Protocol for "External validation of the National Early Warning Score 2 (NEWS2) prediction of in-hospital death in patients with type II respiratory failure: a multi-centre database study"

Date and Version No: 13th February 2018; v1.0

Chief Investigator:

Name	Peter Watkinson
Position	Consultant Intensive Care Physician / Associate Professor of Intensive Care Medicine
Employing organisation	Oxford University Hospitals NHS Trust / University of Oxford
Address	Kadoorie Centre, Level 3, John Radcliffe Hospital, Headley Way, Oxford, OX3 9DU
Telephone	01865 223101
Email	peter.watkinson@ndcn.ox.ac.uk

Co-Investigators:

Gary Collins, University of Oxford

Stephen Gerry, University of Oxford

James Malycha, University of Oxford

Marco Pimentel, University of Oxford

David Prytherch, University of Portsmouth

Oliver Redfern, University of Portsmouth

Paul Schmidt, Portsmouth Hospitals NHS Trust

Gary Smith, Bournemouth University

Chief Investigator Signature:

Conflicts of interest

There are no conflicts of interest

Confidentiality Statement

Abstract

Background: The National Early Warning Score (NEWS), a vital-signs-based early warning score (EWS), has been suggested to perform poorly in patients with hypercapnic respiratory failure. A newly updated version (NEWS2) includes modified peripheral oxygen saturation (S_pO_2) thresholds for such patients and additional weightings for higher S_pO_2 values if they are receiving oxygen therapy. NEWS2 has been mandated for use in acute NHS hospitals. However, these changes await validation. Prior to the publication of NEWS2, the Chronic Respiratory Early Warning Score (CREWS) had been proposed as an adjustment to NEWS for these patients.

Methods: Retrospective observational cohort study using the HAVEN database of admissions to four Oxford University Hospitals January - December 2016 and Portsmouth Hospitals NHS Trust January 2012- December 2016. Our primary endpoint will be the ability of NEWS2, NEWS and CREWS to predict in-hospital death within 24 hours of an observation set in patients with and without type II respiratory failure. Secondary endpoints will include cardiac arrest, intensive care admission and the first of intensive care admission, cardiac arrest or in-hospital death. We will conduct subgroup analyses using different methods for identifying patients with or at risk of type II respiratory failure. We will use the C-statistic to assess discrimination and report sensitivities, specificities and predictive values at recognized cut-offs. To assess the incremental change between NEWS and NEWS2 of the new thresholds for COPD we will present change in the C-statistic.

Conclusion: We will provide the first multi-centre assessment of whether the changes in NEWS2 have led to improved discrimination of adverse outcomes in patients with type II respiratory failure, without the disadvantaging the performance in patients without this condition.

Introduction

Vital signs-based aggregate EWS work by assigning weights to each vital sign according to the degree of deviation from assumed normal values. These weights are then summed to produce the final composite EWS value. Such EWS have been recommended for use in acute patients UK hospitals[1,2]. In 2012 the Royal College of Physicians of London (RCPL) published a proposed National Early Warning Score (NEWS).[3] NEWS has since undergone validation in datasets other than the Portsmouth dataset on which it was derived[4–6]. In NEWS, oxygen saturations receive increasing weights for values of 95% or less and oxygen therapy receives a flat weight. However, guidance for the management of patients with type II respiratory failure suggests lower oxygen saturations (88-92%) should be targeted [7,8]. This guidance also applies to patients “at risk” of type II respiratory

failure prior to blood gas analysis[7]. It has been suggested that NEWS therefore inappropriately weights patients with type II respiratory failure who have oxygen saturations 88-95%[9–11]. Some authors suggest that this weighting risks inappropriate oxygen therapy for these patients, with potential deleterious consequences[9,10].

Recently, the RCPL published an evolution of NEWS, NEWS2[12]. Taking account of concerns about the performance of NEWS in patients with type II respiratory failure, they include a new S_pO_2 scoring system for these patients. Although the derivation of these thresholds is not presented, NEWS2 has been endorsed for use throughout the NHS[13]. Prior to NEWS2, a Chronic Respiratory Early Warning score (CREWS)[11] was published as a suggested adjustment to NEWS for patients with chronic hypoxaemia.

Before widespread adoption occurs, prediction models such as NEWS2 should be evaluated on a separate dataset, representative of the patients on whom the model may be used. Any effect on incremental performance of the additional thresholds should be assessed[14]. We will therefore assess NEWS2 performance in comparison to NEWS and CREWS on a large multi-centre retrospective dataset of vital signs. We will consider performance in three populations: those with documented type II respiratory failure, where NEWS2 is recommended for use; those "at risk" of type II respiratory failure, where target saturations of 88-92% are recommended prior to blood gas analysis (and NEWS2 would score 0 for these saturations) and in patients in neither of these groups, where the use of NEWS2 would be mistaken.

Methods

The database for this study was created with Health Research Authority (reference: 16/SC/0264) approval. We have written this protocol and will report our findings following the TRIPOD statement[14].

Source of data

A database of vital sign observations was collected from adult (aged at least 16 years) acute medical admissions to the Oxford University Hospitals group (January-December 2016) or to Portsmouth Hospitals NHS Trust (January 2011- December 2016) as part of the HAVEN dataset[15]. Clinical staff recorded patients' vital signs at the bedside using the System for Electronic Notification and Documentation (SEND, Drayson Techonolgies, www.draysontechnologies.com)[16] in Oxford and VitalPAC™ (System C Healthcare, www.systemc.com) in Portsmouth[17]. The following data were recorded: date and time of observation (automatically by SEND/VitalPAC™); the manual

measurement of heart rate, systolic and diastolic blood pressure, respiratory rate, body temperature, neurological status using the Alert-Verbal-Painful-Unresponsive (AVPU) scale, S_pO_2; and a record of whether or not supplemental oxygen support was given to the patient at the time of S_pO_2 measurement. The database also contains administrative and patient demographic information, and occurrences of patient death, cardiac arrests, unanticipated ICU admissions and discharge status (dead/alive), and their corresponding timings, identified from the hospital patient administration system and the ICU clinical information system. Prescription data from the electronic patient record is also available within the database for Oxford patients. We calculated NEWS, NEWS2 and CREWS scores retrospectively from the original vital signs data.

Participants

Four Oxford organization hospitals, The John Radcliffe Hospital - a large university hospital, the Horton General Hospital – a small district general hospital, The Churchill hospital – a large university cancer centre and the Nuffield Orthopaedic Hospital (January-December 2016)and Portsmouth Hospital NHS Trust (January 2012- December 2016), a large district general hospital will be included. All patients with electronically recorded vital signs will be considered for inclusion in the study.

Outcome

The primary outcome will be death within 24 hours of an observation set. Secondary outcomes will be cardiac arrest, ICU admission and the first of cardiac arrest, ICU admission or death within twenty-four hours of an observation set, all where sufficient outcomes existed (> 100)[14]. The original source of each of these outcomes lies within hospital patient administration systems. The employees inputting these outcomes were blind to the NEWS, NEWS2 and CREWS scores, which we calculated retrospectively using the original vital signs data.

Predictors

Each vital sign (heart rate, systolic blood pressure, temperature, S_pO_2, respiratory rate and conscious level) along with oxygen administration were recorded by clinical staff using SEND. These vital signs will be retrospectively assigned weights for NEWS, NEWS2 and CREWS scores. The sum of the weights results in the NEWS, NEWS2 and CREWS early warning score for each observation set. Weights and summary scores will be generated by computer algorithm without recourse to the outcomes of interest. NEWS and NEWS2 differ in that, for NEWS2, patients for whom the target oxygen saturation is 88-92% receive lower weights for low oxygen saturations and also receive weights for S_pO_2 considered too high when receiving oxygen (see appendix). SEND in Oxford uses a

modified early warning system, CEWS[18]. The modified CEWS assigns increasing weights to S_pO_2 less than 94% and does not weight S_pO_2 of 94% or above. Clinical staff entering the vital sign data were therefore unaware of the NEWS, NEWS2 or CREWS scores. VitalPAC™ in Portsmouth uses NEWS. In line with most previous vital-signs-based EWS publications, each vital sign set will be analysed as independently associated with the outcome.

Sample size

All available data on the database will be used to maximise the power and generalisability of the results.

Missing data

Where a vital sign was not recorded within a vital sign set, we will multiply impute the missing vital sign. Where the discharge status of the patient episode was not recorded, the episode will be excluded from analysis. Where a patient does not have at least one complete observation set they will be excluded from the analysis. Where a patient has no observation sets in the twenty four hours prior to discharge or death they will be excluded from the analysis. Patients discharged to home will be assumed to be alive at 24 hours.

Statistical analysis methods

We will assess discrimination using the c-statistic. We cannot assess calibration since the NEWS/NEWS2/CREWS do not give an estimate of absolute risk. We will assess the effect of suggested thresholds (a summary NEWS/NEWS2/CREWS score of 5 or 7 or a score from a single vital sign of 3[12]) by reporting sensitivity, specificity and predictive values. We will also plot the distribution of NEWS/NEWS2/CREWS scores for those with and without the outcome. We will plot S_pO_2 distributions for different populations, including those "at risk" of type II respiratory failure in Oxford and Portsmouth. We will assess the incremental change between NEWS and NEWS2, and CREWS and NEWS2, by presenting change in the C-statistic.

Risk groups

We will define three risk groups:

1. Patients with recorded Type II respiratory failure in the *Adult Oxygen Prescription* form of the concurrent admission.

2. Patients “at risk” of Type II respiratory failure identified using the International Statistical Classification of Diseases and Related Health Problems-10 (ICD-10) classification for their concurrent or prior admission, with either of the following groups of diagnosis codes:

    2.1. J40-J44 (typically, 88% coded as J44) - patients with Chronic Obstructive Pulmonary Disease (COPD); or

    2.2. J47 – patients with Bronchiecstasis; or

    2.3. E84 – patients with Cystic Fibrosis; or

    2.4. E66 – patients with obesity and/or obesity hypoventilation syndrome; or

    2.5. G12, G47.3, G70-G71, M95.4, or Q67.8, with J96 – patients with respiratory failure (J96) and one of the following conditions: spinal muscular atrophy and other motor neuron disease (G12), sleep apnea (G47.3), myasthenia gravis and other myoneural disorders (G70), muscular dystrophies (G71), acquired deformity of chest and rib (M95.4), or other congenital deformities of chest (Q67.8).

3. Patients not at risk of Type II respiratory failure, in neither of groups 1 and 2.

Development versus validation datasets.

NEWS was originally developed using a dataset from Portsmouth Hospitals NHS Trust’s Medical Assessment Unit (MAU) (May 2006 - June 2008). The HAVEN database includes separate data from all admissions to Portsmouth Hospital NHS Trust (January 2011- December 2016) and four Oxford organization hospitals, The John Radcliffe Hospital - a large university hospital, the Horton General Hospital – a small district general hospital, The Churchill hospital – a large university cancer centre and the Nuffield Orthopaedic Hospital (January-December 2016). Vital sign data for all sites is present from admission to discharge/death, rather than only in the MAU setting as in the original study. NEWS2 is recommended for use in all the included settings.

Discussion

Main findings

Our study will report NEWS2 performance in comparison to NEWS and CREWS.

Strengths

Our study will be the first multi-centre study to investigate the performance of NEWS2 on the patient groups for which it is intended. Our use of the TRIPOD statement to guide our work will ensure our findings are both robust and interpretable. Robust electronic data capture should allow us to consider subgroups of patients admitted with or at risk of type II respiratory that have not previously been undertaken. Unlike previous studies[19] electronic capture of vital signs will include those undertaken in the emergency department, so capturing the whole patient journey.

Limitations

We will rely on diagnostic codes, oxygen prescriptions and use of non-invasive ventilation to identify and classify patients with type II respiratory failure, so patients may have been misclassified. However, COPD coding has been shown to be relatively reliable[20], suggesting coding for those at risk of type II respiratory failure may be reliable. In the case of oxygen prescriptions, the understanding of the prescribing clinician is recorded, so it seems likely the same understanding would underlie the choice of score used Our database does not include documentation of new confusion, so we could not take account of this change in NEWS2. However, as new confusion was not part of NEWS, our paper will demonstrate the effect of the changes in oxygen saturation scoring for patients with type II respiratory failure clearly.

Interpretation

The endorsement by the RCPL and NHS England of the modified peripheral oxygen saturation (S_pO_2) thresholds and additional weightings for higher S_pO_2 values in NEWS2 for patients at risk of hypercapnic respiratory failure without underpinning evidence of their impact makes our study urgent. The aim of generic EWS, such as NEWS, is to provide a simple tool, applicable across all acutely unwell populations that flags risk of deterioration based on vital signs. Including different scores for diagnostic subgroups creates a more complex protocol. Increasing score complexity has to be shown to improve performance for it to be worthwhile. It should also not (significantly) disadvantage the larger subgroup that does not require the more complex protocol. For computerized early warning systems, this has been demonstrated for adding laboratory values[21], but not so far for diagnostic subgroups. Avoidance of complexity is important for NEWS, which is

designed to be easily undertaken on paper. Here it is clear that chart design significantly affects ability to determine which patients are critically unwell [22,23].

Underlying the changes in NEWS2 is an understanding that the recommendations for these patients to have oxygen therapy limited to 88-92% are robust. One randomized trial of uncontrolled oxygen pre-hospital has shown significant harm from this therapy in patients with COPD, in comparison to oxygen titrated to an 88-92% threshold[24]. However, the differences in care required between the two treatment arms of this study made it highly likely to suffer from a Hawthorne effect (where the undertaking of the study, rather than the intervention, contributes to differences in outcome). Differences in controlled targets *in hospital* have not been effectively tested. Whether NEWS pushes clinicians toward oxygen saturations of 96%, as suggested[9,12] (or modified CEWS to 94%), rather than simply suggesting to them that the patient is at risk is unclear. Even if this were the case, whether running these patients at a controlled target of 94-96% in the monitored environment in hospital would be beneficial or harmful is unclear.

The changes proposed in NEWS2 are essentially those suggested by the Salford-NEWS EWS group[10], who contributed to the NEWS2 report. This has had limited, single-centre analysis previously, in a sub-group restricted to patients with COPD[19]. Our study differs in its multi-centre nature and in our ability to use an extensive electronic database to determine different subgroups of patients with type II respiratory failure, ensuring we will consider all likely groups in which NEWS2 could be applied.

Implications

Our study will be the first multi-centre study able to investigate the performance of NEWS2 in the type II respiratory failure cohort and sub-cohorts for which it is intended. It should inform whether national adoption is indicated.

SUPPLEMENTARY INFORMATION

This protocol can be found on the Oxford University Research Archive (ORA: www.ora.ox.ac.uk), uploaded 28th February 2018.

This work was supported by the NIHR Biomedical Research Centre, Oxford.

References

1. Scotland HI. SIGN 13. Care of deteriorating patients. 2014;(May).
2. NICE Clinical Guidelines. Acutely ill adults in hospital: recognising and responding to deterioration. 2007;(July):1–30.
3. Royal College of Physicians (London). National Early Warning Score (NEWS). 2012;(July).
4. Dawes TR, Cheek E, Bewick V, Dennis M, Duckitt RW, Walker J, Forni LG, Venn R. Introduction of an electronic physiological early warning system: Effects on mortality and length of stay. *British Journal of Anaesthesia*. 2014;113(4):603–609.
5. Silcock DJ, Corfield AR, Gowens PA, Rooney KD. Validation of the National Early Warning Score in the prehospital setting. *Resuscitation*. 2015;89(C):31–35.
6. Churpek MM, Snyder A, Han X, Sokol S, Pettit N, Howell MD, Edelson DP. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside theintensive care unit. *American Journal of Respiratory and Critical Care Medicine*. 2017;195(7):906–911.
7. O'Driscoll BR, Howard LS, Earis J, Mak V. British Thoracic Society Guideline for oxygen use in adults in healthcare and emergency settings. *BMJ Open Respiratory Research*. 2017;4(1):e000170.
8. Beasley R, Chien J, Douglas J, Eastlake L, Farah C, King G, Moore R, Pilcher J, Richards M, Smith S, Walters H. Thoracic Society of Australia and New Zealand oxygen guidelines for acute oxygen use in adults: Swimming between the flags. *Respirology*. 2015;20(8):1182–1191.
9. O'Driscoll BR, Grant K, Green D, Edeghere S, Bakerly ND, Murphy P, Edwards V, Turkington PM. Clinical and scientific letters. *Clinical Medicine*. 2014;14(6):695–696.
10. Kane B, Decalmer S, Murphy P, Turkington P, O'Driscoll B. The proposed National Early Warning System (NEWS) could be hazardous for patients who are at risk of hypercapnic respiratory failure. *Thorax*. 2012;67(Suppl 2):A16.
11. Eccles SR, Subbe C, Hancock D, Thomson N. CREWS: Improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia. *Resuscitation*. 2014;85(1):109–111.
12. Royal College of Physicians. *National Early Warning Score (NEWS) 2. Standardising the assessment of acute-illness severity in the NHS.* London; 2017.
13. NHS England approves NEWS2 to recognise deteriorating patients. www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2, accessed 12/02/2018.
14. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162(1):55–63.
15. HAVEN — Nuffield Department of Clinical Neurosciences. www.ndcn.ox.ac.uk/research/critical-care-research-group-kadoorie-centre/research-studies/hospital-alerting-via-electronic-noticeboard-haven/academic-publications-from-the-haven-project, accessed 12/02/2018.
16. Wong D, Bonnici T, Knight J, Morgan L, Coombes P, Watkinson P. SEND: A system for electronic notification and documentation of vital sign observations. *BMC Medical Informatics and Decision Making*. 2015;15(1):68.

17. Smith GB, Prytherch DR, Schmidt P, Featherstone PI, Knight D, Clements G, Mohammed MA. Hospital-wide physiological surveillance–A new approach to the early identification and management of the sick patient. *Resuscitation*. 2006;71(1):19–28.

18. Tarassenko L, Clifton DA, Pinsky MR, Hravnak MT, Woods JR, Watkinson PJ. Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation*. 2011;82(8):1013–1018.

19. Hodgson LE, Dimitrov BD, Congleton J, Venn R, Forni LG, Roderick PJ. A validation of the National Early Warning Score to predict outcome in patients with COPD exacerbation. *Thorax*. 2017;72(1):23–30.

20. Quint JK, Mullerova H, DiSantostefano RL, Forbes H, Eaton S, Hurst JR, Davis K, Smeeth L. Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). *BMJ Open*. 2014;4(7):e005540–e005540.

21. Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. *Resuscitation*. 2017;123:86–91.

22. Christofidis MJ, Hill A, Horswill MS, Watson MO. Less is more: The design of early-warning scoring systems affects the speed and accuracy of scoring. *Journal of Advanced Nursing*. 2015;71(7):1573–1586.

23. Christofidis MJ, Hill A, Horswill MS, Watson MO. A human factors approach to observation chart design can trump health professionals' prior chart experience. *Resuscitation*. 2013;84(5):657–665.

24. Austin MA, Wills KE, Blizzard L, Walters EH, Wood-Baker R. Effect of high flow oxygen on mortality in chronic obstructive pulmonary disease patients in prehospital setting: randomised controlled trial. *Bmj*. 2010;341(oct18 2):c5462–c5462.