

Predicting outcome in acute respiratory admissions using patterns of National Early Warning Scores

Authors: Sarah Forster,^A Tricia M McKeever,^B Matthew Churpek,^C Sherif Gonem^D and Dominick Shaw^E

ABSTRACT

Aims

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 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 7,487 admissions and the validation cohort included 8,739 admissions. Adding the maximum score in the preceding 24 hours to the most recently recorded NEWS2 improved area under the receiver operating characteristic curve for death in 24 hours from 0.888 (95% confidence interval (CI) 0.881–0.895) to 0.902 (95% CI 0.895–0.909) in the overall respiratory population.

Conclusion

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

KEYWORDS: recognise, rescue, early warning scores, critical care, clinical deterioration

DOI: 10.7861/clinmed.2022-0074

Introduction

The National Early Warning Score (NEWS), now in its second iteration (NEWS2), is deployed in 76% of the 223 acute hospitals trusts and all 10 ambulance trusts across the NHS in England, as well as 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.¹ NEWS2 and its predecessor (NEWS) have been retrospectively validated through several large outcomes-linked vital signs data sets and are more accurate at predicting clinical deterioration than prior early warning score algorithms.^{2–5}

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.^{6,7} 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), inappropriate oxygen use or misplaced clinical reassurance in an unstable patient.^{8,9}

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 – fraction of inspired oxygen (FiO_2) to determine the additional benefit of a pattern in scores if factors (such as FiO_2) were included in a graded manner for when NEWS2 is reviewed in 2023.¹⁰

Authors: ^Aclinical research fellow in respiratory medicine, Nottingham University Hospitals NHS Trust, Nottingham, UK and University of Nottingham School of Medicine, Nottingham, UK; ^Bprofessor of epidemiology and medical statistics, University of Nottingham, Nottingham, UK; ^Cassociate professor in pulmonary and critical care medicine, University of Wisconsin-Madison School of Medicine and Public Health, Madison, USA; ^Dconsultant in respiratory medicine, Nottingham University Hospitals NHS Trust, Nottingham, UK; ^Eprofessor and honorary consultant in respiratory medicine, Nottingham University Hospitals NHS Trust, Nottingham, UK and University of Nottingham School of Medicine, Nottingham, UK

Methods

Source of data

Approval was given by the UK Health Research Authority (HRA; IRAS ID 270837) and the Nottingham University Hospitals NHS Trust's Caldicott guardian, the research and innovation team, and the information governance department (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 01 April 2015 and 31 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, new onset confusion, voice, pain and unresponsive' (ACVPU), respiratory rate measured in breaths per minute, oxygen saturations (%), heart rate (beats per minute), blood pressure (mmHg), temperature (°C), FiO₂ (%) or flow rate (L/min), and urine output (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 International Classification of Diseases 10th revision (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.

Participants

All admitted patients who were aged 18 years or older within the study period admitted to and discharged from respiratory medicine were included. Any vital signs coded as 'end-of-life care' (ie 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.¹¹ 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 dictates an emergency response with continuous monitoring.² 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, identified by presence of an ICD10 code for COPD at any point during admission.¹²

Statistical analysis

The 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 the pattern of NEWS2 score, both over the preceding 24 hours and throughout admission, to assess the ability to predict death within 24 hours of an observation. Death within 24 hours was used

Table 1. Definitions relating to the National Early Warning Score 2 used in the study

Term	Definition
Observation	Set of vital signs recorded at the bedside. Taken together, each set is amalgamated to the NEWS2 score.
Vital sign score	Indicates how far each vital sign deviates from a set normal range, calculated at collection of each observation with a weighting of 0–3.
NEWS2 score	NEWS2 was published by the RCP in 2017 and mandated for use across the NHS in the UK.
Scale 1	NEWS2 is an aggregate early warning score. NEWS2 is a continuous variable from a minimum score of 0 to a maximum of 20.
Scale 2	Oxygen target scale for NEWS2: for use in patients with no evidence of type 2 respiratory failure. Target saturations 94%–98%.
Cut points	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%.
	NEWS2 scores at which certain actions are advised as per the protocol published by the RCP:
	NEWS2 score of 5 or more: Minimum hourly observations; registered nurse to immediately inform medical team and request urgent assessment within 1 hour by a clinician with core competencies in care of acutely ill patients; provide clinical care in an environment with monitoring facilities.
	NEWS2 score of less than 5 but a single score of 3 in any one category: Separate category in original scoring protocol but clinically treated the same as a score of 5 or 6.
	NEWS2 score 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.

COPD = chronic obstructive pulmonary disease; NEWS2 = National Early Warning Score 2; RCP = Royal College of Physicians.

as the outcome rather than intensive care unit (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 Frank and Harrell.¹³ Univariate models were created using the uvs 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 that could be applied in less sophisticated systems, a predictive additive model was created using the maximum NEWS2 score in the preceding 24 hours and most recently recorded NEWS2 score. This additive approach combining the maximum score in the preceding 24 hours and the 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 according to their methods.¹⁰

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 (ROC) curve and area under precision recall (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, a higher area under the PR curve indicates a better 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.

Regulatory approval

This project was provided with HRA approval- IRAS project ID: 270837 Protocol number: 19074. As the study was retrospective, and all data were collected during routine care and anonymised prior to extraction, it was not necessary to gain full REC approval.

Results

Study population

There were 7,487 completed admissions from 5,136 individual patients to the Nottingham University Hospitals NHS Trust respiratory department during the initial 2-year study derivation period from April 2015 to March 2017, and 8,739 admissions from 5,928 individual patients during the validation period from April 2017 to March 2019 (Fig 1). Admission demographics are detailed in Table 2.

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

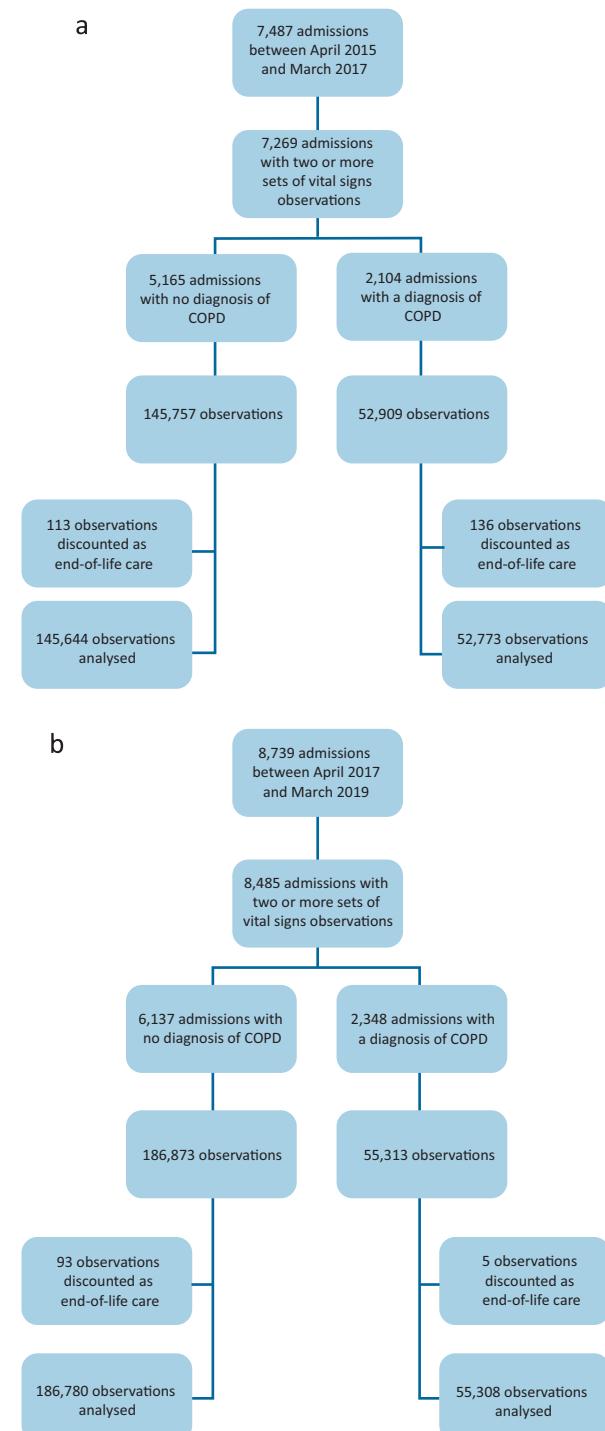


Fig 1. Patients with respiratory disease completing admission. a) Derivation cohort, April 2015 – March 2017. b) Validation cohort, April 2017 – March 2019. COPD = chronic obstructive pulmonary disease.

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.

Table 2. Cohort demographics

	Derivation cohort, April 2015 – March 2017			Validation cohort, April 2017 – March 2019		
	Respiratory (total)	Non-COPD (Scale 1)	COPD (Scale 2)	Respiratory (total)	Non-COPD (Scale 1)	COPD (Scale 2)
Admissions, n	7,269	5,165	2,104	8,485	6,351	2,381
Women, n (%)	3,953 (54.4)	2,775 (53.7)	1,178 (56.0)	4,718 (54.0)	3,402 (53.5)	1,316 (55.3)
Age, years, median (IQR)	71 (61–81)	71 (61–81)	71 (61–76)	71 (56–76)	66 (51–76)	71 (61–76)
Length of stay, days, median (IQR)	4 (2–8)	4 (2–8)	3 (2–7)	3 (1–7)	3 (1–7)	3 (1–6)
In-hospital mortality, n (%)	413 (5.7)	328 (6.4)	85 (4.0)	470 (5.5)	398 (6.5)	72 (3.1)

COPD = chronic obstructive pulmonary disease; IQR = interquartile range.

Area under the ROC curve for NEWS2 in the overall respiratory population was 0.888 (95% confidence interval (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 (ie 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. Sixteen of the 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 with 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.

Thirty-nine observations met the criteria for clinical review/escalation at a cut point of 5 per outcome identified of death within 24 hours. Forty-one 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 area under PR curve was 0.100 in the validation cohort. The odds ratio increase 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.

Using the 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 (Fig 2a).

Table 3. Sensitivity and specificity values of the National Early Warning Score 2

	Derivation cohort, April 2015 – March 2017		Validation cohort, April 2017 – March 2019	
	Sensitivity	Specificity	Sensitivity	Specificity
Cut point ≥5				
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 any 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 (Scale 2)	0.53	0.93	0.57	0.94

COPD = chronic obstructive pulmonary disease; NEWS2 = National Early Warning Score 2.

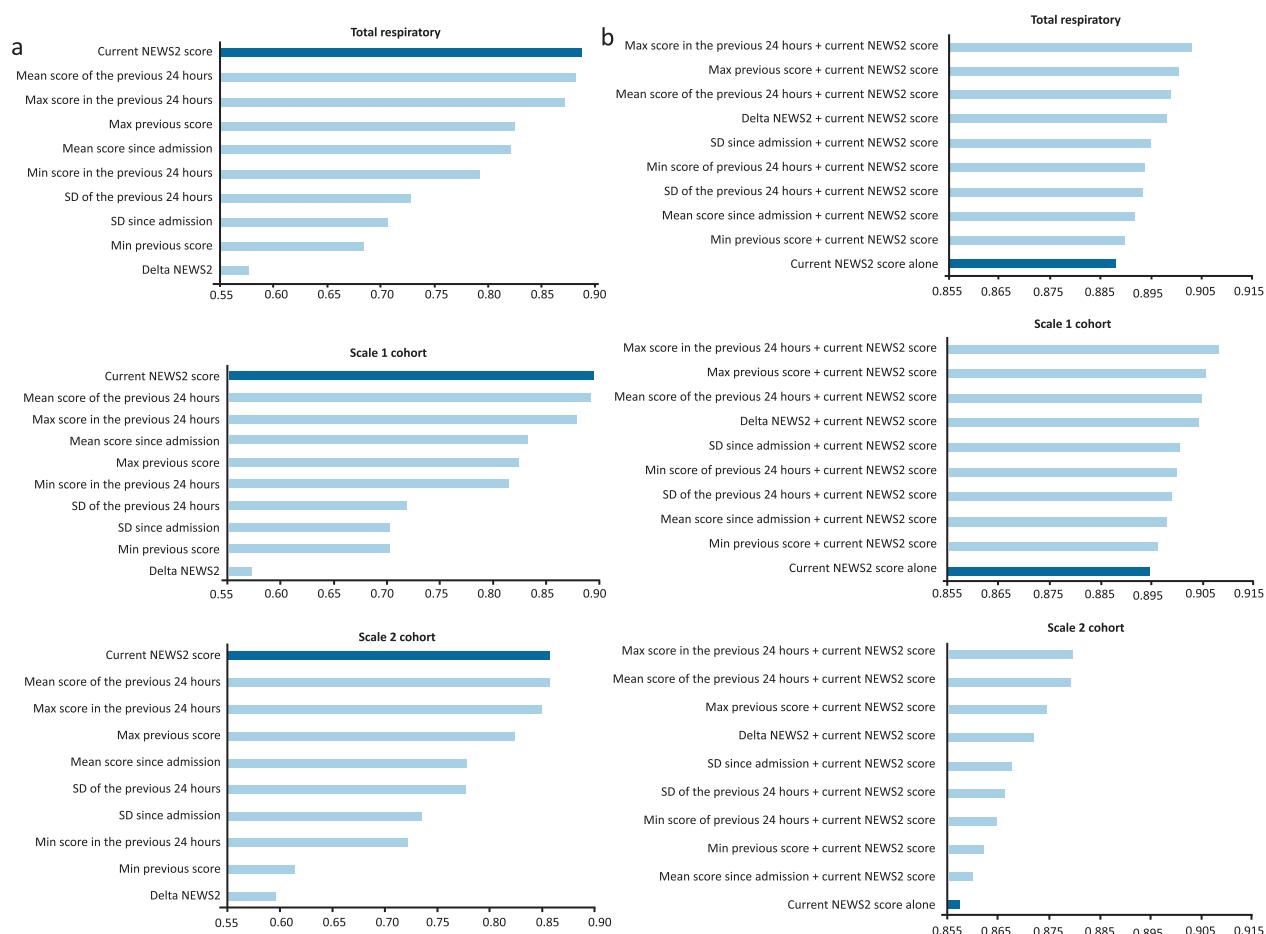


Fig 2. Comparison of area under receiver operating characteristic curves for restricted cubic spline models of National Early Warning Score 2 pattern and existing score for outcome of death within 24 hours. a) Univariate comparison. b) Bivariate comparison. Max = maximum; Min = minimum; NEWS2 = National Early Warning Score 2; SD = standard deviation.

Improvement in prognostic ability was seen in all bivariate restricted cubic spline models compared with NEWS2 alone (Fig 2b). The model with highest prognostic ability for death within 24 hours combined the maximum score in the preceding 24 hours with the 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 the maximum score in the preceding 24 hours and the 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 and 0.880 (95%CI 0.862–0.898) in the Scale 2 cohort. This is also reflected in the area under PR curves shown in Table 4. As PR 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 in place of an equivalent NEWS2 cut point of 5 would result in 7,035 (9.2%) fewer scores meeting the threshold for escalation in the overall population and 1,366 (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.

It has been suggested that the addition of a graded FiO_2 score to future iterations of NEWS could improve risk prediction.¹⁰ In this population, application of a previously described NEWS- FiO_2 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- FiO_2 demonstrated improvement in discrimination when the maximum score in the preceding 24 hours was applied to the total respiratory population and Scale 2 cohorts (supplementary material S1, Table S1).

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 the most recently recorded NEWS2 with the 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 FiO_2 .

Table 4. Areas under receiver operating characteristic curve and precision recall curve for the National Early Warning Score 2 and additive score combining the current National Early Warning Score 2 with the maximum NEWS2 in the preceding 24 hours

	Derivation cohort, April 2015 – March 2017		Validation cohort, April 2017 – March 2019	
	Area under ROC curve (95% CI)	Area under PR curve	Area under ROC curve (95% CI)	Area under PR curve
Total respiratory				
NEWS2	0.888 (0.881–0.895)	0.140	0.880 (0.873–0.887)	0.133
NEWS2 and the maximum score within previous 24 hours	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
NEWS2 and the maximum score within previous 24 hours	0.880 (0.862–0.898)	0.118	0.903 (0.885–0.921)	0.122

CI = confidence interval; NEWS2 = National Early Warning Score 2; PR = precision recall; ROC = receiver operating characteristic.

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 'alert, voice, pain and unresponsive' (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 that 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 the skew between observations and outcomes and, thereby, the potential for excessive workload and unnecessary intervention.

Echevarria *et al* analysed the performance of NEWS2 Scale 2 when applied to patients with COPD.¹² Scale 2 led to a reduction in scores reaching escalation thresholds and an improved discrimination when compared with the original NEWS score (area under ROC curve 0.72 vs 0.65), and it did not fail to identify any outcomes escalated by Scale 1. Pimentel *et al* used a combination of coding and oxygen prescriptions to identify patient cohorts at risk of hypercapnic respiratory failure and confirmed hypercapnic respiratory failure.³ 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 with 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 our 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 outcomes 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 the 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 the maximum NEWS2 in the preceding 24 hours, corresponding in sensitivity to a NEWS2 score of 5, would result in 7,035 (9.2%) fewer scores meeting the threshold for escalation in the overall population and 1,366 (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 (supplementary material S1, Table S2).

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 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 for reporting performance of predictive scores and the STROBE statement for reporting cohort studies were applied through design, analysis and reporting.¹⁶ Lastly, while the 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 PR curves give added information, so both are included.^{5,17–19}

Limitations include that 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 and has become a recognised approach to evaluating early warning scores.^{4,10,11,20,21}

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 the 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 for all healthcare systems that strive to balance the seesaw of resource limitations against the need to predict, react to and prevent clinical deterioration in hospitalised patients. ■

Supplementary material

Additional supplementary material may be found in the online version of this article at www.rcpjournals.org/clinmedicine: S1 – Additional tables.

Funding

Dr Sarah Forster had her salary funded by a Nottingham Hospitals Charity Fellowship. Dr Matthew Churpek reports grants from NIH/NIDA (R01 DA051464), from DOD/PRMRP, W81XWH-21-1-0009, from NIH/NIA (R21 AG068720), from NIH/NIGMS (R01 GM123193), from NIH/NIDDK (R01 DK126933), from EarlySense (Tel Aviv, Israel) and from NIH/NHLBI (R01 HL157262), and has a patent pending (ARCD.P0535US.P2) to The University of Chicago related to clinical deterioration risk prediction algorithms for hospitalised patients.

References

- 1 The King's Fund. *Key facts and figures about the NHS*. The King's Fund, 2019. www.kingsfund.org.uk/audio-video/key-facts-figures-nhs [Accessed 05 July 2021].
- 2 Royal College of Physicians. *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS*. RCP, 2017. www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2
- 3 Pimentel MAF, Redfern OC, Gerry S 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;134:147–56.
- 4 Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. 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;84:465–70.
- 5 Kipnis P, Turk BJ, Wulf DA 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;64:10–19.
- 6 Forster S, Housley G, McKeever TM, Shaw DE. 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:e020269.
- 7 Hodgson LE, Dimitrov BD, Congleton J et al. A validation of the National Early Warning Score to predict outcome in patients with COPD exacerbation. *Thorax* 2017;72:23–30.
- 8 Yiu CJ, Khan SU, Subbe CP, Tofeeck K, Madge RA. Into the night: factors affecting response to abnormal Early Warning Scores out-of-hours and implications for service improvement. *Acute Med* 2014;13:56–60.
- 9 Spangfors M, Molt M, Samuelson K. In-hospital cardiac arrest and preceding National Early Warning Score (NEWS): A retrospective case-control study. *Clin Med* 2020;20:55–60.
- 10 Malycha J, Farajidavar N, Pimentel MAF et al. The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis. *Resuscitation* 2019;139:192–9.
- 11 Jarvis SW, Kovacs C, Briggs J et al. Are observation selection methods important when comparing early warning score performance? *Resuscitation* 2015;90:1–6.
- 12 Echevarria C, Steer J, Bourke SC. Comparison of early warning scores in patients with COPD exacerbation: DECAF and NEWS score. *Thorax* 2019;74:941–6.
- 13 Frank E, Harrell J. *Regression modeling strategies*, 2nd ed. Springer, 2015.
- 14 Brekke IJ, Puntervoll LH, Pedersen PB, Kellett J, Brabrand M. The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. *PLoS One* 2019;14:e0210875.
- 15 Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation* 2016;102:1–5.
- 16 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BJOG* 2015;122:434–43.
- 17 Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10:e0118432.
- 18 Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care* 2015;19:285.
- 19 Fang AHS, Lim WT, Balakrishnan T. Early warning score validation methodologies and performance metrics: a systematic review. *BMC Med Inform Decis Mak* 2020;20:111.
- 20 Hydes TJ, Meredith P, Schmidt PE et al. National Early Warning Score accurately discriminates the risk of serious adverse events in patients with liver disease. *Clin Gastroenterol Hepatol* 2018;16:1657–66.e10.
- 21 Badriyah T, Briggs JS, Meredith P et al. Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS). *Resuscitation* 2014;85:418–23.

Address for correspondence: Dr Sarah Forster, University of Nottingham, Clinical Sciences Building, Edwards Lane, Nottingham NG5 1PB, UK.

Email: sarahforster@nhs.net