Predicting Outcome in Acute Respiratory Admissions Using Patterns of National Early Warning Scores

Sarah Forster[1,2*], Tricia M McKeever[3], Matthew Churpek[4], Sherif Gonem[1], Dominick Shaw[1,2]

Author Affiliations
[1] Nottingham University Hospitals NHS Trust, Nottingham, UK
[2] University of Nottingham School of Medicine, Nottingham, UK
[3] University of Nottingham, Division of Epidemiology and Public Health, Nottingham, UK
[4] University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI, USA

[*]Correspondence to Dr Sarah Forster, sarahforster@nhs.net, Clinical sciences building, Edwards Lane, Nottingham, UK NG5 1PB

Keywords: Early Warning Scores, Critical Care, Clinical deterioration, Recognise and Rescue

Abstract-

Aim

Accurately predicting risk of patient deterioration is vital. Altered physiology in chronic disease affects the prognostic ability of vital signs based early warning score systems. We aimed to assess the potential of early warning score patterns to improve outcome prediction in patients with respiratory disease.

Methods

Patients admitted under respiratory medicine between April 2015 and March 2017 had their National Early Warning Score version 2 (NEWS2) calculated retrospectively from vital sign observations. Prediction models including temporal patterns were constructed and assessed for ability to predict death within 24 hours using all observations collected not meeting exclusion criteria. The best performing model was tested on a validation cohort of admissions from April 2017 to March 2019.

Results

The derivation cohort comprised 7487 admissions and the validation cohort 8379 admissions. Adding maximum score in the preceding 24 hours to most recently recorded NEWS2 improved area under the ROC curve for death in 24 hours from 0·888 (0.881-0.895) to 0·902 (0.895-0.909) in the overall respiratory population.

Conclusion

Combining most recently recorded score and maximum NEWS2 score from the preceding 24 hours demonstrated greater accuracy than using snapshot NEWS2. This simple inclusion of scoring pattern should be considered in future iterations of early warning scoring systems.

Introduction

The National Early Warning Score, now in its second iteration (NEWS2), is deployed in 76% of the 223 acute hospitals Trusts and all 10 ambulance Trusts across the National Health Service(NHS) in England [1], and in hospitals across Europe, the USA, Canada and Asia as a screening tool to categorise patients at risk of deterioration through highlighting deviation of regularly measured vital sign parameters from a predefined physiological range. NEWS2 and its predecessor (NEWS) have been retrospectively validated through several large outcomes-linked vital signs datasets and are more accurate at predicting clinical deterioration than prior early warning score algorithms [2-4].

Respiratory inpatients have a general mix of acute presentations in otherwise well patients and in the setting of chronic disease. Within this population Chronic Obstructive Pulmonary Disease (COPD) represents a paradigm for patients presenting with underlying chronic disease states where baseline vital sign values can be different to the population from which NEWS was derived, and where physiology can react differently to acute pathology [5, 6]. Altered physiology may elevate the baseline NEWS2 score, leading to unnecessary medical interventions in stable patients, alert fatigue in medical staff (reducing clinical response to a high scoring patient[7]), inappropriate oxygen use, or misplaced clinical reassurance in an unstable patient [8].

Concerns regarding the impact of chronic disease on sensitivity and burden of clinical reviews have led to exploration of personalised scores through artificial intelligence and big data analysis. However, there are a limited number of hospitals with the digital maturity to implement such systems, with some NHS trusts still employing paper charts. We therefore set out to determine whether simple temporal patterns in NEWS2 could be used to improve the discrimination of the currently used snapshot score. In order to future proof this approach for prospective iterations of NEWS we also applied this approach to a previously published NEWS-FiO2 to determine additional benefit of pattern in score if factors such as fraction of inspired oxygen were to be included in a graded manner when the score is reviewed in 2023 [9].

Methods

Source of data

Approval was given by the UK Health Research Authority (IRAS ID 270837) and Nottingham University Hospitals Trust's Caldicott guardian, Research and Innovation team and Information Governance department (Ref: DG20-000049-D and IG0025) to establish a database of anonymised, outcomes-linked vital sign data in adults aged 18 years or over admitted to Nottingham University Hospitals NHS

Trust under the care of Respiratory Medicine between 1st April 2015 and March 2019. As the study is limited to use of previously collected, non-identifiable information the HRA did not require research ethics committee review.

Vital signs were recorded at the bedside using the Nervecentre platform, with outcomes and diagnoses linked from the Medway clinical record prior to anonymisation and extraction. A set of vital signs comprised neurological status using Alert-Voice-Pain-Unresponsive (AVPU); new onset confusion (yes/no); respiratory rate measured in breaths per minute; oxygen saturations (%); heart rate in beats per minute; blood pressure in mmHg; temperature in Celsius; fraction of inspired oxygen (%) or flow rate (L/min); urine output in ml per hour (if the patient was catheterised) or passed urine in the preceding 6 hours (yes/no). Any observation set with missing or impossible values was removed from the analysis. Additional data included age, comorbidity score, hospital discharge status and ICD10 codes for admission, dominant and discharge diagnoses. The data set was split into an initial derivation cohort from April 2015 to March 2017, and a validation cohort from April 2017 to March 2019. Data definitions are explained in Table *1*.

Table 1- Definitions relating to NEWS2 used in the study

Term	Definition
Observation	Set of vital signs recorded at bedside. Taken together each set amalgamated to NEWS2 score
Vital Sign Score	Indicates how far each vital sign deviates from set normal range, calculated at collection of each observation with a weighting of 0-3.
NEWS2 score-	National Early Warning Score version 2- Published by the Royal College of Physicians (RCP) in 2017 and mandated for use across the NHS in the UK. NEWS2 is an aggregate early warning score. NEWS2 is a continuous variable from a minimum score of 0 to a maximum of 20.
Scale 1	Oxygen target scale for NEWS2- for use in patients with no evidence of type 2 respiratory failure. Target saturations 94-98%.
Scale 2	Oxygen target scale for NEWS2- for use in patients with evidence of type 2 respiratory failure- used in all patients with diagnosis of COPD in line with clinical practice. Target saturations 88-92%.
Cut points-	NEWS2 scores at which certain actions are advised as per the protocol published by the RCP: **NEWS2 of 5-6**: Minimum hourly observations, registered nurse to immediately inform medical team and request urgent assessment within 1 hour by clinician with core competencies in care of acutely ill patients. Provide clinical care in an environment with monitoring facilities.

	NEWS2 of less than 5 but3 in one category: Separate category in original scoring protocol but clinically treated the same as a score of 5 or more. **NEWS2 of 7 or more**: Continuous monitoring of vital signs; registered nurse to inform registrar or above in medical team, emergency assessment within 30 minutes by team with critical care competencies and advanced airway management skills; consider transfer to level 2 or 3 area with clinical care in an environment with monitoring facilities

Participants

All admissions aged 18 years or older completed within the study period admitted to and discharged from Respiratory Medicine were included. Any vital signs coded as 'End of Life Care' (i.e., interventions aimed at palliation rather than prolonging life) were removed from the analysis.

The NEWS2 score was calculated retrospectively for each set of vital signs observations, with all scores during an admission not coded as end of life care being included in the analysis in line with previous research in this area [10]. Cut points were applied in line with the escalation protocol published with NEWS2 in which a score of 5 or more dictates an urgent response and hourly monitoring, and 7 or more an emergency response with continuous monitoring [11]. NEWS2 oxygen saturation scale 1, with target saturations of 94-98%, was applied to all patients without a diagnosis of COPD. Scale 2 (which adjusts for patients at risk of hypercapnic respiratory failure) was applied to all patients with a diagnosis of COPD in line with previous research [12], identified by presence of an ICD10 code for COPD at any point during admission.

Statistical Analysis

Most recently recorded NEWS2 score was applied as an independent variable and as part of novel bivariate logistic regression models combining most recently recorded NEWS2 score with pattern of NEWS2 score, both over the preceding 24 hours and throughout admission, to assess ability to predict death within 24 hours of an observation. Death within 24 hours was used as the outcome rather than ICU admission as several factors influence ICU admission (bed availability, staffing etc), not just clinical status.

Scoring patterns generated included difference between most recently recorded and previous NEWS2 value, (delta NEWS2), maximum value, minimum value, standard deviation of scores, and mean of scores. The patterns were used to create restricted cubic spline models with three knots, as indicated by the data and to reduce the risk of overfit, at the placement recommended as by Harrell [13]. Univariate models were created using the uvrs package in STATA. Each variable was then combined with most recently recorded NEWS2 score using the mvrs package to create bivariate restricted cubic spline models. As an additional analysis to allow for a score which could be applied in less sophisticated systems a predictive additive model was created using maximum NEWS2 score in the preceding 24 hours and most recently recorded NEWS2 score. This additive approach combining maximum score in the preceding 24 hours and most recently recorded score was also applied to the NEWS-FiO2 proposed by Malycha et al, with FiO2 calculated from flow rate and cut offs applied as per their methods [9].

Ability to predict death was assessed using several approaches. Sensitivity and specificity at the clinical cut points of 5, 5 or a single vital sign score of 3, and 7 were calculated to reflect current clinical application of the score. NEWS2 was also treated as a continuous ordinal and evaluated using area under receiver operating characteristic curve (ROC curve) and area under precision recall curve (PR curve), a plot of precision (positive predictive value) against recall (sensitivity) as appropriate in the whole population, and then in separate cohorts defined by COPD diagnosis. Use of area under the PR curve was used in addition to area under the ROC curve as the latter can be affected disproportionately by small improvements in prognostic ability in the setting of a data set with skewed outcomes, with a very small percentage of observations associated with adverse outcomes, as seen in hospital populations. As with area under ROC curve, the higher the area under the PR curve, the better the model performance.

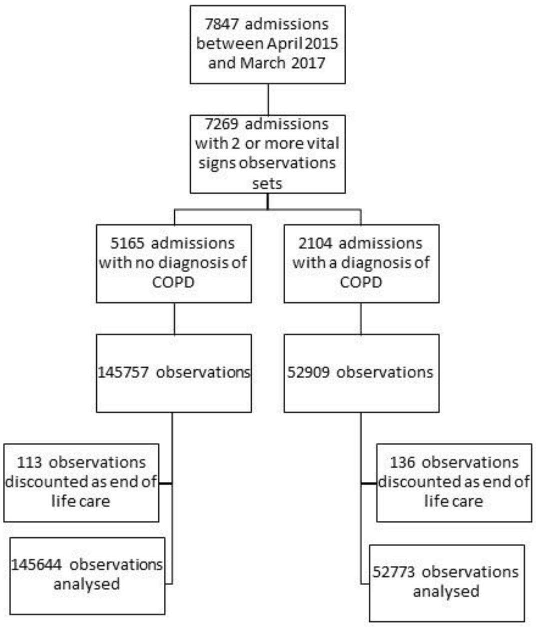
Initial analysis and model building was performed on the initial derivation cohort and analysis to verify findings was performed on the validation cohort. All observations recorded during a patient's stay were included in the analysis.

Results

Study Population

There were 7487 completed admissions from 5136 individual patients to the Nottingham University Hospitals Trust Respiratory Department during the initial two-year study derivation period from April 2015 to March 2017, and 8739 admissions from 5928 individual patients during the validation period from April 2017 to March 2019 (Figure 1). Admission demographics are detailed in

a)



b)

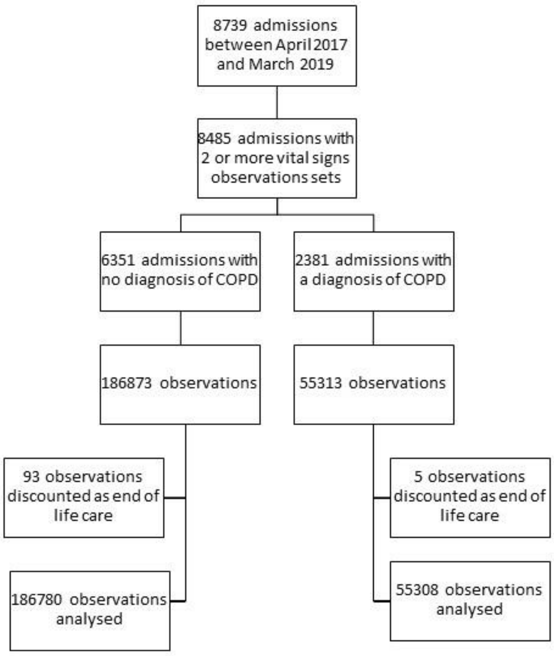
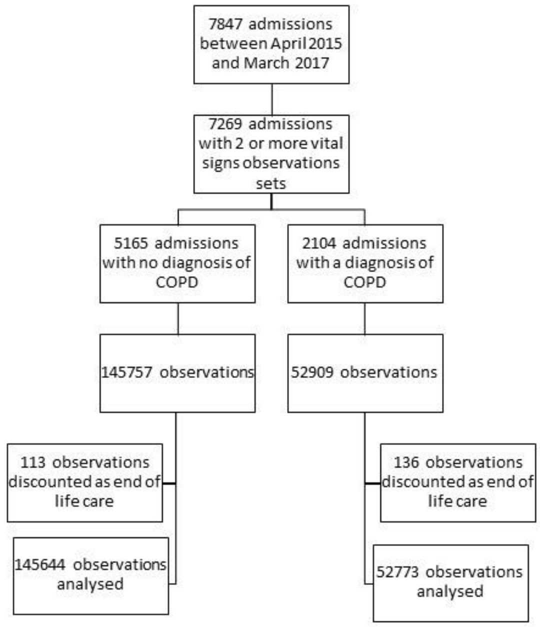


Table *2*.

Figure 1: a- Patients with respiratory disease completing admission between 1st April 2015 and 31st March 2017-derivation cohort; b- Patients with respiratory disease completing admission between 1st April 2017 and 31st March 2019- validation cohort

a)



b)

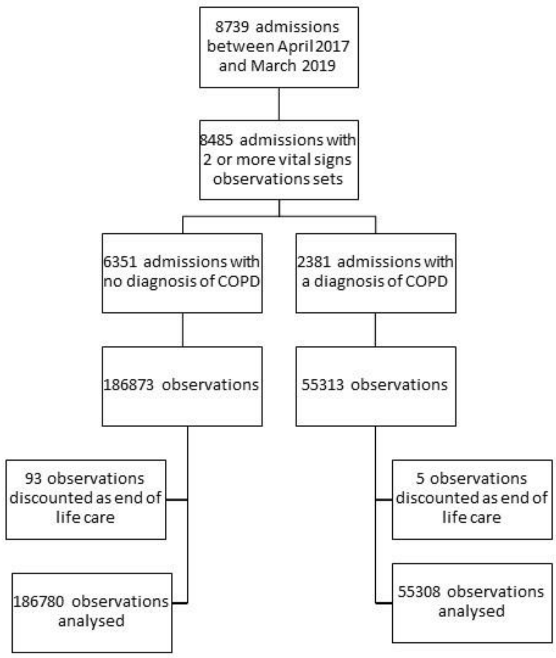


Table 2- Derivation and Validation cohort demographics

	April 2015- March 2017			April 2017- March 2019		
	Respiratory (total)	Non- COPD (scale 1)	COPD (scale 2)	Respiratory (total)	Non- COPD (scale 1)	COPD (scale 2)
Admissions (n)	7269	5165	2104	8485	6351	2381
Female (%)	3953 (54.4)	2775 (53.7)	1178 (56.0)	4718 (54.0)	3402 (53.5)	1316 (55.3)
Median age (IQR)	71 (61- 81)	71 (61- 81)	71 (61- 76)	71 (56-76)	66(51-76)	71 (61-76)
Median Length of Stay in days (IQR)	4 (2-8)	4 (2-8)	3(2-7)	3 (1-7)	3 (1-7)	3 (1-6)
In hospital mortality (%)	413 (5.7)	328 (6.4)	85 (4.0)	470 (5.5)	398 (6.5)	72 (3.1)

NEWS2 Performance in the overall respiratory population

In the overall respiratory population NEWS2 had a sensitivity of 0·87 and specificity of 0·72 at a cut point of 5 for predicting death within 24 hours of an observation set. Sensitivity increased to 0·89 where observations with a single vital sign scoring 3 were added to scores of 5 or more, at the expense of a reduction of specificity to 0·67. At a cut point of 7, sensitivity was reduced to 0·68 and specificity increased to 0·90.

Area under the ROC curve for NEWS2 in the overall respiratory population was 0·888 (95% CI 0·881- 0·895) in the derivation cohort of April 2015 to March 2017 and 0·880 (95% CI 0·873- 0·887) in the validation cohort. Area under the PR curve was 0·140 in the derivation cohort and 0·133 in the validation cohort. Each point increase in NEWS2 score increased the odds ratio for death within 24 hours of an observation by 1·72 (95% CI 1·69-1·74) in the derivation cohort and 1·70 (95% CI 1·68 - 1·72) in the validation cohort.

Workload

The additional clinical workload (i.e. patient review by nurse or doctor) that high NEWS scores led to can be seen in the number of observations reaching the threshold for review that were then not followed by death within 24 hours. For example, 32 observations met the criteria for escalation and clinical review for every observation followed by death within 24 hours of that score at a cut point of 5, meaning there were 31 scores requiring clinical review that were not followed by death within 24 hours. This increased to 38 if observations scoring 3 in a single vital sign were included. 16 observations per outcome identified met the criteria for escalation at a cut point of 7. These values were similar to those seen in the validation cohort (Table *3*).

NEWS2 performance in patients with a diagnosis of COPD, applying oxygen target saturation scale 2

Sensitivity at a cut point of 5 was reduced to 0·77 In the scale 2 cohort, with a higher specificity of 0·77 when compared to the scale 1 cohort. Adding observations with scores of 3 in one vital sign increased sensitivity to 0·81 with specificity reduced to 0·74. For a cut point of 7 sensitivity was 0·53 and specificity was 0·93.

39 observations met the criteria for clinical review/escalation at a cut point of 5 per outcome identified of death within 24 hours. 41 observations per outcome identified met the criteria for escalation if observations containing a single vital sign scoring 3 were included and 17 observations at a cut point of 7.

Area under the ROC curve analysis was 0·857 (95% CI 0·838-0·877) and area under the PR curve was 0·114 in the derivation cohort. Area under ROC curve was 0·878 and are under PR curve was 0·100 in the validation cohort. The odds ratio per point increase in NEWS2 score was 1·70 (95% CI 1·65-1·76) in the derivation cohort and 1·76 (95%CI 1·70-1·83) in the validation cohort.

Table 3- Sensitivity, Specificity values of NEWS2 at cut points of 5 and 7 in the derivation and validation cohorts

	Derivation cohort April 2015- March 2017		**Validation cohort April 2017- March 2019**	
Cut point 5	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**
NEWS2 in total respiratory population	0.87	0.72	0.88	0.69
NEWS2 in COPD (Scale 2)	0.77	0.77	0.81	0.79
Cut point 5 or single vital sign score of 3				
NEWS2 in total respiratory population	0.89	0.67	0.89	0.64
NEWS2 in COPD (Scale 2)	0.81	0.74	0.85	0.76
Cut point 7				
NEWS2 in total respiratory population	0.68	0.90	0.70	0.88
NEWS2 in COPD (Scale2)	0.53	0.93	0.57	0.94

Using NEWS2 pattern to enhance risk prediction

Maximum and mean NEWS2 in the preceding 24 hours demonstrated similar area under ROC curve analysis to stand-alone NEWS2 for outcome of death in 24 hours (Figure 2a).

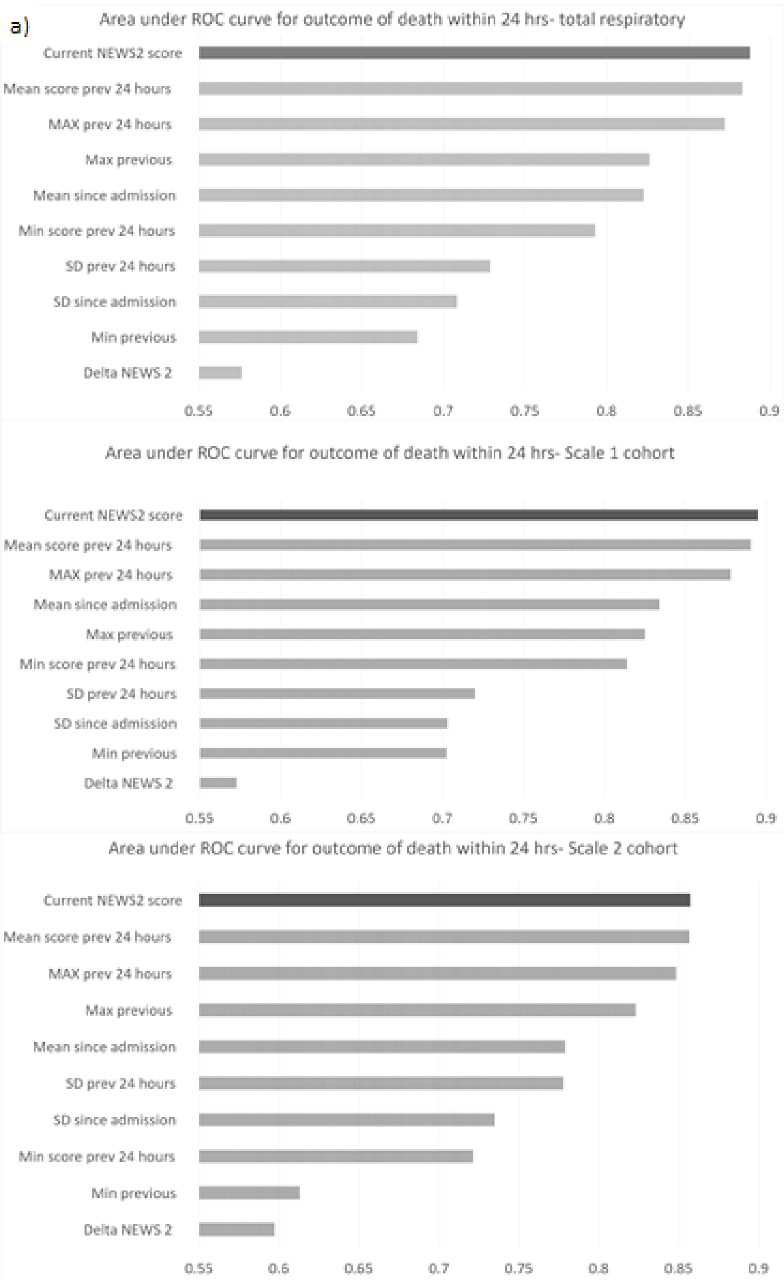
Improvement in prognostic ability was seen in all bivariate restricted cubic spline models compared to NEWS2 alone (Figure 2b). The model with highest prognostic ability for death within 24 hours combined maximum score in the preceding 24 hours with most recently recorded score, giving a ROC curve value of 0·903 (95%CI 0·896-0·910)in the total population and 0·880 (95%CI 0·862- 0·897) in the Scale 2 cohort.

A simple additive model using maximum score in the preceding 24 hours and most recently recorded score, had equal prognostic ability to the spline model using the same components, with ROC curves for outcome of 0·902 (95%CI 0·895-0·909) in the overall population (95%CI 0·862-0·898) in the Scale 2 cohort. This is also reflected in the area under precision recall curves shown in Table 3. As precision recall curves incorporate positive predictive value, improvement here indicates the potential to reduce escalated scores without sacrificing sensitivity.

Applying a cut point of 12 to the additive model to be used in place of an equivalent NEWS2 cut point of 5 would result in 7035 (9.2%) fewer scores meeting the threshold for escalation in the overall population and 1366 (11·2%) fewer scores reaching the threshold for escalation in the scale 2 cohort with a diagnosis of COPD, without reducing sensitivity in identifying outcome of death within 24 hours in either group in the validation cohort.

Figure 2: a- Comparison of area under ROC curves for univariate restricted cubic spline models of NEWS2 pattern and existing score for outcome of death within 24 hours; b- Comparison of area under

ROC curves for bivariate restricted cubic spline models combining *most recently recorded* NEWS2 and various pattern variables for outcome of death within 24 hours



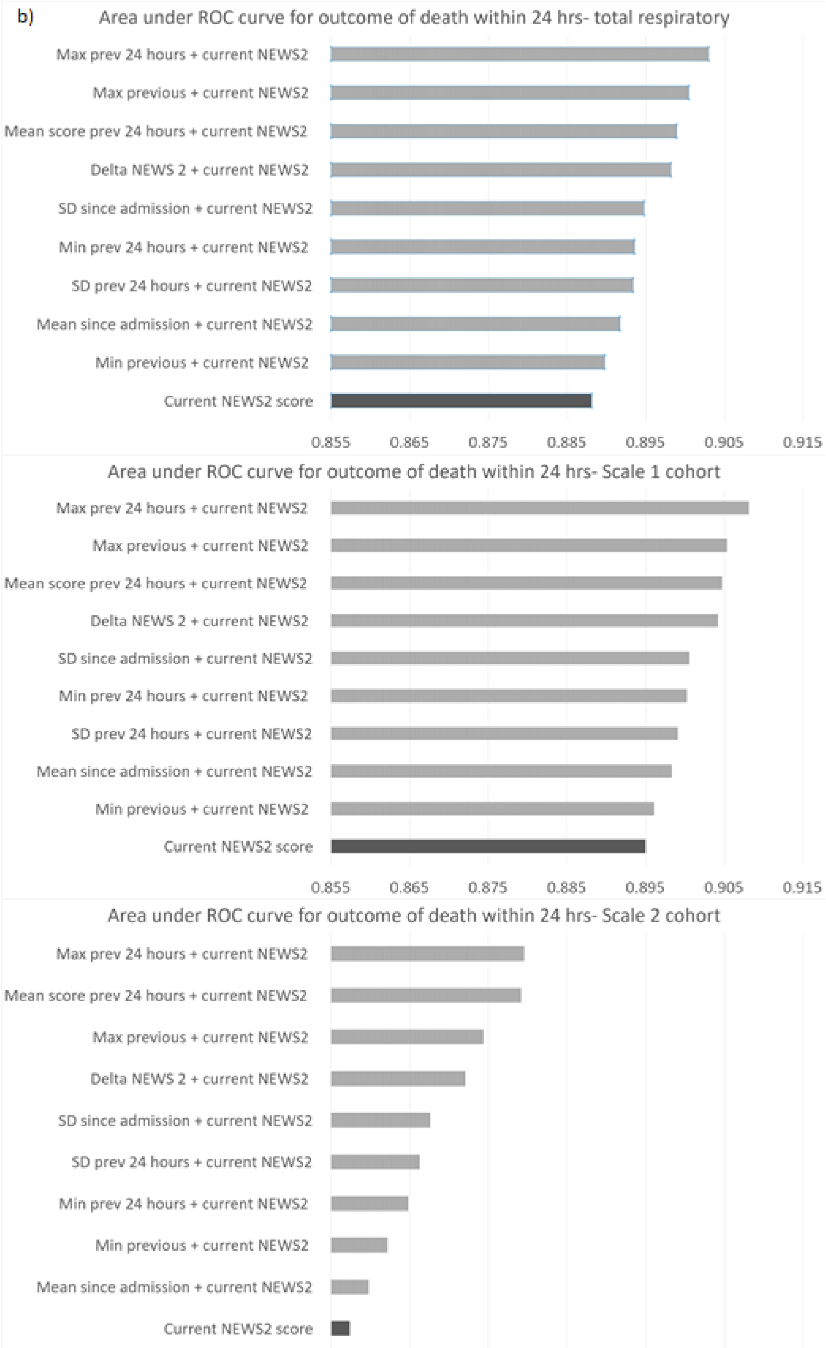


Table 4- Area under ROC and PRC for NEWS2 and additive score combining *most recently recorded* NEWS2 and maximum NEWS2 in the preceding 24 hours

Population	**Metric**	**Derivation cohort April 2015- March 2017**		**Validation cohort April 2017- March 2019**	
		Area under ROC curve	Area under PR curve	Area under ROC curve	Area under PR curve
Total Respiratory	NEWS2	0.888 (0.881-0.895)	0.140	0.880 (0.873-0.887)	0.133
	Additive NEWS2 + Maximum previous 24hrs	0.902 (0.895-0.909)	0.144	0.898 (0.891-0.904)	0.144
Scale 2 Cohort	NEWS2	0.857 (0.838-0.877)	0.114	0.878 (0.859-0.897)	0.099
	Additive NEWS2 + Maximum previous 24hrs	0.880 (0.862-0.898)	0.118	0.903 (0.885-0.921)	0.122

It has been suggested that the addition of a graded FiO2 score to future iterations of NEWS could improve risk prediction [9]. In this population, application of a previously described NEWS-FiO2 did not provide significant improvement in area under the ROC curve in predicting outcome of death within 24 hours. However, this may be attributed to the small number of outcomes present in the study population. Both the original NEWS2 and NEWS-FiO2 demonstrated improvement in discrimination when maximum score in the preceding 24 hours was applied to the total respiratory population and Scale 2 cohorts (Appendix table 1)

Discussion

In our study, NEWS2 had good prognostic ability for predicting death within 24 hours in the overall respiratory population, but a reduced prognostic ability in patients with a diagnosis of COPD. We also created a simple additive model combining most recently recorded NEWS2 with maximum score in the preceding 24 hours that could be used to reduce the number of observations reaching the threshold for escalation without affecting sensitivity for predicting which observations would be followed by death within 24 hours. A similar improvement in prognostic accuracy was indicated if the same approach was applied to a score incorporating FiO2

Following the release of the original NEWS in 2012 there has been ongoing evaluation of the score with the result that a second oxygen scale and additions to the AVPU criteria were made for NEWS2. While Scale 2 mitigated concerns regarding hyperoxia in patients at risk of type 2 respiratory failure it did not account for other baseline characteristics of these patients which impact on the ability of the score to predict which patients are at risk of deterioration. In addition, patients admitted to hospital with COPD have a lower mortality than the overall respiratory population (4·0% vs 5·7% in the derivation cohort and 3·1% vs 5·5% in the validation cohort). This makes the positive predictive value even more important due to skew between observations and outcomes and thereby the potential for excessive workload and unnecessary intervention

Echevarria et al. [12] analysed the performance of NEWS2 scale 2 when applied to patients with COPD. Scale 2 led to a reduction in scores reaching escalation thresholds, improved discrimination when compared to the original NEWS score (area under ROC curve 0·72 vs 0·65) and did not fail to identify any outcomes escalated by scale 1. Pimentel et al. [2] used a combination of coding and oxygen prescriptions to identify patient cohorts at risk of hypercapnic respiratory failure and confirmed hypercapnic respiratory failure. The performance of NEWS and the scale 2 component of NEWS2 (the modified AVPU component of NEWS2 was not applied) was compared in these cohorts to respiratory patients without risk factors for hypercapnia. As in our study, NEWS2 had worse predictive ability in the cohort with hypercapnic respiratory failure. These findings, and ours, suggest that the underlying physiological changes from chronic respiratory disease make NEWS2 less effective in patients at risk, or with hypercapnic respiratory failure, including those with COPD.

Using trends in vital signs observations has been shown to improve predictive ability [14, 15]. In this study, novel variables created from the pattern of NEWS2 scores preceding the most recently recorded set of observations were demonstrated to be independent predictors of outcome, and enhanced the prognostic ability of NEWS2 when combined with most recently recorded NEWS2 score in bivariate models.

This demonstrates the potential to further improve NEWS without having to change either the mode of data collection or the observations recorded, and providing additional value even where additional factors such as Fio2 are included. Furthermore, use of maximum score in the preceding 24 hours would be possible in a paper-based system, while additional modelling could potentially combine multiple variables to improve accuracy in an electronic system.

Our study is the first to examine the possible impact on workload of adding an additional layer of risk assessment. Applying a cut point of 12 to the additive model combining NEWS2 and maximum NEWS2 in the preceding 24 hours, corresponding in sensitivity to a NEWS2 score of 5, would result in 7035 (9·2%) fewer scores meeting the threshold for escalation in the overall population and 1366 (11·2%) fewer scores reaching the threshold for escalation in the scale 2 cohort with a diagnosis of COPD, without reducing sensitivity in predicting death within 24 hours (see appendix Table 2) .

The size and completeness of our data set (all observations were input directly onto devices at bedside with a very small percentage of missing or impossible entries) strengthens confidence in our findings. Other strengths include the fact that all elements of the NEWS2 score were incorporated in the vital signs observations sets collected and that ICD10 coding made it possible for patients to be assigned to the appropriate oxygen scale. The TRIPOD checklist [16] for reporting performance of predictive scores and the STROBE statement for reporting cohort studies were applied through design, analysis and reporting. Lastly while area under the ROC curve is the most commonly used measure applied in studies relating to predictive models such as NEWS2, it is now recognised that due to the small percentage of outcome (death) within a hospital population, area under precision recall curves give added information, so both are included [4, 17-19].

Limitations included the fact data were retrospective and from a single centre. It was not possible to retrospectively apply scale 2 to all patient groups who might be managed using scale 2 throughout the entire study period, therefore the decision to apply to patients with a diagnosis of COPD was a pragmatic approach to ensure consistency.

. The relatively small number of outcomes also represents a higher risk of type 2 error in examining the statistical discrimination of these models. While the use of multiple vital signs from an individual care episode could at first glance appear to be a limitation, this approach has been validated in the literature [10] and has become a recognised approach to evaluating early warning scores [3, 9, 20, 21].

In conclusion, chronic pathophysiological changes, such as those found in respiratory disease, affect the prognostic ability of NEWS2. This prognostic ability can be improved without the need for additional changes in data collection or major changes to existing systems by addition of the maximum score in the preceding 24 hours to the most recently recorded NEWS2 and could be applied to future iterations of NEWS if other variables such as graded FiO2 were to be included; this approach could easily be tested in other centres. This simple and scalable improvement could have beneficial implications all healthcare systems which strive to balance the seesaw of resource limitations versus the need to predict, react to, and prevent clinical deterioration in hospitalized patients.

Data Sharing Statement

Due to the permissions and information governance restrictions at an NHS Trust level we are unable to make the deidentified participant data available.

Data Dictionary and Study protocol will be made available with publication

Regulatory approval

This project was provided with HRA approval- IRAS project ID: 270837 Protocol number: 19074. Due to the nature of the data as collected during routine care and anonymisation it was not necessary to gain full REC approval.

Appendix

Table 1- Area under ROC and PRC for NEWS2, additive score combining NEWS2 and maximum score in the preceding 24 hours, NEWS-FiO2 and additive score combining current NEWS-FiO2 and maximum NEWS-FiO2 in the preceding 24 hours

Prediction of outcome of death within 24 hours of an observation		NEWS2	Additive NEWS2 + max NEWS2 in previous 24 hrs	NEWS2 fio2	Additive NEWS2 fio2 + max NEWS2 fio2 in preceding 24 hrs
All Respiratory 2015-2017	Area under ROC	0.888 (0.881-0.895)	0.902 (0.895-0.909)	0.890 (0.882-0.897)	0.901 (0.894-0.908)
	Area under PR curve	0.140	0.144	0.158	0.167
Scale 2 Cohort 2015-2017	Area under ROC	0.857 (0.838-0.877)	0.880 (0.862-0.898)	0.865 (0.847-0.884)	0.883 (0.866-0.900)
	Area under PR curve	0.115	0.118	0.123	0.132
All Respiratory 2017-2019	Area under ROC	0.880 (0.873-0.887)	0.898 (0.892-0.905)	0.887 (0.881-0.894)	0.900 (0.894-0.907)
	Area under PR curve	0.134	0.144	0.145	0.155
Scale 2 cohort 207-2019	Area under ROC	0.878 (0.860-0.897)	0.903 (0.885-0.921)	0.880 (0.861-0.899)	0.899 (0.881-0.918)
	Area under PR curve	0.100	0.121	0.102	0.128

Appendix Table 2 a)- Cut points for escalation with additive score combining maximum score in previous 24 hours and current NEWS2 matched to the NEWS2 score with equivalent sensitivity

2015-2017 Derivation Cohort						2017-2019 Validation Cohort					
Additive Score			NEWS 2			Additive Score			NEWS 2		
Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity
0	1.00	0.00	0	1.00	0.00	0	1.00	0.00	0	1.00	0.00
1	1.00	0.02				1	1.00	0.01			
2	1.00	0.06				2	1.00	0.04	1	1.00	0.09
3	1.00	0.12	1	1.00	0.12	3	1.00	0.09			
4	1.00	0.19				4	0.99	0.16			
5	0.99	0.27	2	0.99	0.27	5	0.99	0.23	2	0.99	0.24
6	0.99	0.35				6	0.99	0.31			
7	0.98	0.43	3	0.97	0.43	7	0.98	0.39			
8	0.96	0.51				8	0.97	0.47	3	0.97	0.40
9	0.95	0.58	4	0.93	0.60	9	0.95	0.55			
10	0.93	0.65				10	0.94	0.62	4	0.93	0.56

11	0.90	0.72				11	0.91	0.69			
12	**0.87**	**0.78**	**5**	**0.87**	**0.72**	**12**	**0.89**	**0.75**	**5**	**0.88**	**0.69**
13	0.83	0.83	6	0.80	0.82	13	0.85	0.80			
14	0.79	0.87				14	0.80	0.85	6	0.80	0.80
15	**0.72**	**0.90**	**7**	**0.68**	**0.90**	**15**	**0.74**	**0.88**	**7**	**0.70**	**0.88**
16	0.64	0.92				16	0.68	0.91			
17	0.57	0.94	8	0.57	0.94	17	0.61	0.94	8	0.59	0.93
18	0.50	0.96	9	0.45	0.97	18	0.53	0.95	9	0.46	0.96
19	0.43	0.97				19	0.44	0.97			
20	0.36	0.98	10	0.32	0.98	20	0.36	0.98	10	0.32	0.98
21	0.28	0.99				21	0.29	0.99			
22	0.22	0.99	11	0.20	0.99	22	0.23	0.99	11	0.23	0.99
23	0.16	1.00				23	0.18	0.99			
24	0.13	1.00	12	0.13	1.00	24	0.14	1.00	12	0.15	1.00
25	0.08	1.00	13	0.07	1.00	25	0.11	1.00			
26	0.06	1.00				26	0.09	1.00	13	0.09	1.00
27	0.04	1.00	14	0.04	1.00	27	0.06	1.00	14	0.04	1.00
28	0.02	1.00	15	0.02	1.00	28	0.04	1.00			
29	0.02	1.00				29	0.02	1.00	15	0.02	1.00
30	0.01	1.00				30	0.01	1.00			
31	0.01	1.00	16	0.00	1.00	31	0.01	1.00	16	0.00	1.00
32	0.00	1.00				32	0.00	1.00	17	0.00	1.00
33	0.00	1.00	17	0.00	1.00	33	0.00	1.00			

Appendix Table 2 b)- Cut points for Additive NEWS2 score matched to NEWS2 score with closest matched sensitivity for death in 24 hours- Scale 2 cohort with a diagnosis of COPD

2015-2017 Derivation Cohort	2017-2019 Validation Cohort

Additive Score			NEWS 2			Additive Score			NEWS 2		
Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity
0	1.00	0.00	0	1.00	0.00	0	1.00	0.00	0	1.00	0.00
1	1.00	0.02				1	1.00	0.01	1	1.00	0.13
2	1.00	0.05				2	1.00	0.04			
3	1.00	0.12	1	1.00	0.14	3	0.98	0.11			
4	0.99	0.20				4	0.98	0.18			
5	0.99	0.28	2	0.97	0.30	5	0.98	0.27			
6	0.97	0.37				6	0.98	0.37			
7	0.95	0.46	3	0.95	0.47	7	0.97	0.46	2	0.97	0.30
8	0.93	0.55				8	0.95	0.56	3	0.95	0.47
9	0.91	0.63				9	0.93	0.65	4	0.90	0.66
10	0.87	0.71	4	0.87	0.64	10	0.89	0.73			
11	0.82	0.77				11	0.85	0.80			
12	**0.77**	**0.83**	**5**	**0.77**	**0.77**	**12**	**0.83**	**0.85**	**5**	**0.81**	**0.79**
13	0.70	0.87	6	0.68	0.87	13	0.76	0.89	6	0.71	0.88
14	0.66	0.91				14	0.70	0.92			
15	**0.58**	**0.94**	**7**	**0.53**	**0.93**	**15**	**0.63**	**0.95**	**7**	**0.57**	**0.94**
16	0.51	0.96				16	0.54	0.96			
17	0.44	0.97	8	0.43	0.96	17	0.42	0.98	8	0.42	0.97
18	0.38	0.98				18	0.35	0.98	9	0.28	0.99
19	0.32	0.99	9	0.32	0.98	19	0.25	0.99	10	0.19	0.99
20	0.28	0.99	10	0.24	0.99	20	0.17	0.99			
21	0.22	0.99				21	0.13	1.00	11	0.12	1.00
22	0.17	1.00	11	0.15	1.00	22	0.12	1.00			
23	0.11	1.00	12	0.10	1.00	23	0.10	1.00			
24	0.07	1.00	13	0.06	1.00	24	0.08	1.00	12	0.07	1.00
25	0.05	1.00				25	0.07	1.00			
26	0.03	1.00	14	0.03	1.00	26	0.05	1.00	13	0.05	1.00

27	0.02	1.00				27	0.03	1.00	14	0.03	1.00
28	0.01	1.00	15	0.01	1.00	28	0.02	1.00			
29	0.01	1.00				29	0.01	1.00	15	0.01	1.00
30	0.00	1.00	16	0.00	1.00	30	0.00	1.00			

References

1. Fund, T.K. *Key facts and figures about the NHS*. 2019 05/07/2021]; Available from: https://www.kingsfund.org.uk/audio-video/key-facts-figures-nhs.
2. Pimentel, M.A.F., et al., *A comparison of the ability of the National Early Warning Score and the National Early Warning Score 2 to identify patients at risk of in-hospital mortality: A multi-centre database study.* Resuscitation, 2019. **134**: p. 147-156.
3. Smith, G.B., et al., *The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death.* Resuscitation, 2013. **84**(4): p. 465-70.
4. Kipnis, P., et al., *Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU.* J Biomed Inform, 2016. **64**: p. 10-19.
5. Forster, S., et al., *Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to Nottingham University Hospitals Trust over a 2-year period.* BMJ Open, 2018. **8**(7): p. e020269.
6. Hodgson, L.E., et al., *A validation of the National Early Warning Score to predict outcome in patients with COPD exacerbation.* Thorax, 2017. **72**(1): p. 23-30.
7. Yiu, C.J., et al., *Into the night: factors affecting response to abnormal Early Warning Scores out-of-hours and implications for service improvement.* Acute Med, 2014. **13**(2): p. 56-60.
8. Spangfors, M., M. Molt, and K. Samuelson, *In-hospital cardiac arrest and preceding National Early Warning Score (NEWS): A retrospective case-control study.* Clin Med (Lond), 2020. **20**(1): p. 55-60.
9. Malycha, J., et al., *The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis.* Resuscitation, 2019. **139**: p. 192-199.
10. Jarvis, S.W., et al., *Are observation selection methods important when comparing early warning score performance?* Resuscitation, 2015. **90**: p. 1-6.
11. Royal College of Physicians of, L., *National Early Warning Score (NEWS) 2*, in *Standardising the assessment of acute-illness severity in the NHS*. 2018, Royal College of Physicians: Royal College of Physicians Website.
12. Echevarria, C., J. Steer, and S.C. Bourke, *Comparison of early warning scores in patients with COPD exacerbation: DECAF and NEWS score.* Thorax, 2019. **74**(10): p. 941-946.
13. Frank E. Harrell , J., *Regression Modeling Strategies*. 2nd ed. Springer Series In Statistics. 2015: Springer.
14. Brekke, I.J., et al., *The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review.* PLoS One, 2019. **14**(1): p. e0210875.
15. Churpek, M.M., R. Adhikari, and D.P. Edelson, *The value of vital sign trends for detecting clinical deterioration on the wards.* Resuscitation, 2016. **102**: p. 1-5.
16. Collins, G.S., et al., *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement.* BJOG, 2015. **122**(3): p. 434-43.
17. Saito T, R.M., *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets.* PLoS One, 2015. **4**(10).
18. Romero-Brufau S, H.J., Escobar GJ, Liebow M. , *Why the C-statistic is not informative to evaluate early warning scores and what metrics to use.* Crit Care., 2015. **13**(19).
19. Fang, A.H.S., W.T. Lim, and T. Balakrishnan, *Early warning score validation methodologies and performance metrics: a systematic review.* BMC Med Inform Decis Mak, 2020. **20**(1): p. 111.

20. Hydes, T.J., et al., *National Early Warning Score Accurately Discriminates the Risk of Serious Adverse Events in Patients With Liver Disease.* Clin Gastroenterol Hepatol, 2018. **16**(10): p. 1657-1666 e10.
21. Badriyah, T., et al., *Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS).* Resuscitation, 2014. **85**(3): p. 418-23.