|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Supplement e1: NEWS and MEWS scores simulation to calculate positive predictive value**  A simulation was run for the 48 hours prior to adverse events (AE). The National Early Warning Score (NEWS) and Modified Early Warning Score (MEWS) were calculated at each time point to obtain the sensitivity, specificity, and resulting positive predictive value (PPV). Tables eTable1 and eTable2 describe these results for each NEWS and MEWS score respectively. To maintain conformity between NEWS, MEWS, and the logistic model, we chose a specificity as close to 0.98 as possible. This corresponds to a NEWS of 7 and a MEWS of 4.  **eTable 1. National Early Warning Score (NEWS) results 48 hours prior to an Adverse Event**   |  |  |  |  | | --- | --- | --- | --- | | **NEWS** | **Sensitivity** | **Specificity** | **PPV** | | 1 | 0.987 | 0.220 | 0.009 | | 2 | 0.922 | 0.354 | 0.016 | | 3 | 0.819 | 0.543 | 0.030 | | 4 | 0.651 | 0.764 | 0.063 | | 5 | 0.488 | 0.911 | 0.138 | | 6 | 0.352 | 0.968 | 0.254 | | **7** | **0.219** | **0.987** | **0.349** | | 8 | 0.123 | 0.994 | 0.405 | | 9 | 0.065 | 0.997 | 0.424 |   *PPV - positive predictive value*  **eTable 2. Modified Early Warning Score (MEWS) results 48 hours prior to an Adverse Event**   |  |  |  |  | | --- | --- | --- | --- | | **MEWS** | **Sensitivity** | **Specificity** | **PPV** | | 1 | 1.000 | 0.117 | 0.004 | | 2 | 0.898 | 0.407 | 0.019 | | 3 | 0.639 | 0.857 | 0.109 | | **4** | **0.388** | **0.973** | **0.309** | | 5 | 0.182 | 0.992 | 0.436 | | 6 | 0.074 | 0.998 | 0.491 | | 7 | 0.027 | 0.999 | 0.554 |   *PPV - positive predictive value* |  |  |

**Supplement e2: Transforming the raw data variables**

In order to perform a logistic regression, each continuous variable was modified using an arithmetic transformation such that the modified variable has a linear relationship with the log-odds of an adverse event (AE). The log-odds for each continuous variable was calculated and displayed for manual verification. An example is shown in Figure e1 displaying the change in respiratory rate (RR) from baseline (RRTr BL). In order to linearise RRTr BL, the absolute value of RRTr BL is taken before setting all values of RRTr BL greater than 15 to 15. This produces a roughly linear relationship between the modified variable and the log-odds of an adverse event.

**eFigure1. Raw change in respiratory rate (breaths per minutes) from baseline vs the log-odds of an adverse outcome.**

![Chart, scatter chart

Description automatically generated]()

**eFigure2. Arithmetic transformation of the change in respiratory rate (breaths per minute - absolute value) vs the log-odds of an adverse outcome.**

![Chart, scatter chart

Description automatically generated]()

**Supplement e3: Choosing the ideal logistic model**

As described in the Results section of the manuscript, three logistic models were created using different data sets - Model 1 used all data up to and including data recorded at T0, Model 2 used all data up to and including data recorded at T-1, and Model 3 used all data up to and including data recorded at T-4 . A separate model was also created with vital signs alone i.e. with no laboratory values or demographics. Each of these models was then compared at different time points prior to an adverse event. The results are described below in table e4. Model 2 has the best balance of predictive accuracy across the different time intervals and so was chosen for the remainder of the study. In spite of the simplicity of the ‘vital signs only’ model, it still outperformed both NEWS and MEWS at all time points.

|  |  |  |  |
| --- | --- | --- | --- |
| **eTable3. AUC for each model and MEWS and NEWS prior to an adverse event** | | | |
|  | **T0** | **T-1** | **T-4** |
| **Model 1 (T0)** | 0.92 (0.9-0.94) | 0.87 (0.85-0.89) | 0.85 (0.83-0.87) |
| **Model 2 (T-1)** | 0.92 (0.9-0.94) | 0.89 (0.87-0.91) | 0.87 (0.85-0.89) |
| **Model 3 (T-4)** | 0.9 (0.88-0.92) | 0.88 (0.86-0.9) | 0.88 (0.86-0.9) |
| **Model - vital signs only** | 0.86 (0.84-0.88) | 0.79 (0.77-0.81) | 0.78 (0.76-0.8) |
| **NEWS (7)** | 0.79 (0.77-0.82) | 0.72 (0.70-0.74) | 0.70 (0.68-0.72) |
| **MEWS (4)** | 0.79 (0.77-0.82) | 0.74 (0.72-0.76) | 0.71 (0.69-0.73) |

*MEWS – Modified Early Warning Score, NEWS – National Early Warning Score, AE – Adverse Event, T - the reference time of an adverse outcome/control in hours prior to the event*

|  |  |  |  |
| --- | --- | --- | --- |
| **eTable4. Sensitivity for each model and MEWS and NEWS prior to an adverse event** | | | |
|  | **T0** | **T-1** | **T-4** |
| **Model 1 (T0)** | 0.59 | 0.37 | 0.35 |
| **Model 2 (T-1)** | 0.58 | 0.41 | 0.37 |
| **Model 3 (T-4)** | 0.46 | 0.4 | 0.39 |
| **Model - vital signs only** | 0.36 | 0.31 | 0.21 |
| **NEWS (7)** | 0.14 | 0.09 | 0.05 |
| **MEWS (4)** | 0.27 | 0.17 | 0.1 |

*MEWS – Modified Early Warning Score, NEWS – National Early Warning Score, AE – Adverse Event, T - the reference time of an adverse outcome/control in hours prior to the event*

**eTable5. logistic regression results when run against subtypes of adverse events. The logistic regression was run with each adverse event isolated, in order to predict events at least one hour in advance.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Patients** | **Specificity** | **Sensitivity** | **AUC** |
| **Death** | 1313 | 0.99 | 0.64 | 0.94 |
| **MET call** | 3462 | 0.98 | 0.5 | 0.88 |
| **ICU admission** | 2911 | 0.98 | 0.54 | 0.89 |
| **Return to theatre** | 316 | 0.98 | 0.6 | 0.91 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Supplement e4, eTable6: The final logistic model variables and coefficients** | | | |  |  |
| **Demographics** | **coefficient** | **P value** | **Current laboratory values** | **coefficient** | **P value** |
| Age | 2.84E-02 | <0.001 | Hb | -3.77E-02 | <0.001 |
| Sex | -1.87E-01 | <0.001 | WCC | 1.24E-03 | <0.001 |
| General Anaesthetic | 9.99E-01 | <0.001 | Urea | 6.19E-02 | <0.001 |
|  |  |  | eGFR | 1.43E-02 | <0.001 |
|  |  |  |  |  |  |
| **Current vital signs** |  |  | **Trend in laboratory values** |  |  |
| SBP | -1.07E-01 | <0.001 | Hb (TRBL) | 7.61E-03 | <0.001 |
| DBP | -3.06E-04 | <0.001 | Hb (TR1) | -2.45E-02 | <0.001 |
| RR | 9.31E-02 | 0.038 | WCC (TRBL) | -2.56E-02 | <0.001 |
| SpO2 | -7.49E-02 | <0.001 | WCC (TR1) | 6.33E-02 | <0.001 |
| HR | 2.56E-04 | <0.001 | Urea (TRBL) | 2.65E-01 | <0.001 |
| Conscious state | 8.12E-01 | <0.001 | Urea (TR1) | 4.89E-02 | <0.001 |
| Temperature (0C) | 8.94E-01 | <0.001 | eGFR (TRBL) | -2.36E-02 | <0.001 |
| Total IV fluids (24hrs) | 5.94E-04 | <0.001 | eGFR (TR1) | 2.21E-02 | <0.001 |
|  |  |  |  |  |  |
| **Trends in vital signs** |  |  | **Previous vital signs** |  |  |
| SBP (TRBL) | 1.00E-03 | 0.028 | SBP | -3.44E-02 | <0.001 |
| SBP (TR1) | -1.08E-02 | 0.002 | SpO2 | -9.40E-02 | <0.001 |
| SBP (TR2) | -1.12E-02 | <0.001 | HR | 2.45E-02 | <0.001 |
| DBP (TR2) | -1.41E-03 | <0.001 | Conscious state | -7.62E-01 | <0.001 |
| RR (TRBL) | -5.04E-02 | <0.001 | RR | 1.52E-01 | <0.001 |
| RR (TR1) | 5.29E-02 | <0.001 |  |  |  |
| RR (TR2) | 9.12E-02 | <0.001 |  |  |  |
| SpO2 (TR1) | -5.80E-01 | <0.001 |  |  |  |
| SpO2 (TR2) | 3.20E-02 | <0.001 |  |  |  |
| SpO2 (TRBL) | 1.42E-01 | <0.001 |  |  |  |
| HR (TRBL) | 7.38E-03 | 0.002 |  |  |  |
| HR (TR1) | 3.23E-02 | <0.001 |  |  |  |
| HR (TR2) | -1.05E-04 | <0.001 |  |  |  |
| Temperature (TRBL) | 1.37E-01 | <0.001 |  |  |  |
| Temperature (TR1) | 3.46E-01 | <0.001 |  |  |  |

*SBP - systolic blood pressure, DBP - diastolic blood pressure, RR - respiratory rate, SpO2 - oxygen saturation normalised for supplement O2, HR - heart rate, Hb – hemoglobin, WCC – white cell count, eGFR – estimated glomerular filtration rate. Trend from baseline - TRBL, trend from previous measurement - TR1, trend from two measurements prior - TR2.*

**Supplement e5: Excluded MET calls and demographics.**

**cTable7. Outcomes for MET calls that were excluded where the MET calls were prompted by isolated hypertension, bradycardia, or hypopnoea alone.**

|  |  |  |
| --- | --- | --- |
|  | **Excluded MET** | **Included MET** |
| Number | 530 | 3462 |
| Mortality | 6 (1.1%) | 503 (14.5%) |
| Unplanned ICU | 44 (8.3%) | 787 (22.7%) |
| Unplanned Surgery | 15 (2.8%) | 162 (4.7%) |

|  |
| --- |
| **cTable8. Patient demographics** |

|  |  |  |
| --- | --- | --- |
|  | Hospital 1 | Hospital 2 |
| Controls | 75855 | 176992 |
| Cases | 3902 | 4100 |
| Deaths | 803 | 510 |
| MET/RRT | 1538 | 1924 |
| Unplanned ICU admission | 1400 | 1511 |
| Unplanned surgery | 161 | 155 |
| Mean Age (years) | 70.1 | 75.8 |
| % female | 47 | 49 |

*MET – Medical Emergency Team call, RRT – Rapid Response Team call, ICU – Intensive Care Unit*

**Supplement e6: Positive predictive values**

Table e9 shows the positive predictive values (PPV) calculated for the logistic model as well as NEWS, MEWS, and Yellow flag alerts (from the “Between the Flags” protocol). These were determined from the 48-hour simulation prior to an AE or control time stamp. The PPV was calculated for 3 separate simulations: the entire 48 hours prior to an AE T-(48-0), 48 to 1 hour prior T-(48-1), and 48 to 4 hours prior - T-(48-4). This final measure that records the PPV up to 4 hours prior reflects how well suited each test is at determining an unwell patient at least 4 hours in advance of an AE.

The logistic model has a PPV roughly twice that of the NEWS and MEWS, and an order of magnitude more than the yellow flag alerts. We focused on the PPV because the higher the PPV, the lower the risk of alert fatigue for clinical staff. At a PPV of 0.6, our logistic model correctly predicts an AE in 60% of cases and is more likely to gain acceptance than the Yellow flag at a PPV of 0.071, which will be correct in only 7% of cases.

**eTable9. Positive predictive value for alerts prior to an adverse outcome:**

**Logisitic model vs NEW vs MEWS vs Yellow flag**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Logistic model** | **NEWS** | **MEWS** | **Yellow flag** |
| **T-48-0** | 0.64 | 0.35 | 0.31 | 0.083 |
| **T-48-1** | 0.63 | 0.32 | 0.28 | 0.076 |
| **T-48-4** | 0.6 | 0.29 | 0.26 | 0.071 |

*T – time in hours prior to an adverse event*

**Supplement e7: Alert Timing.**

Figure e3 below shows the sensitivity for predicting an AE as a function of the logistic model output (probability of an AE) at various time points. The simulation was run for 12, 24, and 48 hours prior to an AE and each of these curves plotted together. This demonstrates that there aren’t a significant number of patients detected at 48 hours that have no further alerts.

**eFigure3. The sensitivity for the 48-hour simulation compared to the sensitivity of only the most recent 24 hours prior to an AE and the sensitivity of only the most recent 12 hours prior to an AE.**



**Supplement e8: Alert Timing.**

Figure e4 shows the cumulative sensitivity for the 48 hours prior to an AE for both amber and red alerts generated by the logistic model.

**eFigure4. The cumulative sensitivity for the 48-hour simulation for Amber and Red alerts**



**Supplement e9: Deterioration Index**

As described in the body of the manuscript, we developed a deterioration index (DI) such that each value reflects a linear relationship to the positive predictive value (PPV) of an AE. eTable8 describes the characteristics of each DI level in a patient-centric manner, with the sensitivity calculated based on how many patients had an alert prior to an AE. TP (patients) is the number of patients with at least one true positive alert and FN (patients) is the number of patients with an AE but no prior alert. The final two columns describe the total number of TP and FP alerts for each DI value. DI levels 6-7 trigger an amber alert’ (to flag that clinical review is required) and DI levels 8-10 trigger a ‘red alert’ (to flag that rapid response is required).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **eTable10. Characteristics of the Deterioration Index (DI) levels** | | | | | | | | | |
| **DI** | **Alert** | ***P*(*AE*)** | **PPV** | **Sens** | **Spec** | **TP (patients)** | **FN (patients)** | **TP alerts** | **FP alerts** |
| **2** | none | 0.01 | 0.07 | 0.98 | 0.23 | 7854 | 142 | 59763 | 826879 |
| **3** | " | 0.04 | 0.18 | 0.87 | 0.58 | 6977 | 1019 | 39397 | 181145 |
| **4** | " | 0.08 | 0.27 | 0.78 | 0.75 | 6275 | 1721 | 30887 | 82823 |
| **5** | " | 0.15 | 0.39 | 0.69 | 0.87 | 5536 | 2460 | 23451 | 37188 |
| **6** | amber | 0.26 | 0.51 | 0.59 | 0.94 | 4707 | 3289 | 17111 | 16552 |
| **7** | " | 0.38 | 0.59 | 0.51 | 0.97 | 4042 | 3954 | 12905 | 8960 |
| **8** | red | 0.55 | 0.67 | 0.4 | 0.98 | 3231 | 4765 | 8746 | 4322 |
| **9** | " | 0.75 | 0.72 | 0.27 | 0.99 | 2184 | 5812 | 4879 | 1939 |
| **10** | " | 0.88 | 0.74 | 0.17 | 1 | 1363 | 6633 | 2633 | 918 |

*P(AE) – probability of an adverse event, the PPV - positive predictive value, sens – sensitivity, spec – specificity, TP – true positive, FN – false negative, FP – false positive*

|  |  |  |
| --- | --- | --- |
| **eTable11. Percentage of adverse events predicted by alert type** | | |
|  |  |  |
|  | **Adverse events detected (%)** | |
| **Hours prior** | **DI ≥ 6 (amber alert)** | **DI ≥ 8 (red alert)** |
| **<1** | 42.7 | 32.9 |
| **1-4** | 33.9 | 25.2 |
| **4-8** | 26.4 | 18.4 |
| **8-12** | 21.9 | 14.7 |
| **12-18** | 18.3 | 12.0 |
| **18-24** | 14.8 | 9.4 |
| **24-36** | 12.0 | 7.6 |
| **36-48** | 8.2 | 5.0 |
|  |  |  |

*DI – Deterioration Index*

**Supplement e10: 48, 24, and 12 hour Simulations**

Simulations were run at 48 hours out, 24 hours out and 12 hours out to show that the effect of removing earlier alerts (which may have been due to imputed results) on true positives and sensitivities was negligible. The simulation run at 24 hours out ignores all alerts before 24 hours. The simulation run 12 hours out ignores all alerts before 12 hours.

In this manner we have shown that there a very few alerts occurring over 12 hours prior that are not followed with alerts within 12 hours. This reduces the chance of seeing an alert at 48 hours that was a chance occurrence and therefore labelled as a true positive incorrectly. The other aspect that this addresses in the bias that imputed values may have. If the imputed values for missing data 48 hours ago resulted in a positive bias for the model, then the 12 hour simulation should have a greatly reduced sensitivity which we have no observed.

**eTable 12. Sensitivity at detecting an adverse event for Amber and Red Alerts with the simulation run for various lengths prior to an adverse outcome.**

|  |  |  |
| --- | --- | --- |
| Simulation | Amber alert (DI 6-7) | Red Alert (DI 8-10) |
| 48 hours | 0.61 | 0.44 |
| 24 hours | 0.60 | 0.43 |
| 12 hours | 0.58 | 0.41 |