning Score (NEWS). EDI was trained using the most significant variables in predicting deterioration by leveraging the knowledge from a large dataset through data mining. It was implemented electronically for continuous automatic computation. The discriminative performance of EDI, MEWS, and NEWS was calculated before deterioration using the area under the receiver operating characteristic curve (AUROC). Additionally, the performance of the 3 scores for 24 h prior to deterioration were computed. EDI was a better discriminator of deterioration than MEWS or NEWS; AUROC values for the validation dataset were: EDI – 0.7655, NEWS – 0.6569, MEWS – 0.6487. EDI also identified more patients likely to deteriorate for the same specificity as NEWS or MEWS. EDI had the best performance among the 3 scores for the last 24 h of the patient stay.

Conclusion: EDI detects more deteriorations for the same specificity as the other two scores. Our results show that EDI performs better at predicting deterioration than commonly used NEWS and MEWS.

Introduction

Deterioration of patients in hospitals is typically preceded by changes in vital signs. More than 80% of these patients can be identified 24 h before the adverse event [1,2]. However, signs of deterioration often go unnoticed due to subtle changes in vital signs. Early warning score (EWS) systems are currently used to assess risk of deterioration. These systems use commonly measured vital signs to evaluate risk. New scores are calculated whenever a new set of vital signs measurement is obtained and alert clinicians when a preset threshold is reached.

Individual vital signs such as blood pressure or heart rate by themselves provide an incomplete picture of the patient condition. To solve this issue, several EWS systems have been developed [3–10]. These systems have a scoring table which assigns a risk score to each vital sign measurement. The total deterioration risk is obtained by adding all the individual risk scores. These tables have been derived by physician consensus or through iterative testing and can differ by hospital and by region. The risk score obtained from EWS systems is typically linked to an action table which specifies actions to be performed at each risk score.

Two widely used EWS systems are MEWS and NEWS. MEWS was proposed by Subbe et al. in 2001 [3] and uses 5 parameters. NEWS has been adopted for standardized assessment of acute illness severity by the NHS [4] and uses 7 parameters. The scoring table for NEWS has been standardized and unified for easier adoption. These systems have been developed for quick, manual scoring by clinicians. Hence, vital sign values are divided into large groups (bins) which are assigned the same risk. In addition, sharp cutoffs in dividing the bins can result in assigning similar patients to vastly different risks scores. For example, MEWS assigns a zero risk to systolic blood pressure (SBP) between 101 and 199 mm Hg and a risk score of 2 for values greater than 199 mm Hg (Fig. 1). A patient whose SBP is 199 mm Hg is assigned zero risk from blood pressure while a patient whose SBP is 200 mm Hg is assigned a risk score of

☆ A Spanish translated version of the summary of this article appears as Appendix in the final online version at https://doi.org/10.1016/j.resuscitation.2017.10.026.

* Corresponding author at: Acute Care Solutions, Philips Research North America, 2 Canal Park Suite 3, Cambridge, MA, United States.
E-mail address: erina.ghosh@philips.com (E. Ghosh).

https://doi.org/10.1016/j.resuscitation.2017.10.026
0300-9572/

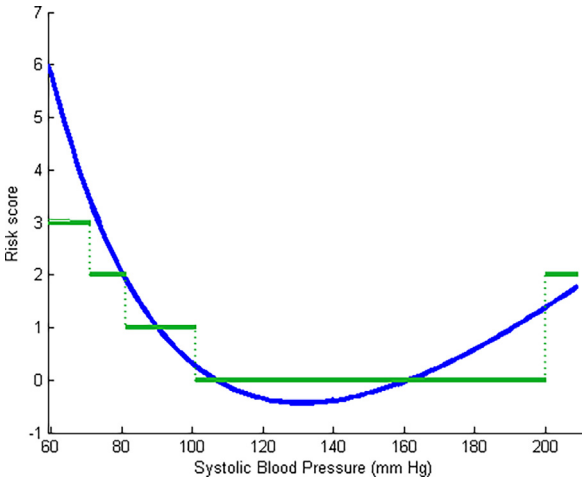

Fig. 1. EDI (blue) and MEWS (green) risk scoring curve for systolic blood pressure. Using a data- driven approach, EDI has a smooth curve while the MEWS scores are integer values. EDI also assigns negative risk scores for stable values. See text for details. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. These large risk score bins may reduce the sensitivity of EWS. In addition, EWS may contribute to alarm fatigue due to higher false alarm rates [11].

The predictive performance of EWS scores might be improved by making optimal use of the available data. EDI was developed to overcome the challenges mentioned above by using a data-driven approach to create more informative risk scoring tables. In this work, we describe how EDI was developed and calculated. We also validate EDI on a separate dataset and compare its performance to MEWS and NEWS. We quantify the performance of scores over time and map EDI scores to MEWS and NEWS.

Methods

Clinical data collection

Data collection was done in 2 phases – in the first phase data from 9265 patients (11,864 encounters) was collected and in the second phase, data from 2097 patients (2418 encounters) was collected. Thus, data from a total of 11,362 patients (14,282 encounters) was collected retrospectively from a community hospital with 78 general ward beds in Phoenix, AZ. The data was collected from hospital admissions over 2 years (2012–2013). The data included patients' spot-checked vital signs documented in the electronic medical records (Cerner Millennium, Kansas, US). The study was approved by the Institutional Review Board (IRB) of Banner Health (Phoenix, AZ, USA). More information on data measurement frequency is provided in the Data in Brief article [12].

Outcome definition and cohort labels

Deterioration was defined as death or transfer to higher level of care such as progressive care unit (PCU) or ICU [3,11,13]. We classified episodes into "stable" and "unstable" classes based on their discharge location.

Development of EDI

The algorithm was developed using data collected during the first phase and this included data from 11,864 encounters. The training dataset contained 1,731,648 stable measurements and 12,571 unstable measurements from the Development arm of the cohort (Fig. 2A).

Computation of feature risk curve

We generated curves for each feature indicating the risk of deterioration at each feature value. All the measurements from stable patients and all measurements for unstable patients taken up to 24 h prior to deterioration event were classified as "stable". Fig. 2B green band shows the data points classified as "stable" for risk curve generation. All the measurements taken in the period 90 min before a patient deteriorated were labeled as "unstable" class, since vital signs measured during this period reflected deterioration. Fig. 2B red band shows the unstable data points. The data was randomly divided into training and test sets, stratified over the stable and deteriorating classes and risk curves were created using Naïve Bayes classification [14]. One example is shown in Fig. 1 as the solid blue line. More information on risk curve generation is provided in the Appendix A.

EDI feature selection and score calculation

We next evaluated how well the features worked in combination and how much each contributed. A single set of measurements closest in time but before the middle of a stable encounter or 3 h prior to deterioration for unstable encounters were selected (Fig. 2B). If patient stay was shorter than 6 h, a set of measurements in the middle of the stay was chosen. A raw risk score was calculated by summing feature scores from the tables. Feature importance was evaluated by comparing the AUROC scores when the feature was included and left out. Since the EDI scores are not integers and can be either positive or negative, they look quite unlike MEWS and NEWS scores. However, we would expect EDI to be used like MEWS and NEWS are used, with recommended actions corresponding to specified thresholds. In order to give the EDI scores intuitive meaning, we mapped them into probabilities of instability (0–1.0) by applying logistic regression to the sum of feature scores.

EDI validation

The algorithm performance was validated using data collected from the second phase and it consisted of 2418 encounters (Validation arm of the cohort in Fig. 2A). Stable and unstable episodes were defined the same as described in 'Outcome definition and cohort labels' section above.

Calculation and comparison of EDI, MEWS and NEWS

The EDI score was calculated for each of these hospital encounters as described above (shown in Fig. 2B by blue line). The sensitivity and specificity at different thresholds in predicting deterioration was evaluated. MEWS and NEWS scores were calculated using the same data. The performance of the 3 scores were evaluated using the AUROC curve. The receiver–operator characteristic curve compares the sensitivity (detection rate) to specificity (1 – false positive rate) for each score and the area under the curve represents the discriminative power [15,16].

Mapping EDI to NEWS and MEWS

We mapped EDI to NEWS and MEWS to investigate the discriminative performance of these scores. This was done by calculating the sensitivity and specificity of NEWS at each value and then calculating the EDI score which had the same specificity as each of the NEWS values. The sensitivity of EDI and NEWS scores were compared for all NEWS values from 0 to 11. A similar method was used to map EDI to MEWS and compare EDI and MEWS sensitivity at all MEWS values.

Discriminative performance of the scores over time

In addition to evaluating the performance of EDI, MEWS, and NEWS at a specific time point, we also evaluated the performance of the 3 scores in the last 24 h of the patient's general ward stay