RESUSCITATION 188 (2023) 109821

Available online at ScienceDirect

Resuscitation

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

Clinical paper

The implementation of a real time early warning system using machine learning in an Australian hospital to improve patient outcomes

Levi Bassin[a,b,*], *Jacques Raubenheimer*[c], *David Bell*[a]

Abstract

Background: Early Warning Scores (EWS) monitor inpatient deterioration predominantly using vital signs. We evaluated inpatient outcomes after implementing an Artificial Intelligence (AI) based intervention in our local EWS.

Methods: A prior study calculated a Deterioration Index (DI) with logistic regression utilising demographics, vital signs, and laboratory results at multiple time points to predict any major adverse event (MAE—all cause mortality, ICU admission, or medical emergency team activation). The current study is a single hospital, pre-post study in Australia comparing the DI plus the existing EWS (Between the Flags-BTF) to only BTF. Data were collected on all eligible inpatients (≥16 years, admitted ≥24 hours, in general non-palliative wards). Controls were inpatients in the same hospital between January and December 2019. The DI was integrated into the electronic medical record and alerts were sent to senior ward nurse phones (July 2020–April 2021).

Results: We enrolled 28,639 patients (median age 73 years, IQR: 60–83) with 52.3% female. The intervention and control groups did not show any statistically significant differences apart from reduced admissions via the emergency department in the intervention group (40.4% vs 41.6%, $P = 0.03$). Risk for an MAE was lower in intervention than control (RR: 0.81; 95%CI: 0.74–0.89). Length of hospital stay was significantly reduced in the intervention group (3.74 days, IQR 1.84–7.26) compared to the control group (3.86 days, IQR 1.86–7.86, $P = 0.002$)

Conclusions: Implementing the DI in one hospital in Australia was associated with some improved patient outcomes. Future RCTs are needed for further validation.

Keywords: Medical informatics, Early Warning Score, Patient Deterioration, Patient Monitoring, Implementation

Introduction

Delayed recognition of hemodynamic and respiratory deterioration during hospitalization may increase morbidity and mortality rates.[1–3] There are measures that can improve the recognition of acute deterioration and reduce the likelihood of major adverse events such as all-cause mortality and unplanned ICU admission. These measures include the implementation of Early Warning Scores (EWS) such as the National and Modified Early Warning Scores (NEWS and MEWS) which generate aggregate scores based on single point in time observations[4–5] and have also been implemented in automated computerized systems.[6–8] The Between The Flags (BTF) alerting system which is implemented in all hospitals in the state of New South Wales, Australia is a two tiered alert system, and triggers alerts based on single point in time observation values.[9] In this system, any single vital sign outside of the normal reference range will trigger a nursing or medical review.

We previously described the development and validation of the Deterioration Index (DI) using a logistic regression model which was designed to predict the risk of Medical emergency team (MET) activation, unplanned ICU admission, and all-cause mortality.[10] We showed that our DI was highly predictive of patient outcomes after incorporating patient demographics, lab values, as well as the change in demographics and lab values. In a 48 hour simulation of 86,989 patient prior to a MAE, The DI had a larger area under the curve (AUC) for the receiver operator characteristic of 0.87 at predicting a MAE compared with the values for MEWS and NEWS of 0.71 and 0.74 respectively. The prior study included a

* Corresponding author at: Royal North Shore Hospital, Sydney, Australia.
E-mail address: levi.bassin@gmail.com (L. Bassin).

https://doi.org/10.1016/j.resuscitation.2023.109821
Received 22 August 2022; Received in Revised form 22 April 2023; Accepted 24 April 2023

six month silent trial, but the DI has to be evaluated in a real-world setting where its alerts are acted upon by the medical care team.

In this article we present our outcome measures after implementing our previously validated DI in the EWS of a single hospital. We examine whether implementing DI in conjunction with BTF will improve our patient outcomes compared to BTF alone. We examine whether the intervention produced significant improvements in the primary clinical outcome of a major adverse event (MAE): a composite of all cause mortality, MET calls, or unplanned ICU admission, in general hospital wards.

Methods

Study design

We obtained institutional ethics approval (2019/ETH00557), for a pre-post trial to examine the deployment of the DI in conjunction with BTF to BTF alone in a single Australian hospital. Eligible participants included all patients older than 16 years of age, admitted for more than 24 hours to all general wards during either the control or intervention periods. Palliative patients were excluded. Fig. 1 describes the inclusion of eligible participants. The control group data were extracted from hospital records for the period January to April 2019 and July to December 2019, when only BTF was in use. From January to April 2020 the DI was integrated and tested in a silent fashion where the system was running, but only sending alerts to one or two wards. Nursing and medical staff education also took place during this period. Following this, a staggered deployment took place over May and June 2020 where active alerts were 'turned on' for one or two wards per week until the entire hospital had the system in full production mode. Post-implementation and intervention data were collected from July 2020 to April 2021, during which time both BTF and DI were in use. To allow for a month-to-month comparison (to control for the potential confounder of seasonal variation in the cause for admission), control data were collected from the same hospital for the months January–April 2019, and July–December 2019, so that each period contained data for ten months (Supplementary figure 3).

EMR integration and alerting

A DI score from 1-10 was calculated using the previously described logistic regression model output[10] which produced a predicted probability of a MAE – P(*AE*) as a value between 0–1. This was converted to a score between 1–20 such that $DI = 10 \times P(AE)^{DC}$, where DC is the deterioration coefficient which can be adjusted.[10] This created a more interpretable score for clinical staff, and also allowed for real time calibration of the balance between sensitivity and specificity of the DI (Supplementary figure 1.1) by easily adjusting the DC. Increasing the DC would increase the DI's specificity but lower its sensitivity and vice versa. The DC was adjusted during the silent phase targeting two alerts per ward per day.

The model was installed into a rules engine, Rippledown (Beamtree Holdings, Australia) that was integrated with the EMR for a real time two-way data feed. The output of the model, a DI, was calculated in real time for every eligible patient within the hospital. If the DI was equal to 6 or 7, an 'amber' alert was triggered. If the DI was greater than or equal to 8 a 'red' alert was triggered. Each alert was accompanied by an automated clinical message which included the patient's heart rate (HR), blood pressure (BP) and respiratory rate (RR), as well as other variables including oxygen saturation, the use of supplemental oxygen, Haemoglobin (Hb), white cell count (WCC), urea (Ur) and creatinine (Cr) if these laboratory results were noted to be deteriorating (Supplementary figure 2.2). Changes in these variables over time were also incorporated into the clinical comment. All alerts were written into the electronic medical record (EMR) and sent via phone text message to the senior nurse for the given patient's ward. Amber alerts were reviewed by nursing staff with discretion to escalate and red alerts required immediate attention of a house medical officer. There were no additional interventions provided. Specifically, no additional staff were employed to review or react to DI alerts. In addition to this, all DIs along with their corresponding clinical summary, were displayed on the patient dashboard of the EMR (Supplementary figure 2.1) with the complete workflow demonstrated in Appendix 2.

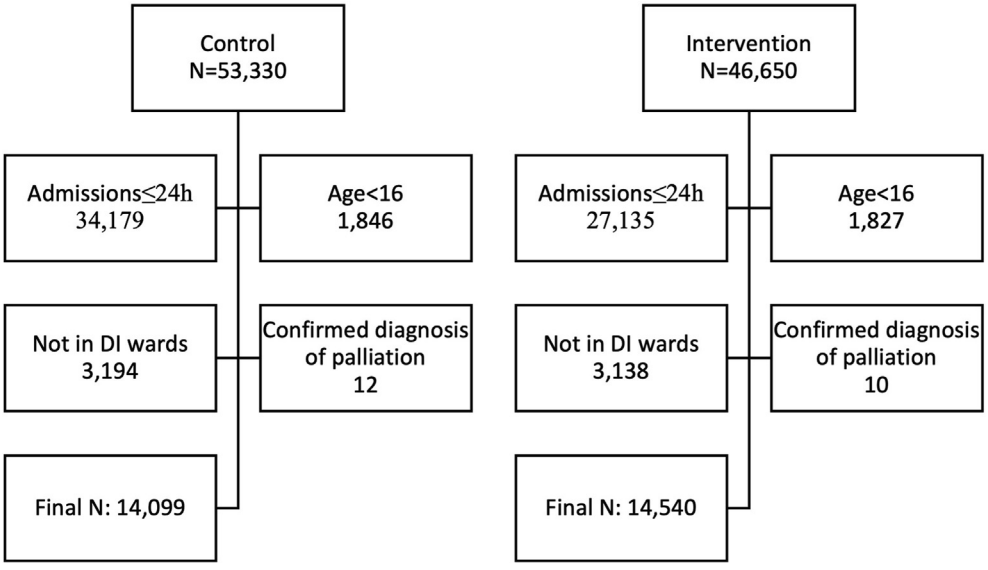

Fig. 1 – Eligible patients for the control and intervention groups.

Analysis

The primary outcome measure was a major adverse event (MAE) defined by 1) all-cause mortality, or 2) unplanned ICU admission or 3) MET activation. Secondary outcomes included 1) length of hospital stay (recorded in the EMR in hours, summarised into days), 2) length of ICU stay for unplanned ICU admissions (recorded in days), as well as 3) significant hypotension defined by a systolic blood pressure (SBP) $\leq$ 80 mmHg at any point in the hospital admission, 4) renal dysfunction defined by a fall in the estimated glomerular filtration rate (eGFR) $\geq$ 15 ml/min/1.73 m^2 between admission and discharge, and 5) elevated lactate defined by a lactate $\geq$3 mmol/L at any point during the admission, with the latter three recorded in binary fashion. The values defining the binary outcomes of hypotension, fall in eGFR, and high lactate were determined as being markers of clinically significant deterioration. To analyse the ability of the DI and BTF to correctly predict an adverse event, true positives were defined as any alert followed by a MAE within 48 hours and false positives were defined as any alert not followed by a MAE within 48 hours. False negatives were defined as adverse events which occurred without a prior alert. All patients who had no alerts and no adverse events were counted as true negatives.

Baseline characteristics for the study cohort were evaluated using chi-squared tests (using Yates' continuity correction for 2x2 tables) for categorical variables and the Mann-Whitney test for continuous variables. We provided analysis for the primary composite outcome and each included outcome. The Poisson regression models used Pearson scaling to adjust for some underdispersion, and the log of the length of total hospital stay was used as an offset. The chi-square test (again with Yates' correction) was used for secondary outcomes as these outcomes could have occurred at any time during the patient's stay but were recorded only once. Together with the chi-square values, we calculated the relative risk with 95% continuity-adjusted confidence limits. Analyses were done using SAS/STAT version 15.1 of the SAS software system for Windows version 9.4 and the online JavaStat—2-way Contingency Table Analysis tool.[11]

Results

Of the 53,330 and 47,053 patients admitted during the control (pre-DI) and intervention (DI) periods, respectively, 14,099 and 14,540 patients were eligible for inclusion in the control and intervention groups. A total of 352 patients had multiple adverse events (202 in control and 150 in intervention). The baseline characteristics of the two samples are shown in Table 1 and the study flow chart is shown in Fig. 1. Admissions via the emergency department (ED) were significantly less in the intervention vs the control group [40.4% (5,871/14,540) vs 41.6% (5,871/14,099), P = 0.03]. The impact of COVID-19 on the study population was minimal, with very few cases in Australia during the study period. This is reflected by the similar number of patient admission for each month between the control and study groups, including those undergoing surgery. Additionally there were very few COVID-19 admissions in this private hospital.

The estimates for each of the Poisson models testing one of the outcome variables are shown in Table 2, with the intervention group showing reduced probability for every outcome barring death. At the event level there were significantly less MAEs [5.7 % (852/14,738)] vs [7.1% (1,024/14,352), P < 0.001] in the intervention group

Table 1 – Baseline characteristics for control and intervention groups.

	Control (N = 14,099)		Intervention (N = 14,540)		P
	N	%	N	%	
Month					
January	1,231	8.7	1,237	8.5	0.430
February	1,386	9.8	1,460	10.0	
March	1,534	10.9	1,625	11.2	
April	1,338	9.5	1,434	9.9	
July	1,476	10.5	1,448	10.0	
August	1,422	10.2	1,499	10.3	
September	1,409	10.0	1,528	10.5	
October	1,515	10.8	1,520	10.5	
November	1,496	10.6	1,462	10.1	
December	1,292	9.2	1,327	9.1	
Sex					
Male	6,782	48.1	6,874	47.3	0.162
Female	7,317	51.9	7,666	52.7	
Admission via Emergency Department					
No	8,228	58.4	8,669	59.6	0.030
Yes	5,871	41.6	5,871	40.4	
General Anaesthesia					
No	7,375	52.3	7,505	51.6	0.241
Yes	6,724	47.7	7,035	48.4	
	N	**Median (IQR)**	**N**	**Median (IQR)**	**P**
Age (years)	14,099	74 (60–83)	14,540	73 (59–83)	0.356
Baseline haemoglobin	7,249	126 (112–138)	6,924	127 (112–139)	0.031
Baseline White cell count	7,248	8.3 (6.5–10.8)	6,923	8.3 (6.5–10.7)	0.877
Baseline eGFR	7,332	69 (52–82)	6,997	71 (53–83)	<0.0001

Table 2 – Poisson Regression Models Comparing Main Outcomes by Intervention/Control.

Model	Outcome	Intercept	Estimate	StdErr	IRR (95%CL)	P
1	Major Adverse event	−4.47	−0.167	0.0487	0.846 (0.769–0.931)	<0.001
2	Unplanned ICU	−6.52	−0.212	0.1049	0.809 (0.659–0.994)	0.044
3	MET call	−4.92	−0.165	0.0580	0.847 (0.756–0.950)	0.004
4	Death	−6.39	−0.110	0.1144	0.896 (0.716–1.121)	0.337

(Table 3). Mann-Whitney comparisons (Table 4) show that the intervention group had a significantly shorter length of total hospitalisation (3.74 days, IQR 1.84–7.26) compared to the control group (3.86 days, IQR 1.86–7.86, $P = 0.002$).

The true and false positive rates were determined for the DI and the BTF system in the intervention arm for the 48 hours prior to an AE. The sensitivity for the MAE at various cumulative time points is shown in Fig. 2. The sensitivity was greater for the DI at every time point apart from the prediction of MET activation at 4 hours prior. There was a striking drop in sensitivity for BTF as the duration prior to the AEs increased. In contrast, the DI maintained excellent sensitivity, even up to 24 hours prior to an AE. This is seen particularly with alerts prior to a mortality, so that even while the sensitivity of the BTF system came close to that of the DI at 4 hours (0.69 vs 0.77), the sensitivity of the BTF system dropped considerably with increasing times prior to an AE. The number of alerts not followed by an AE within 48 hours was more than 10 times higher for BTF than for the DI (66,101 vs 5,273). In terms of workload generated, the DI yielded 17.4 alerts not followed by an AE per day across 13 wards (1.34 per ward per day), whilst BTF yielded 218.2 alerts not followed by an AE per day across 13 wards (16.8 per ward per day).

We noted earlier that the intervention group had significantly fewer ED admissions. A subgroup analysis of the intervention by ED admission status showed that both ED and non-ED admissions had a significant reduction in the MAE following intervention ($P = 0.031$ and <0.001, Supplementary Table 4.1). Although baseline eGFR measurements showed a statistically significant difference between control and intervention groups, the difference is not clinically meaningful.

Discussion

We have shown that the addition of a machine learning derived DI alerts in conjunction with the established BTF system, was associated with a reduced composite major adverse event of all-cause mortality, unplanned ICU admission, or MET calls. Importantly the introduction of the DI system was also associated with improved patient hemodynamics with significant reduction in renal dysfunction, hypotension, and hyperlactatemia.

This DI is the first machine learning derived EWS that is paired with automated clinical summaries (Supplementary Figure 2.2) and integrated into the EMR. The ward-based patient dashboard with the DI viewable for all patients enables clinicians and managers to assess which patients are at the highest risk of deterioration. We believe that this is an improvement on current EMR dashboards,

Table 3 – Incidence and Relative Risk for the Main Outcomes by Intervention/Control.

Outcome	N (Rate/1000)		Relative Risk (95%CI)	P[a]
Event level (N = 29,090)	**Intervention N = 14,738**	**Control N = 14,352**		
Major Adverse event	852 (57.81)	1,024 (71.35)	0.81 (0.742–0.885)	<0.001
Unplanned ICU	105 (7.12)	132 (9.20)	0.775 (0.595–1.008)	0.057
MET call	544 (36.91)	653 (45.50)	0.811 (0.725–0.908)	<0.001
Patient level (N = 28,639)	**N = 14,540**	**N = 14,099**		
Death	133 (9.15)	151 (10.71)	0.854 (0.673–1.084)	0.202
Lactate >3 mmol/L	229 (15.75)	280 (19.86)	0.793 (0.664–0.946)	0.010
SBP <80 mmHg	426 (29.30)	528 (37.45)	0.782 (0.689–0.889)	<0.001
eGFR fall >15 ml/min/1.73 m^2	153 (10.52)	209 (14.82)	0.71 (0.574–0.878)	0.001

a: Continuity-adjusted chi-square.

Table 4 – Mann-Whitney Comparisons for Duration of Stay and ICU duration by Intervention/Control.

Variable	Group	N	Mean (SD)	Median (IQR)	P
Length of stay (days)	Control	14,099	6.36 (8.37)	3.86 (1.86–7.86)	0.002
	Intervention	14,540	6.06 (7.47)	3.74 (1.84–7.26)	
Days in ICU	Control	132	4.19 (5.48)	2.00 (2.00–4.50)	0.356
	Intervention	65	3.86 (3.11)	3.00 (2.00–5.00)	

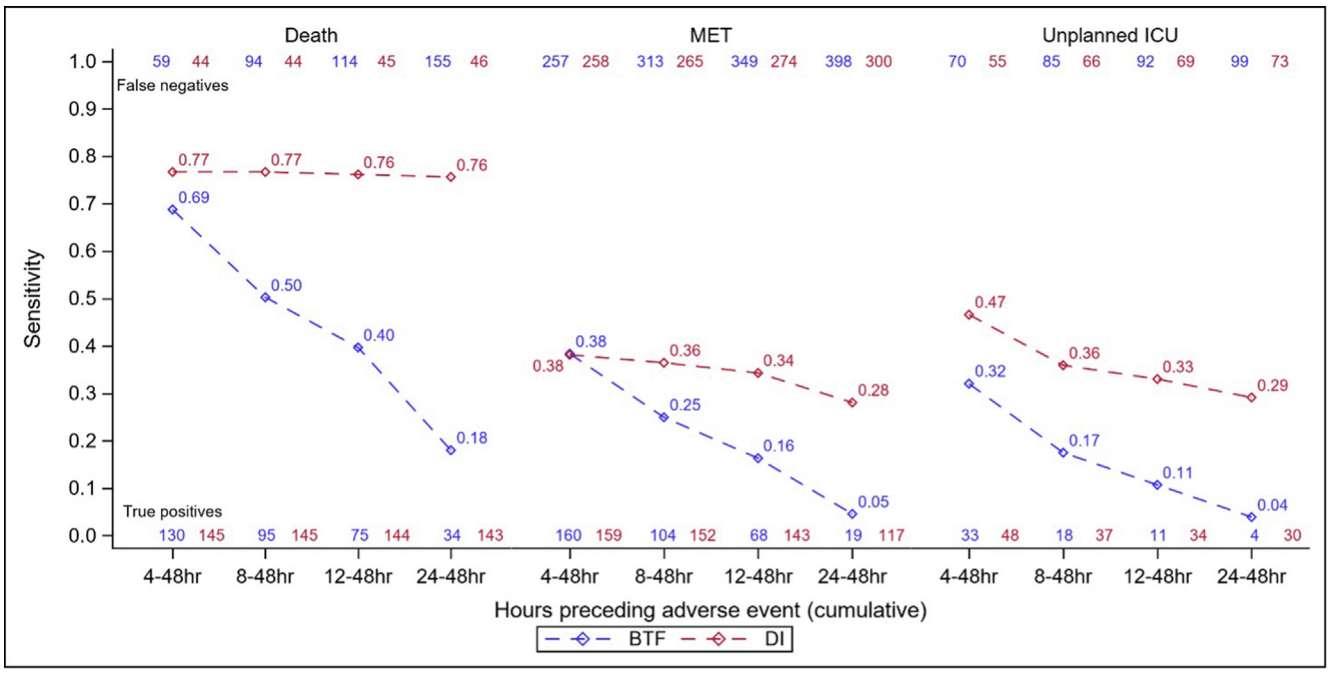

Fig. 2 – Sensitivity of the Between The Flags (BTF) system and the Deterioration Index (DI) by adverse event, at times prior to the adverse events.

and facilitates improved situational awareness as well as enhanced oversite. The clinical message was added to the alert system to improve communication between nursing and medical staff – often a significant factor in the delayed detection of inpatient deterioration.[12]

Escobar et al's recently published study demonstrated significant reduction in inpatient mortality with the use of a statistically derived deterioration alert system.[13] That system, however, required additional nursing staff who remotely monitored and manually reviewed records of patients identified at high risk. It may not be practical or feasible for every hospital to employ dedicated staff to monitor a patient dashboard.

The results of the BTF rapid response system that is used across the state of NSW (including the study hospital) showed a continued declining trend of cardiac arrests and particular benefit in low risk patients.[14] In the present study, a direct comparison of DI to BTF (where both are running side by side), the DI had a higher sensitivity at predicting an AE than BTF at anything greater than 8 hours in advance, and with far less false alarms (by a factor of 10). The sensitivity of the DI at 24–48 hours prior to an AE gives plausibility to the improvement in outcomes with the addition of the DI system, giving clinicians time to intervene.

In a study similar to ours which evaluated the impact of automated alerts using NEWS in general wards, patient outcomes were assessed before (n = 42,402) and after implementation (n = 42,920).[15] That study found no significant change in the primary outcome of mortality or unplanned ICU admission with the automated alert system. Implementation of our DI may present a more viable alternative.

Limitations

The DI was developed using data from two Australian hospitals[10] and in this study, implemented in one of the two, and therefore may not be as effective in other jurisdictions and patient populations. To further validate the success of this system it should be tested widely in different hospitals and on different patient populations.

Conclusion

We have shown that the use of a machine learning EWS in conjunction with the established BTF system can significantly reduce the composite major adverse event of unexpected mortality, unplanned ICU admissions, or MET activation over the use of BTF alone. Further validation of the system on diverse patient groups within a variety of clinical environments is required. We recommend further evaluation of the DI system, possibly using several hospitals in cluster-randomised or stepped-wedge trials, with clusters defined at the hospital level (defining wards as clusters may be confounded by transfers between wards).

Statement of Contribution

This is to certify that all three authors (DB, JR, LB) have made substantial contributions to the manuscript. DB and LB conceived and designed the study, performed the trial and extracted the data. All authors contributed to the analysis, the drafting the manuscript, and have given final approval for the submitted version. We all are all accountable for all the aspects of the work.

CRediT authorship contribution statement

Levi Bassin: Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review &

editing. **Jacques Raubenheimer:** Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing, Conceptualization. **David Bell:** .

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: David Bell and Levi Bassin are the co-developers of the Deterioration Index described in this manuscript as previously published.[10] The intellectual property of the Deterioration Index is the property of Beamtree Holdings Australia, a publicly listed company. David and Levi are employees of Beamtree.

Jacques Raubenheimer is an academic statistician at the University of Sydney, Faculty of Medicine and Health, School of Medical Science. He has performed the statistical analysis as an independent statistician, as a consultant for Beamtree.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.resuscitation.2023.109821.

Author details

[a]*Sydney Adventist Hospital, Sydney, Australia* [b]*Royal North Shore Hospital, Sydney, Australia* [c]*The University of Sydney, Faculty of Medicine and Health, School of Medical Sciences, Biomedical Informatics and Digital Health, Sydney, Australia*

REFERENCES

1. Harris S, Singer M, Rowan K, Sanderson C. Delay to admission to critical care and mortality among deteriorating ward patients in UK hospitals: a multicentre, prospective, observational cohort study. Lancet 2015;385:S40.
2. van Galen LS, Struik PW, Driesen BE, Merten H, Ludikhuize J, van der Spoel JI, et al. Delayed Recognition of Deterioration of Patients in General Wards Is Mostly Caused by Human Related Monitoring Failures: A Root Cause Analysis of Unplanned ICU Admissions. PLoS One 2016;11:e0161393.
3. Harris S, Singer M, Sanderson C, Grieve R, Harrison D, Rowan K. Impact on mortality of prompt admission to critical care for deteriorating ward patients: an instrumental variable analysis using critical care bed strain. Intensive Care Med 2018;44:606–15.
4. RCoPo L, editor National early warning score (NEWS): standardising the assessment of acute-illness severity in the NHS—report of a working party. Royal College of Physicians; 2012.
5. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. Qjm 2001;94:521–6.
6. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Critical Care 2017;21:1–9.
7. Heller AR, Mees ST, Lauterwald B, Reeps C, Koch T, Weitz J. Detection of deteriorating patients on surgical wards outside the ICU by an automated MEWS-based early warning system with paging functionality. Ann Surg 2020;271:100–5.
8. Mestrom E, De Bie A, Mvd S, Driessen M, Atallah L, Bezemer R, et al. Implementation of an automated early warning scoring system in a surgical ward: practical use and effects on patient outcomes. PloS one 2019;14:e0213402.
9. Commission CE. Recognition and management of patients who are clinically deteriorating. 2013.
10. Bell D, Baker J, Williams C, Bassin L. A trend-based early warning score can be implemented in a hospital electronic medical record to effectively predict inpatient deterioration. Critical care Med 2021;49:e961.
11. 2-way Contingency Table Analysis [cited 7th March, 2023]. Available from: (https://statpages.info/ctab2x2.html).
12. Agency NPS. Recognising and Responding Appropriately to Early Signs of Deterioration in Hospitalised Patients: November 2007. National Patient Safety Agency; 2007.
13. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. New England J Med 2020;383:1951–60.
14. Chen J, Ou L, Flabouris A, Hillman K, Bellomo R, Parr M. Impact of a standardized rapid response system on outcomes in a large healthcare jurisdiction. Resuscitation 2016;107:47–56.
15. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal impact of implemented early warning score and best practice alert for patient deterioration. Critical Care Med 2019;47:49.