BMJ Open

The Impact of Early Warning Scores on Managing Patients with Respiratory Disease

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020269
Article Type:	Research
Date Submitted by the Author:	25-Oct-2017
Complete List of Authors:	Forster, Sarah; University of Nottingham School of Medicine, Respiratory Medicine Housley, Gemma; Nottingham University Hospitals NHS Trust McKeever, Tricia; University of Nottingham, Division of Epidemiology and Public Health Shaw, Dominick; University of Nottingham, Respiratory Medicine
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Evidence based practice
Keywords:	Thoracic medicine < INTERNAL MEDICINE, Adult intensive & critical care < INTENSIVE & CRITICAL CARE, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

The Impact of Early Warning Scores on Managing Patients with Respiratory Disease

Sarah Forster[1,2], Gemma Housley[3], Tricia M McKeever[4], Dominick E Shaw[2,3]

Author Affiliations

[1] NIHR Academic Clinical Fellow, University of Nottingham, Nottingham, UK

[2] Respiratory Research Unit, Division of Respiratory Medicine, University of Nottingham, Nottingham, UK

[3] Medical Informatics, East Midlands Academic Health Sciences Network, Nottingham, UK

[4] Division of Epidemiology, University of Nottingham, Nottingham, UK

Correspondence to Dr Sarah Forster, sarahforster@nhs.net

Word count: 1693

Abstract

Objective

Early Warning Scores (EWS) are used to monitor patients for signs of imminent deterioration. Although used in respiratory disease, EWS have not been well studied in this population, despite the underlying cardiopulmonary pathophysiology often present. We set out to examine the performance of two scoring systems in patients with respiratory disease.

Design A retrospective cohort analysis of vital signs observations was performed on all patients admitted to a tertiary respiratory unit over a 2 year period. To establish the performance of the National EWS (NEWS) we linked scores to outcome data, and retrospectively compared results to a locally adapted EWS.

Setting Nottingham University Hospitals NHS Trust respiratory wards. Data were collected from an integrated electronic observation and task allocation system employing a local EWS, which generates immediate mandatory referrals when vital signs scoring thresholds are met.

Outcome Measures We examined both the actual (for local EWS) and projected workload (for NEWS) created by the scores, and the sensitivity and specificity of the scores in predicting mortality based on outcome within 24 hours of a score generated by vital signs observations.

Results 8812 individual patient episodes occurred during the study period. Overall mortality was 5.9%. Applying NEWS retrospectively (versus local EWS) generated a significant eight fold increase in mandatory escalations, but had a higher sensitivity for predicting mortality at the cut points applied by the protocol.

Conclusions This study highlights the issues surrounding the use of EWS in patients with respiratory disease. The higher sensitivity and lower specificity of NEWS means that it acts like a d-dimer; a low score is useful in ruling out deterioration, a high score is less helpful in predicting mortality. Further work on the significance of changes in vital signs in patients with complex underlying comorbidity and greater understanding of the pathophysiology involved is needed.

Strengths and Limitations of this Study

- Data were obtained from a large clinical vital signs database with clear identification of specialty allowing for subgroup analysis.
- Granularity of data collection in the database allowed for reliable identification of patients meeting the exclusion criteria.
- Only 0.2% of the observations recorded during the study period were identified as being incomplete.
- The retrospective nature of study precludes conclusions relating to impact of introducing NEWS on mortality.

Background

In 2012 the Royal College of Physicians published the NEWS protocol in an attempt to standardise processes for identifying patients at risk of imminent deterioration [1]. EWS protocols guide decisions around patient care by mandating when a patient with evidence of pathophysiology, in the form of deranged vital signs, should be reviewed by a clinical member of staff, and therefore influence overall clinical workload and resource allocation for all in-patients. Patients with respiratory disease make up a large proportion of a hospital's in-patient population, however a previous small study found that chronic physiological disturbance caused by COPD may render NEWS

less discriminative than in an unselected medical population [2]; consequently attempts have been made to improve the score in this population [3].

Nottingham University Hospitals NHS Trust employs an electronic observations system with mandatory escalation based on an adapted EWS. We compared the potential impact in terms of workload and mortality prediction of using a locally designed EWS versus NEWS (see Figure 1) in patients with respiratory disease.

Figure 1- Escalation protocol for NEWS and Nottingham University Hospitals Early Warning Score

Methods

We performed a single centre retrospective analysis of all patients admitted to the respiratory department at Nottingham University Hospitals NHS Trust between 01/04/2015 and 31/03/2017. This is a tertiary referral centre for respiratory medicine, with one specialist admissions ward and 3 inpatient wards. Data from the integrated electronic observation and communication system comprising respiratory rate, oxygen saturations, heart rate, blood pressure, temperature, conscious level (AVPU score), and urine output were analysed. The same system also automatically generates mandated escalation and referral at set scoring thresholds via a pre-determined protocol. Scores from the local EWS were linked to demographics and mortality outcomes. NEWS criteria were applied retrospectively to determine how many patients would have been escalated if the NEWS system were followed. Results were analysed using STATA 15. The entire data set was analysed for measurement of escalation patterns, analysis of workload and sensitivity and specificity in predicting death within 24 hours of an observation [4]. Observations coded as end of life care following clinical decision were excluded from mortality analysis (see Figure 2).

Figure 2. Cohort flow diagram of exclusion criteria

Results

236,840 observation sets were recorded during 8812 inpatient episodes (53.1% female- see Table1) involving 6091 individuals. In-hospital mortality for respiratory patients was 5.9% (n=521) and median length of stay was 4 days (range 0-175).

Characteristics of the patients in this study	
Numbers (%)	
Male	3824 (47)
Female	4438 (53)
Total	8812
Mean age in years	
Male	63.7
Female	62.7
Total	63.1
Vital Signs (mean +/- SD)	
Heart Rate (beats per minute)	87 (16)
Respiratory rate (breaths per minute)	19 (3)
Systolic BP (mmHg)	130 (22)
Temperature (°C)	36.6 (1)
Oxygen saturations (%)	94 (6)

Table 1. Population characteristics of study cohort

59,434 (25.1%) observations sets were recorded between the hours of 0900-1700 Monday to Friday (excluding bank holidays). 177,406 (74.9%) were recorded outside of these hours. The local EWS and escalation protocol led to a median of 36 (range 1-148) scores per day that triggered a medical review (Table 3). This included a median of 5 (range 0-41) automated referrals to the resident on call senior clinician (medical registrar) every day.

If NEWS criteria were applied to the same population it would have generated a median of 98 (range 12-270) escalations to a doctor per day ($p<0.001$ for difference between scores), with 38 (range 2-158) scores generating automatic referral to the registrar ($p<0.001$ for difference between scores) per day.

Sensitivity and specificity for predicting in-hospital mortality based on death within 24 hours of a set of vital signs observations point is shown in Table 2. At each clinically equivalent band, the sensitivity and specificity in predicting mortality of all patients scoring at and above that cut point are shown. At each cut point, NEWS would have had a higher sensitivity than the local EWS (i.e. a higher percentage of patients who went on to die were flagged as requiring escalation), but a lower specificity.

Figure 3 plots sensitivity in predicting mortality, against median number of mandated clinician alerts per day for both EWS types. It demonstrates that for a sensitivity of 0.7, NEWS generates a higher number of mandated escalations. At both extremes of sensitivity (0 and 1) the number of escalations is the same: i.e. Mandating an escalation at a NEWS or EWS of 0 would mean all patients were escalated, and each score would have 100% sensitivity for predicating mortality (as everyone who died would have been reviewed). Likewise only escalating patients with a maximum EWS or NEWS score would lead to very few patients being escalated.

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

Discussion

In this study we examined the effect of two different EWS systems in patients admitted with respiratory disease. We analysed the number of mandatory escalations generated and the sensitivity and specificity of both scores in predicting imminent in-hospital mortality. Our data shows that at the scores' cut points for escalation, NEWS would have generated a significantly higher workload due to a lower specificity, with a higher sensitivity for predicting imminent deterioration, when compared with the locally used EWS.

Although previous work suggested that NEWS was worse at predicting deterioration in patients with respiratory disease, compared to a population of unselected medical admissions [2], NEWS has not been studied in large numbers of respiratory patients across an entire admission.

Our study faced similar limitations to others, namely the low prevalence of mortality in the patient population and the difficulty of studying patients in real time. However, our observed findings of an increase workload generated are both novel and important as, when used as part of a system which employs automatic escalation of threshold scores, NEWS leads to a significant impact on work load in a resource pressured environment, with little evidence of improved clinical outcome.

Although there is a difference in the workload generated by both scoring systems, this relates to the cut points for escalation mandated by the protocols, rather than the scores themselves; unsurprisingly overall both scores perform similarly (they are based on similar clinical observations) however the mandated cut points differ. The difference in protocol design relates to the way in which the scores are used clinically, and can be explained as follows:

The first approach is seen in the scoring thresholds dictated by NEWS. Its cut points for each layer of clinical intervention, i.e. escalation to nurse, clinician or registrar, have a higher sensitivity which acts to rule out imminent clinical deterioration in those patients whose vital signs do not meet scoring thresholds, meaning clinicians can be confident that patients with a low score are very unlikely to be

at imminent risk. This is akin to a d-dimer where a low value in an individual with low clinical suspicion effectively excludes a venous thromboembolism. This approach works well in a setting with less highly trained staff delivering the first layer of monitoring. However, if this approach is applied in an unfiltered manner, such as can exist with systems employing electronic monitoring with automatic mandatory escalation, the workload generated by escalations from patients who never go on to deteriorate has significant resource and operational implications, as well as increased likelihood of unnecessary intervention for patients. Again this is analogous to using a d-dimer which has been demonstrated to have a positive predictive value of approximately 20% in going on to find radiological evidence of a venous thromboembolism [5].

The second approach, used by the local EWS, is one of high specificity in the cut points for escalation, with a relatively lower sensitivity. This approach acts to highlight potential imminent clinical deterioration in those meeting the escalation criteria, but does not always rule out deterioration in those who score under the cut point. This may seem a less preferable approach, however a recent study of rapid response systems indicated that staff clinical concern in the absence of a qualifying score was responsible for escalation in 47% of calls [6], highlighting the role of staff education and empowerment, over EWS protocols. It may also mean that in resource limited environments (such as during out of hours care) if patients are highlighted as needing intervention, staff are more likely to be available to intervene as they are not being used to review patients who are triggering a review but clinically stable.

Despite the mandated and widespread uptake of EWS, there has been minimal prospective validation of their use. Efforts to improve precision in predicting outcome through scrutiny of large datasets has largely employed analyses utilising area under the receiver operating characteristic curves which is limited by the low prevalence of mortality in the population [7]. Before and after studies have largely, but not universally [8-10] highlighted the efficacy of EWS, however no randomised controlled trials have been performed. Consequently evidence of the scores' real impact on clinical outcomes, such as mortality, transfer to higher level of care or length of stay, or on workforce outcomes such as workload through excessive task generation and alarm fatigue, has only been obtained from observational studies. These are all limited by significant confounders.

This evidence gap around the clinical and workforce implications of EWS systems will become increasingly important as hospitals move towards automated systems with mandated referral of patients who reach a threshold score. Continuing integration of more data into digital healthcare systems via continuous monitoring, dynamic measures of fitness, and electronic health records will further highlight this gap, as without an understanding of how these data can be applied it will be

difficult to differentiate the signal from the noise. Given the growing complexity of the inpatient population more work is urgently required to understand the wider impact of EWS on outcomes such as mortality and length of stay, task burden, working patterns and cost. This is particularly important in patients with respiratory disease where physiology is often chronically deranged and less responsive to intervention.

NEWS band	Mandated escalation to:	% of observations in each band	Median number per day (range)	Sensitivity for predicting death within 24 hours	Specificity for predicting death within 24 hours
0	Nil	17.86	32 (3-75)	100.00	0.00
1 to 4	Nurse	67.34	180 (21-457)	99.44	15.09
5 to 6	Doctor	8.82	60 (10-184)	88.64	74.51
7 or more	Registrar	5.97	38 (2-158)	68.53	91.16
NUH EWS band	Mandates escalation to:	% of observations in each band	Median number per day (range)	Sensitivity for predicting death within 24 hours	Specificity for predicting death within 24 hours
0	Nil	56.11	174 (20-409)	100.00	0
1 to 2	Nurse	31.83	99 (16-300)	95.49	56.32
3	Nurse/ Doctor	5.39	16 (1-116)	76.65	88.24
4 to 5	Doctor	4.74	14 (1-55)	63.33	93.58
6 or more	Registrar	1.94	5 (0-41)	41.91	98.24

Table 2- Observations falling within each escalation band and sensitivity for each band relating to in hospital mortality

Collaborators Dr Mark Simmonds was instrumental in delivering the electronic observations system at Nottingham University Hospitals Trust and has liaised with our group in relation to our work on EWS

Author Contributorship Statement

Sarah Forster was first author and performed initial data analysis, literature review and created initial document and revisions.

Gemma Housley is an NHS data analyst who extracted the data for the article and provided editorial input.

Tricia Mckeever is a University of Nottingham Statistician who provided statistical advice and oversight with editorial input.

Dominick Shaw is the supervising author. In collaboration with the first author he developed the protocol, advising on data sources and research question while providing editorial input at all stages of the manuscript's development.

Competing Interests None declared

Data Sharing Statement

Unpublished metadata from the dataset analysed for this work is available through contacting the authors. However, due to the nature of the raw data it is not possible to make it freely available due to the limitations placed on the use of the dataset by NHS Information Governance procedures and approvals.

Funding Statement There is no funding to declare for this project

Corresponding Author Email

Dr Sarah Forster, sarahforster@nhs.net

References

1. Physicians, R.C.o., *National Early Warning Score (NEWS): Standardising the assessment of acute illness severity in the NHS.* 2012, Royal College of Physicians.
2. Hodgson, L.E., B.D. Dimitrov, J. Congleton, 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.
3. Eccles, S.R., C. Subbe, D. Hancock, and N. Thomson, *CREWS: improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia.* Resuscitation, 2014. **85**(1): p. 109-11.
4. Jarvis, S.W., C. Kovacs, J. Briggs, et al., *Are observation selection methods important when comparing early warning score performance?* Resuscitation, 2015. **90**: p. 1-6.
5. Owaidah, T., N. AlGhasham, S. AlGhamdi, et al., *Evaluation of the usefulness of a D dimer test in combination with clinical pretest probability score in the prediction and exclusion of Venous Thromboembolism by medical residents.* Thromb J, 2014. **12**(1): p. 28.
6. McGaughey, J., P. O'Halloran, S. Porter, J. Trinder, and B. Blackwood, *Early warning systems and rapid response to the deteriorating patient in hospital: A realist evaluation.* J Adv Nurs, 2017.
7. Romero-Brufau, S., J.M. Huddleston, G.J. Escobar, and M. Liebow, *Why the C-statistic is not informative to evaluate early warning scores and what metrics to use.* Crit Care, 2015. **19**: p. 285.
8. Subbe, C.P., R.G. Davies, E. Williams, P. Rutherford, and L. Gemmell, *Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions.* Anaesthesia, 2003. **58**(8): p. 797-802.
9. Booth, C., *Effect on outcome of an early warning score in acute medical admissions.* British Journal of Anaesthesia, 2003. **90**(4): p. 1.
10. Moon, A., J.F. Cosgrove, D. Lea, A. Fairs, and D.M. Cressey, *An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR.* Resuscitation, 2011. **82**(2): p. 150-4.

Weighted score	3	2	1	0	1	2	3
Respiratory rate	<8	-	9-11	12-20	-	21-24	>25
Oxygen saturation (NEWS only)	<91	92-93	94-95	>96	-	-	-
Supplemental oxygen	-	NEWS-Yes	-	NEWS-No Notts-0-9L/min	Notts-10-14L/min		Notts->15L/min
Heart Rate	<40	-	41-50	51-90	91-110	111-130	>131
Blood Pressure	<90	91-100	101-110	111-219	-	-	>220
Temperature	<35	-	35.1-36	36.1-38	38.1-39	>39.1	-
Urine output (Notts only)	-	Not PU'd in 12 hours or <20ml/hour	Not PU'd 6 hours or 20-30	PU'd in last 6 hours or 31-199ml/hour	>200ml/hr	-	-
AVPU	Both-Unresponsive	Notts- Pain	Notts-Voice	Both- Alert			

	NEWS		Nottingham Early Warning Score	
	Monitoring frequency	Clinical Response	Monitoring frequency	Clinical Response
0	Min 12 hourly	Continue routine NEWS monitoring	Min 12 hourly	
1	Min 4-6 hourly	Inform RN who must assess patient RN to decide re. increased monitoring or escalation	Min 4-6 hourly	
2				
3			Hourly for minimum 2 hours Consider fluid balance monitoring	RN to recheck Inform NiC Manual BP if deranged
4			Hourly with fluid balance monitoring	RN to check EWS F1/SHO to discuss management with SpR or consultant
5	Min 1 hourly	RN to urgently inform medical team Urgent assessment by clinician with core competencies Clinical care in environment with monitoring facilities		
6			Minimum ½ hourly with hourly fluid balance. Consider GCS	SpR Review If no improvement following intervention for discussion with consultant SpR to contact critical care for advice
>7	Continuous	RN to immediately inform medical team at minimum SpR Emergency assessment by clinical team with critical care competencies Consider transfer to level 2/3 care		

Figure 1- Escalation protocol for NEWS and Nottingham University Hospitals Early Warning Score

121x155mm (96 x 96 DPI)

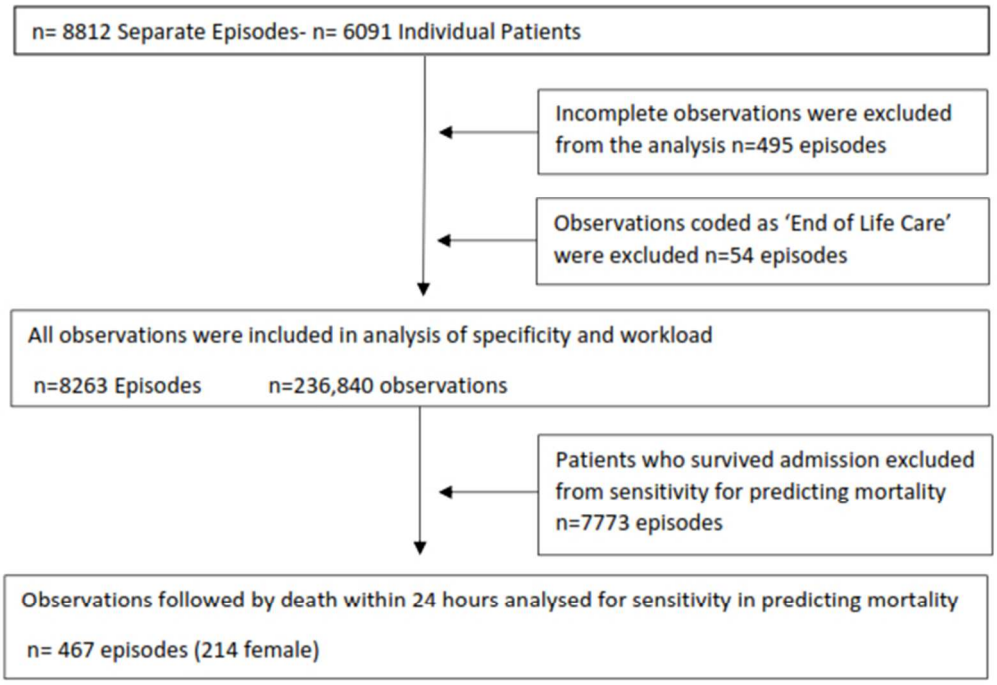

Figure 2. Cohort flow diagram of exclusion criteria

158x112mm (96 x 96 DPI)

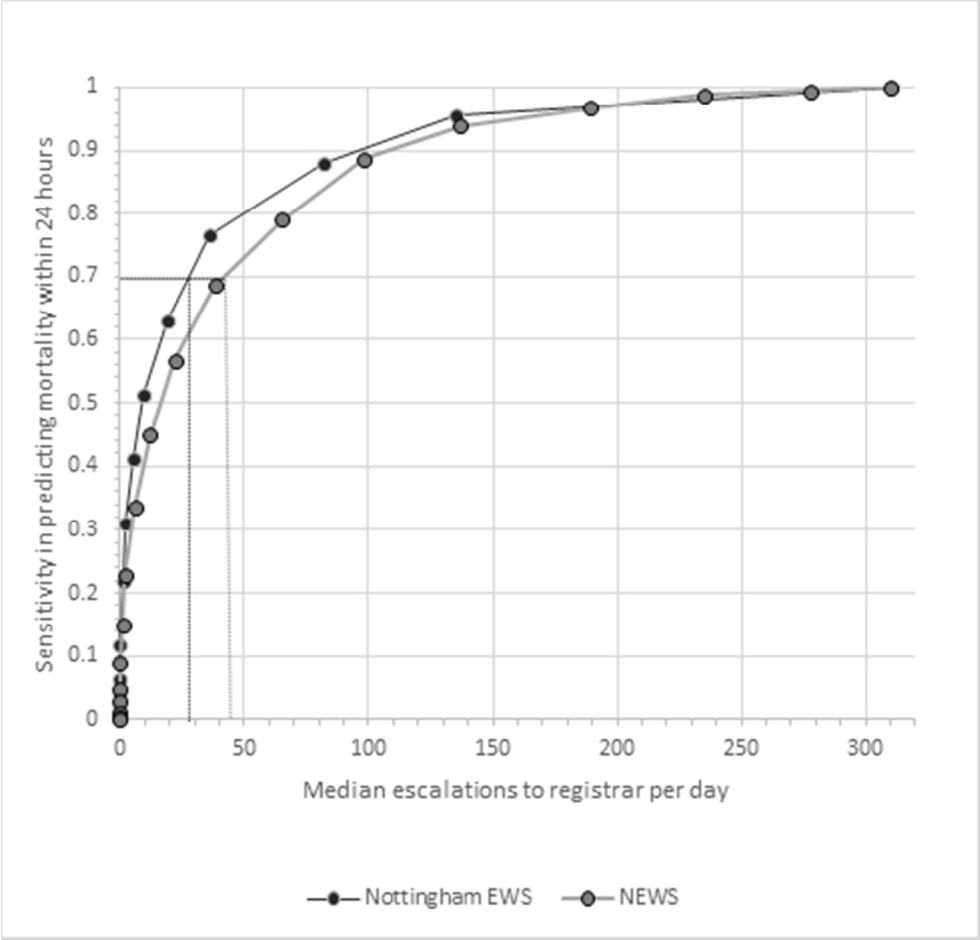

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

129x124mm (96 x 96 DPI)

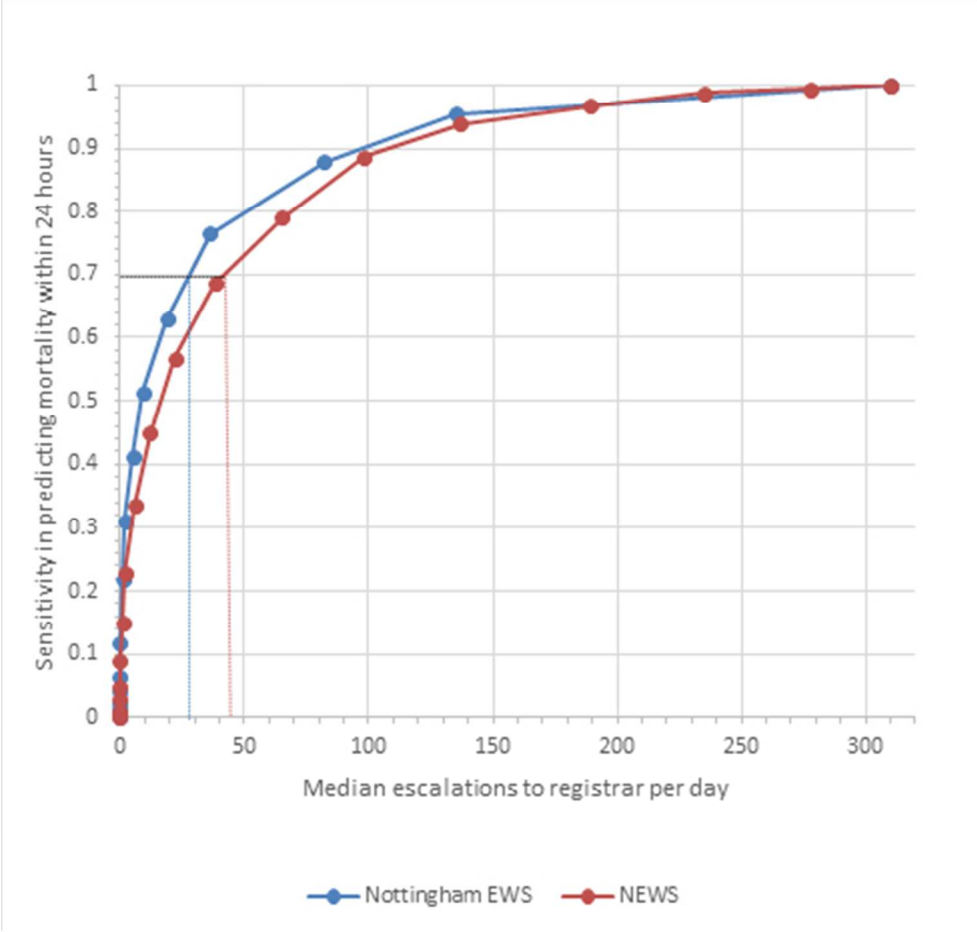

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

129x124mm (96 x 96 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **[See design section of abstract pages 1-2]**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **[See abstract pages 1-2]**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **[See abstract and Background sections- pages 1-2]**
Objectives	3	State specific objectives, including any prespecified hypotheses **[See objective section in abstract]**
Methods		
Study design	4	Present key elements of study design early in the paper **[methods in abstract – expanded in methods section of main article on page 3]**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **[Methods and Setting in abstract, Methods page 3]**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **[Methods section page 3]**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **[N/A]**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **[Background, Methods, Discussion]**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **[Methods page 3]**
Bias	9	Describe any efforts to address potential sources of bias **[N/A]**
Study size	10	Explain how the study size was arrived at **[Methods page 3]**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **[Methods page 3]**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **[Methods page 3]**
		(*b*) Describe any methods used to examine subgroups and interactions **[Methods page 3]**
		(*c*) Explain how missing data were addressed **[Cohort Diagram- Figure 2, Methods page 3]**
		(*d*) If applicable, explain how loss to follow-up was addressed **[n/a]**
		(*e*) Describe any sensitivity analyses **[Methods page 3]**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **[Results page 4]**
		(b) Give reasons for non-participation at each stage **[N/A]**
		(c) Consider use of a flow diagram **[Cohort flow diagram in methods section]**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **[Table 1- characteristics of**

		study population]
		(b) Indicate number of participants with missing data for each variable of interest **[See cohort flow diagram- figure 2 in methods page 3]**
		(c) Summarise follow-up time (eg, average and total amount) **[N/A]**
Outcome data	15*	Report numbers of outcome events or summary measures over time **[Abstract and results]**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **[N/A]**
		(*b*) Report category boundaries when continuous variables were categorized **[See figure 1 + Table 2]**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **[n/a]**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **[Methods and results page 3-4]**
Discussion		
Key results	18	Summarise key results with reference to study objectives **[Discussion page 5 onwards and abstract]**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **[Strengths and weaknesses page 1; Discussion page 5 onwards]**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **[Background page 2 and Discussion page 5 onwards]**
Generalisability	21	Discuss the generalisability (external validity) of the study results **[Discussion page 5]**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **[Included after main body of text before references]**

*Give information separately for exposed and unexposed groups.

BMJ Open

Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to a tertiary respiratory referrals centre over a 2 year period.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020269.R1
Article Type:	Research
Date Submitted by the Author:	25-Jan-2018
Complete List of Authors:	Forster, Sarah; University of Nottingham School of Medicine, Respiratory Medicine Housley, Gemma; Nottingham University Hospitals NHS Trust McKeever, Tricia; University of Nottingham, Division of Epidemiology and Public Health Shaw, Dominick; University of Nottingham, Respiratory Medicine
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Evidence based practice
Keywords:	Thoracic medicine < INTERNAL MEDICINE, Adult intensive & critical care < INTENSIVE & CRITICAL CARE, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to a tertiary respiratory referrals centre over a 2 year period.

Sarah Forster[1,2], Gemma Housley[3], Tricia M McKeever[4], Dominick E Shaw[2,3]

Author Affiliations

[1] NIHR Academic Clinical Fellow, University of Nottingham, Nottingham, UK

[2] Respiratory Research Unit, Division of Respiratory Medicine, University of Nottingham, Nottingham, UK

[3] Medical Informatics, East Midlands Academic Health Sciences Network, Nottingham, UK

[4] Division of Epidemiology, University of Nottingham, Nottingham, UK

Correspondence to Dr Sarah Forster, sarahforster@nhs.net

Word count: 2976

Abstract

Objective

Early Warning Scores (EWS) are used to monitor patients for signs of imminent deterioration. Although used in respiratory disease, EWS have not been well studied in this population, despite the underlying cardiopulmonary pathophysiology often present. We examined the performance of two scoring systems in patients with respiratory disease.

Design

Retrospective cohort analysis of vital signs observations of all patients admitted to a respiratory unit over a 2 year period. Scores were linked to outcome data to establish the performance of the National EWS (NEWS) compared results to a locally adapted EWS.

Setting

Nottingham University Hospitals NHS Trust respiratory wards. Data were collected from an integrated electronic observation and task allocation system employing a local EWS, also generating mandatory referrals to clinical staff at set scoring thresholds.

Outcome Measures

Projected workload, and sensitivity and specificity of the scores in predicting mortality based on outcome within 24 hours of a score being recorded.

Results

8812 individual patient episodes occurred during the study period. Overall mortality was 5.9%. Applying NEWS retrospectively (versus local EWS) generated an eight fold increase in mandatory escalations, but had higher sensitivity in predicting mortality at the protocol cut points.

Conclusions

This study highlights issues surrounding use of scoring systems in patients with respiratory disease. NEWS demonstrated higher sensitivity for predicting death within 24 hours, offset by reduced specificity. The consequent workload generated may compromise the ability of the clinical team to respond to patients needing immediate input. The locally adapted EWS has higher specificity but lower sensitivity. Statistical evaluation suggests this may lead to missed opportunities for intervention, however this does not account for clinical concern independent of the scores, nor ability to respond to alerts based on workload. Further research into the role of warning scores and the impact of chronic pathophysiology is urgently needed.

Strengths and Limitations of this Study

- Data were obtained from a large clinical vital signs database with clear identification of specialty allowing for subgroup analysis. All observations were included in the analysis, regardless of whether there had previously been a high score which may have resulted in a change of management by the clinical team
- Granularity of data collection in the database allowed for reliable identification of patients meeting the exclusion criteria. Only 0.2% of the observations recorded during the study period were identified as being incomplete.
- The retrospective nature of study precludes conclusions relating to impact of introducing NEWS on mortality.
- DNACPR decisions were not linked as part of the analysis.
- Inherent inaccuracy in recording time of death in hospital records means 24 hour cut off may not be always be exact

Background

Early Warning Scores combine vital sign measures into a composite score in order to identify patients at risk of clinical deterioration, guide early intervention and reduce avoidable mortality. Scores have evolved over the last 30 years following the recognition that patients experiencing a serious adverse event, such as unplanned transfer to intensive care, in hospital cardiac arrest or death, showed evidence of pathophysiology in their vital signs observations in the hours leading up to overt deterioration. Initially this information was captured in the form of single parameter scores where significant derangement in a single vital sign or clinical concern triggered a set clinical response. In the UK this led to the development of aggregate weighted scores, whereby each vital sign is given a weighting depending on how far outside the predetermined normal range it falls; the sum of these scores is then used to guide response.

In 2012 the Royal College of Physicians published the NEWS protocol in an attempt to standardise processes for identifying patients at risk of imminent deterioration [1]. EWS protocols guide decisions around patient care by mandating when a patient with evidence of pathophysiology, in the form of deranged vital signs, should be reviewed by a clinical member of staff, and therefore influence overall clinical workload and resource allocation for all in-patients. Patients with respiratory disease make up a large proportion of a hospital's in-patient population, however it is recognised that chronic physiological disturbance caused by COPD may render NEWS less

discriminative when compared to an unselected medical population [2]. This has significant implications for patients, in terms of increased observations and interventions, and to clinical staff in terms of workload and potential for alert fatigue. Consequently attempts have been made to improve the score in this population [3].

Nottingham University Hospitals NHS Trust employs an electronic observations system with mandatory escalation based on an adapted EWS. The Nottingham EWS, unlike NEWS, does not score oxygen saturations and has a graduated approach to weighting for both oxygen delivery and level of consciousness. As a more general marker of morbidity it also employs urine output. We compared the sensitivity and specificity of the two scores in predicting mortality within 24 hours of a set of observations being recorded at the clinical cut points determined by the associated protocols and examined the potential impact in terms of workload of using the locally designed EWS versus NEWS (see Figure 1) in patients with respiratory disease based on analysis of the vital signs observations and outcomes of patients admitted to the respiratory department in Nottingham over a 2 year period. We then went on to answer the same questions in a subgroup of patients who were admitted with a diagnosis of COPD to examine the performance of the two scores in this cohort.

Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham University Hospitals Early Warning Score

Methods

We performed a single centre retrospective analysis of all patients admitted to the respiratory department at Nottingham University Hospitals NHS Trust between 01/04/2015 and 31/03/2017. This is a tertiary referral centre for respiratory medicine, with one specialist admissions ward and 3 inpatient wards. The analysis included all adults admitted with respiratory disease not transferred to a higher level of care, i.e. high dependency or intensive care, greater than 24 hours before death as these areas are not currently employing electronic observations, long term ventilator dependent patients were also excluded as hospital policy dictates that these patients are always admitted to the high dependency unit. Following approval from the NHS Information Governance Lead, and in line with existing permissions within the East Midlands Academic Health Sciences network, data from the integrated electronic observation and communication system comprising respiratory rate, oxygen saturations, heart rate, blood pressure, temperature, conscious level (AVPU score), and urine output were anonymised by an NHS data analyst prior to extraction from the clinical server. The same system also automatically generates mandated escalation and referral at set scoring thresholds via a pre-determined protocol. Scores from the local EWS were linked to demographics and mortality

outcomes prior to extraction. NEWS criteria were applied retrospectively to determine how many patients would have been escalated if the NEWS system were followed. Results were analysed using STATA 15. The entire data set was analysed for measurement of escalation patterns, analysis of workload and sensitivity and specificity in predicting death within 24 hours of an observation [4]. A chi-square analysis was performed to demonstrate whether the difference in escalations was significant. The statistical analysis involved the use of all vital signs observations recorded throughout admission, which were linked to outcome to determine whether they were followed by death within 24 hours of the observation timestamp created by the input devices at the bedside. Observations coded as end of life care following clinical decision were excluded from mortality analysis (see Figure 2). A further subgroup analysis was then performed on patients coded as having chronic obstructive pulmonary disease at any point in their admission as per ICD10 codes (see Figure 3) in order to further assess the statistical performance of the two scores in the presence of chronic pathophysiology.

Figure 2. Cohort flow diagram of exclusion criteria

Results

236,840 observation sets were recorded during 8812 inpatient episodes (53.1% female- see Table1) involving 6091 individuals. In-hospital mortality for respiratory patients was 5.9% (n=521) and median length of stay was 4 days (range 0-175).

Characteristics of the patients in this study	
	Numbers (%)
Male	3824 (47)
Female	4438 (53)
Total	8812
Mean age in years	
Male	63.7
Female	62.7
Total	63.1
Vital Signs	Mean (+/- SD)
Heart Rate (beats per minute)	87 (16)
Respiratory rate (breaths per minute)	19 (3)
Systolic BP (mmHg)	130 (22)
Temperature (°C)	36.6 (1)
Oxygen saturations (%)	94 (6)

Table 1. Population characteristics of study cohort

59,434 (25.1%) observations sets were recorded between the hours of 0900-1700 Monday to Friday (excluding bank holidays). 177,406 (74.9%) were recorded outside of these hours. The local EWS and escalation protocol led to a median of 36 (range 1-148, calculated from the raw data of scores between 3 and 5 each day) scores per day that triggered a medical review (Table 2). This included a median of 5 (range 0-41) automated referrals to the resident on call senior clinician (medical registrar) every day. Direct comparison of workload generated to other members of the clinical team was not possible as the escalation protocol for both scores is only directly comparable at registrar level, however the workload generated at each of the clinically applied cut points can be seen in tables 2 and 3.

If NEWS criteria were applied to the same population it would have generated a median of 98 (range 12-270) escalations to a doctor per day ($p<0.001$ for difference between scores), with 38 (range 2-

158) scores generating automatic referral to the registrar (p<0.001 for difference between scores) per day.

Sensitivity and specificity for predicting in-hospital mortality based on death within 24 hours of a set of vital signs observations point is shown in Table 2. At each clinically equivalent band, the sensitivity and specificity in predicting mortality of all patients scoring at and above that cut point are shown. At each cut point, NEWS would have had a higher sensitivity than the local EWS (i.e. a higher percentage of patients who went on to die were flagged as requiring escalation), but a lower specificity.

Figure 3 plots sensitivity in predicting mortality, against median number of mandated clinician alerts per day for both EWS types. It demonstrates that for a sensitivity of 0.7, NEWS generates a higher number of mandated escalations. At both extremes of sensitivity (0 and 1) the number of escalations is the same: i.e. mandating an escalation at a NEWS or EWS of 0 would mean all patients were escalated, and each score would have 100% sensitivity for predicting mortality (as everyone who died would have been reviewed). Likewise only escalating patients with a maximum EWS or NEWS score would lead to very few patients being escalated.

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

Further subgroup analysis was performed on admissions with an ICD10 code for COPD at any point. This yielded 56,345 observations from 2207 episodes by 1365 individual patients. Using the local EWS protocol led to median of 0 (range 0-19) escalations to the registrar, while applying NEWS would have generated a median of 6 (0-47) scores being escalated to the registrar each day. As in the unselected respiratory cohort, NEWS was more sensitive in predicting imminent mortality than the local EWS but with a significantly inferior specificity at each clinical cut point applied (see table 3).

Discussion

In this study we examined the effect of two different early warning score systems in patients admitted with respiratory disease. We analysed the number of mandatory escalations generated and the sensitivity and specificity of both scores in predicting imminent in-hospital mortality in an unselected respiratory population and in a subgroup analysis of patients with COPD. Our data shows

that at the scores' cut points for escalation, NEWS would have generated a significantly higher workload due to a lower specificity, with a higher sensitivity for predicting imminent deterioration, when compared with the locally used EWS. This was accentuated in patients with COPD, an observation we believe is due to chronic changes in the underlying physiology which influences the way in which these patients respond to acute pathological processes.

Although previous work suggested that NEWS was less discriminative in predicting deterioration in patients with respiratory disease, compared to a population of unselected medical admissions [2], NEWS has not been studied in large numbers of respiratory patients across an entire admission.

Our study faced similar limitations to others, namely the low prevalence of mortality in the patient population and the difficulty of studying patients in real time. However, our observed findings of an increase workload generated are both novel and important as, when used as part of a system which employs automatic escalation of threshold scores, NEWS leads to a significant impact on work load in a resource pressured environment, with little evidence of improved clinical outcome.

Although there is a difference in the workload generated when comparing the scoring systems both in a general respiratory population and in patients with COPD, this relates to the cut points for escalation mandated by the protocols, rather than the scores themselves; unsurprisingly overall both scores perform similarly when the individual scores are plotted (they are based on similar clinical observations) however the mandated cut points differ. The difference created by the protocol design relates to the way in which the scores are used clinically, and can be explained as follows:

The first approach is seen in the scoring thresholds dictated by NEWS. Its cut points for each layer of clinical intervention, i.e. escalation to nurse, clinician or registrar, have a higher sensitivity which acts to rule out imminent clinical deterioration in those patients whose vital signs do not meet scoring thresholds, meaning clinicians can be confident that patients with a low score are very unlikely to be at imminent risk. This is akin to a d-dimer where a low value in an individual with low clinical suspicion effectively excludes a venous thromboembolism [5, 6] . This high sensitivity approach works well in a setting with less highly trained staff delivering the first layer of monitoring. However, if this approach is applied in an unfiltered and automated manner, the workload generated by escalations from patients who never go on to deteriorate will have significant resource and operational implications, as well as increasing the likelihood of unnecessary intervention for patients.

The second approach, used by the local EWS, is one of high specificity in the cut points for escalation, with a relatively lower sensitivity. This approach acts to highlight potential imminent clinical deterioration in those meeting the escalation criteria, but does not always rule out deterioration in those who score under the cut point. This may seem a less preferable approach. However a recent study of rapid response systems indicated that staff clinical concern in the absence of a qualifying score was responsible for escalation in 47% of calls [7], highlighting the role of staff education and empowerment, over and above EWS protocols The variability in physiological normal baselines created by patient specific factors such as comorbidity or fitness means that using vital signs observations alone as the basis for a score leading to mandatory escalation will always require a trade-off between sensitivity in accurately identifying patients potentially at risk of deterioration and staff alarm fatigue generated by patients who do not go onto deteriorate. This is particularly pertinent in resource limited environments (such as during out of hours care),

Despite the mandated and widespread uptake of EWS, there has been minimal prospective validation of their use. Efforts to improve precision in predicting outcome through scrutiny of large datasets has largely employed analyses utilising area under the receiver operating characteristic curves which are limited by the low prevalence of mortality in the population [8]. Before and after studies have largely, but not universally [9-11] highlighted the efficacy of EWS, however no randomised controlled trials have been performed. Consequently evidence of the scores' real impact on clinical outcomes, such as mortality, transfer to higher level of care or length of stay, or on workforce outcomes such as workload from excessive task generation and alarm fatigue, has only been obtained from observational studies. These are all limited by significant confounders.

This evidence gap around the clinical and workforce implications of EWS systems will become increasingly important as hospitals move towards automated systems with mandated referral of patients who reach a threshold score. Continuing integration of more data into digital healthcare systems via continuous monitoring, dynamic measures of fitness, and electronic health records will further highlight this gap, as without an understanding of how these data can be applied it will be difficult to differentiate the signal from the noise. Given the growing complexity of the inpatient population more work is urgently required to understand the wider impact of EWS on outcomes such as mortality and length of stay, task burden, working patterns and cost. There is also need to reconsider the role of clinical concern in monitoring patients and how this can be further promoted to prevent future systems depending purely on scores rather than integrating staff skills and intuition into the decision making process. Early warning scores should not be developed in isolation based on statistical performance as this fails to recognise that they are a component within the complex clinical environment and therefore need to be designed to enhance, not complicate, the

clinical decision making process. This is particularly important in patients with respiratory disease where physiology is often chronically deranged and less responsive to intervention and a greater understanding of the contributory clinical factors and more individualised approach is required.

NEWS band	Mandated escalation to:	% of observations in each band	Median number per day (range)	Sensitivity for predicting death within 24 hours	Specificity for predicting death within 24 hours
0	Nil	17.86	32 (3-75)	100.00	0.00
1 to 4	Nurse	67.34	180 (21-457)	99.44	15.09
5 to 6	Doctor	8.82	60 (10-184)	88.64	74.51
7 or more	Registrar	5.97	38 (2-158)	68.53	91.16
NUH EWS band	Mandates escalation to:	% of observations in each band	Median number per day (range)	Sensitivity for predicting death within 24 hours	Specificity for predicting death within 24 hours
0	Nil	56.11	174 (20-409)	100.00	0
1 to 2	Nurse	31.83	99 (16-300)	95.49	56.32
3	Nurse/ Doctor	5.39	16 (1-116)	76.65	88.24
4 to 5	Doctor	4.74	14 (1-55)	63.33	93.58
6 or more	Registrar	1.94	5 (0-41)	41.91	98.24

Table 2- Workload predictions and sensitivity and specificity in predicting death within 24 hours for NEWS and local EWS for unselected respiratory population

NEWS band	Mandated Escalation	% of observations in each band	Median number (range)	Sensitivity for death within 24 hours	Specificity for death within 24 hours
0	Nil	7.96	5 (0-23)	100.00	0.00
1 to 4	Nurse	59.3	43 (4-112)	100.00	7.99
5 to 6	Doctor	22.2	16 (1-59)	89.85	67.47
7 or more	Registrar	10.54	6 (0-47)	71.07	89.68
NUH EWS band	Mandated Escalation	% of observations in each band	Median number (range)	Sensitivity for death within 24 hours	Specificity for death within 24 hours

0	Nil	53.89	39 (1-101)	100.00	0.00
1 to 2	Nurse	35.05	26 (4-90)	92.39	54.06
3	Nurse/Doctor	5.46	3 (0-30)	70.56	88.50
4 to 5	Doctor	4.31	2 (0-18)	58.38	94.20
6 or more	Registrar	1.28	0 (0-19)	38.07	98.75

Table 3- Workload predictions and sensitivity and specificity in predicting death within 24 hours for NEWS and local EWS for patients with COPD

Collaborators

Dr Mark Simmonds was instrumental in delivering the electronic observations system at Nottingham University Hospitals Trust and has liaised with our group in relation to our work on EWS

Author Contributorship Statement

Sarah Forster was first author and performed initial data analysis, literature review and created initial document and revisions.

Gemma Housley is an NHS data analyst who extracted the data for the article and provided editorial input.

Tricia Mckeever is a University of Nottingham Statistician who provided statistical advice and oversight with editorial input.

Dominick Shaw is the supervising author. In collaboration with the first author he developed the protocol, advising on data sources and research question while providing editorial input at all stages of the manuscript's development.

Competing Interests

None declared

Data Sharing Statement

Unpublished metadata from the dataset analysed for this work is available through contacting the authors. However, due to the nature of the raw data it is not possible to make it freely available due to the limitations placed on

the use of the dataset by NHS Information Governance procedures and approvals.

Funding Statement There is no funding to declare for this project

Corresponding Author Email

Dr Sarah Forster, sarahforster@nhs.net

References

1. Physicians, R.C.o., *National Early Warning Score (NEWS): Standardising the assessment of acute illness severity in the NHS.* 2012, Royal College of Physicians.
2. Hodgson, L.E., B.D. Dimitrov, J. Congleton, 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.
3. Eccles, S.R., C. Subbe, D. Hancock, and N. Thomson, *CREWS: improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia.* Resuscitation, 2014. **85**(1): p. 109-11.
4. Jarvis, S.W., C. Kovacs, J. Briggs, et al., *Are observation selection methods important when comparing early warning score performance?* Resuscitation, 2015. **90**: p. 1-6.
5. Stein, P.D., R.D. Hull, K.C. Patel, et al., *D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism: a systematic review.* Ann Intern Med, 2004. **140**(8): p. 589-602.
6. Owaidah, T., N. AlGhasham, S. AlGhamdi, et al., *Evaluation of the usefulness of a D dimer test in combination with clinical pretest probability score in the prediction and exclusion of Venous Thromboembolism by medical residents.* Thromb J, 2014. **12**(1): p. 28.
7. McGaughey, J., P. O'Halloran, S. Porter, J. Trinder, and B. Blackwood, *Early warning systems and rapid response to the deteriorating patient in hospital: A realist evaluation.* J Adv Nurs, 2017.
8. Romero-Brufau, S., J.M. Huddleston, G.J. Escobar, and M. Liebow, *Why the C-statistic is not informative to evaluate early warning scores and what metrics to use.* Crit Care, 2015. **19**: p. 285.
9. Subbe, C.P., R.G. Davies, E. Williams, P. Rutherford, and L. Gemmell, *Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions.* Anaesthesia, 2003. **58**(8): p. 797-802.
10. Booth, C., *Effect on outcome of an early warning score in acute medical admissions.* British Journal of Anaesthesia, 2003. **90**(4): p. 1.
11. Moon, A., J.F. Cosgrove, D. Lea, A. Fairs, and D.M. Cressey, *An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR.* Resuscitation, 2011. **82**(2): p. 150-4.

	NEWS		Nottingham Early Warning Score	
	Monitoring frequency	Clinical Response	Monitoring frequency	Clinical Response
0	Min 12 hourly	Continue routine NEWS monitoring	Min 12 hourly	
1	Min 4-6 hourly	Inform RN who must assess patient RN to decide re. increased monitoring or escalation	Min 4-6 hourly	
2				
3			Hourly for minimum 2 hours Consider fluid balance monitoring	RN to recheck Inform NIC Manual BP if deranged
4			Hourly with fluid balance monitoring	RN to check EWS F1/SHO to discuss management with SpR or consultant
5	Min 1 hourly	RN to urgently inform medical team Urgent assessment by clinician with core competencies Clinical care in environment with monitoring facilities		
6			Minimum ½ hourly with hourly fluid balance. Consider GCS	SpR Review If no improvement following intervention for discussion with consultant SpR to contact critical care for advice
>7	Continuous	RN to immediately inform medical team at minimum SpR Emergency assessment by clinical team with critical care competencies Consider transfer to level 2/3 care		

Weighted score	3	2	1	0	1	2	3
Respiratory rate	<8	-	9-11	12-20	-	21-24	>25
Oxygen saturation (NEWS only)	<91	92-93	94-95	>96	-	-	-
Supplemental oxygen	-	**NEWS-**Yes	-	**NEWS-**No **Notts-** 0-9L/min	**Notts-** 10-14L/min		**Notts-** ≥15L/min
Heart Rate	<40	-	41-50	51-90	91-110	111-130	≥131
Blood Pressure	<90	91-100	101-110	111-219	-	-	≥220
Temperature	<35	-	35.1-36	36.1-38	38.1-39	>39.1	-
Urine output (Notts only)	-	Not PU'd in 12 hours or <20 ml/hour	Not PU'd 6 hours or 20-30	PU'd in last 6 hours or 31-199ml/hour	>200ml/hr	-	-
AVPU	**Both-** Unresponsive	**Notts-** Pain	**Notts-** Voice	**Both-** Alert			

Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham University Hospitals Early Warning Score

105x143mm (300 x 300 DPI)

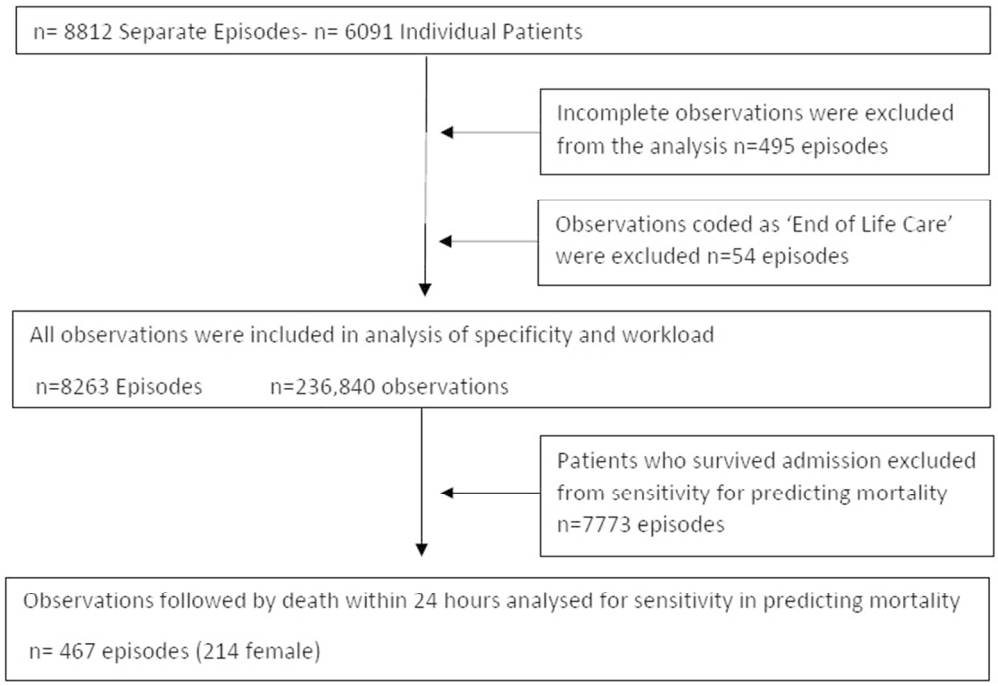

Figure 2. Cohort flow diagram of exclusion criteria

72x49mm (300 x 300 DPI)

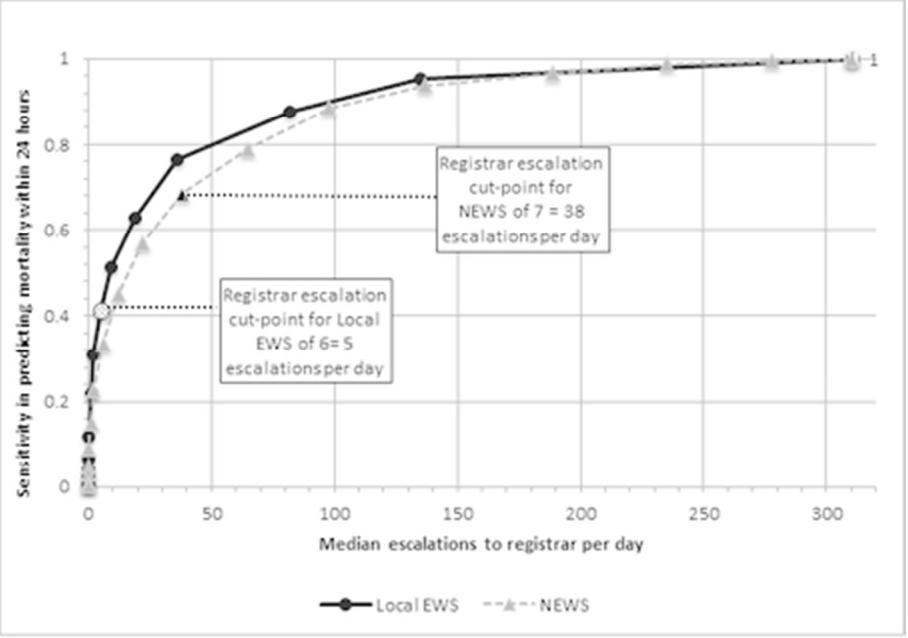

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

38x29mm (300 x 300 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **[See design section of abstract pages 1-2]**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **[See abstract pages 1-2]**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **[See abstract and Background sections- pages 1-2]**
Objectives	3	State specific objectives, including any prespecified hypotheses **[See objective section in abstract]**
Methods		
Study design	4	Present key elements of study design early in the paper **[methods in abstract – expanded in methods section of main article on page 3]**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **[Methods and Setting in abstract, Methods page 3]**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **[Methods section page 3]**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **[N/A]**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **[Background, Methods, Discussion]**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **[Methods page 3]**
Bias	9	Describe any efforts to address potential sources of bias **[N/A]**
Study size	10	Explain how the study size was arrived at **[Methods page 3]**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **[Methods page 3]**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **[Methods page 3]**
		(*b*) Describe any methods used to examine subgroups and interactions **[Methods page 3]**
		(*c*) Explain how missing data were addressed **[Cohort Diagram- Figure 2, Methods page 3]**
		(*d*) If applicable, explain how loss to follow-up was addressed **[n/a]**
		(*e*) Describe any sensitivity analyses **[Methods page 3]**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **[Results page 4]**
		(b) Give reasons for non-participation at each stage **[N/A]**
		(c) Consider use of a flow diagram **[Cohort flow diagram in methods section]**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **[Table 1- characteristics of**

		study population]
		(b) Indicate number of participants with missing data for each variable of interest **[See cohort flow diagram- figure 2 in methods page 3]**
		(c) Summarise follow-up time (eg, average and total amount) **[N/A]**
Outcome data	15*	Report numbers of outcome events or summary measures over time **[Abstract and results]**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **[N/A]**
		(*b*) Report category boundaries when continuous variables were categorized **[See figure 1 + Table 2]**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **[n/a]**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **[Methods and results page 3-4]**
Discussion		
Key results	18	Summarise key results with reference to study objectives **[Discussion page 5 onwards and abstract]**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **[Strengths and weaknesses page 1; Discussion page 5 onwards]**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **[Background page 2 and Discussion page 5 onwards]**
Generalisability	21	Discuss the generalisability (external validity) of the study results **[Discussion page 5]**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **[Included after main body of text before references]**

*Give information separately for exposed and unexposed groups.

BMJ Open

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.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020269.R2
Article Type:	Research
Date Submitted by the Author:	19-Mar-2018
Complete List of Authors:	Forster, Sarah; University of Nottingham School of Medicine, Respiratory Medicine Housley, Gemma; Nottingham University Hospitals NHS Trust McKeever, Tricia; University of Nottingham, Division of Epidemiology and Public Health Shaw, Dominick; University of Nottingham, Respiratory Medicine
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Evidence based practice
Keywords:	Thoracic medicine < INTERNAL MEDICINE, Adult intensive & critical care < INTENSIVE & CRITICAL CARE, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

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.

Sarah Forster[1,2], Gemma Housley[3], Tricia M McKeever[4], Dominick E Shaw[2,3]

Author Affiliations

[1] NIHR Academic Clinical Fellow, University of Nottingham, Nottingham, UK

[2] Respiratory Research Unit, Division of Respiratory Medicine, University of Nottingham, Nottingham, UK

[3] Medical Informatics, East Midlands Academic Health Sciences Network, Nottingham, UK

[4] Division of Epidemiology, University of Nottingham, Nottingham, UK

Correspondence to Dr Sarah Forster, sarahforster@nhs.net

Word count: 3217

Abstract

Objective

Early Warning Scores (EWS) are used to monitor patients for signs of imminent deterioration. Although used in respiratory disease, EWS have not been well studied in this population, despite the underlying cardiopulmonary pathophysiology often present. We examined the performance of two scoring systems in patients with respiratory disease.

Design

Retrospective cohort analysis of vital signs observations of all patients admitted to a respiratory unit over a 2 year period. Scores were linked to outcome data to establish the performance of the National EWS (NEWS) compared results to a locally adapted EWS.

Setting

Nottingham University Hospitals NHS Trust respiratory wards. Data were collected from an integrated electronic observation and task allocation system employing a local EWS, also generating mandatory referrals to clinical staff at set scoring thresholds.

Outcome Measures

Projected workload, and sensitivity and specificity of the scores in predicting mortality based on outcome within 24 hours of a score being recorded.

Results

8812 individual patient episodes occurred during the study period. Overall mortality was 5.9%. Applying NEWS retrospectively (versus local EWS) generated an eight fold increase in mandatory escalations, but had higher sensitivity in predicting mortality at the protocol cut points.

Conclusions

This study highlights issues surrounding use of scoring systems in patients with respiratory disease. NEWS demonstrated higher sensitivity for predicting death within 24 hours, offset by reduced specificity. The consequent workload generated may compromise the ability of the clinical team to respond to patients needing immediate input. The locally adapted EWS has higher specificity but lower sensitivity. Statistical evaluation suggests this may lead to missed opportunities for intervention, however this does not account for clinical concern independent of the scores, nor ability to respond to alerts based on workload. Further research into the role of warning scores and the impact of chronic pathophysiology is urgently needed.

Strengths and Limitations of this Study

- Data were obtained from a large clinical vital signs database with clear identification of specialty allowing for subgroup analysis. All observations were included in the analysis, regardless of whether there had previously been a high score which may have resulted in a change of management by the clinical team
- Granularity of data collection in the database allowed for reliable identification of patients meeting the exclusion criteria. Only 0.2% of the observations recorded during the study period were identified as being incomplete.
- The retrospective nature of study precludes conclusions relating to impact of introducing NEWS on mortality.
- DNACPR decisions were not linked as part of the analysis.
- Inherent inaccuracy in recording time of death in hospital records means 24 hour cut off may not be always be exact

Background

Early Warning Scores combine vital sign measures into a composite score in order to identify patients at risk of clinical deterioration, guide early intervention and reduce avoidable mortality. Scores have evolved over the last 30 years following the recognition that patients experiencing a serious adverse event, such as unplanned transfer to intensive care, in hospital cardiac arrest or death, showed evidence of pathophysiology in their vital signs observations in the hours leading up to overt deterioration. Initially this information was captured in the form of single parameter scores where significant derangement in a single vital sign or clinical concern triggered a set clinical response. In the UK this led to the development of aggregate weighted scores, whereby each vital sign is given a weighting depending on how far outside the predetermined normal range it falls; the sum of these scores is then used to guide response.

In 2012 the Royal College of Physicians published the NEWS protocol in an attempt to standardise processes for identifying patients at risk of imminent deterioration [1]. EWS protocols guide decisions around patient care by mandating when a patient with evidence of pathophysiology, in the form of deranged vital signs, should be reviewed by a clinical member of staff, and therefore influence overall clinical workload and resource allocation for all in-patients. Patients with respiratory disease make up a large proportion of a hospital's in-patient population, however it is recognised that chronic physiological disturbance caused by COPD may render NEWS less

discriminative when compared to an unselected medical population [2]. This has significant implications for patients, in terms of increased observations and interventions, and to clinical staff in terms of workload and potential for alert fatigue. Consequently attempts have been made to improve the score in this population [3].

Nottingham University Hospitals NHS Trust employs an electronic observations system with mandatory escalation based on an adapted EWS. The Nottingham EWS, unlike NEWS, does not score oxygen saturations and has a graduated approach to weighting for both oxygen delivery and level of consciousness. As a more general marker of morbidity it also employs urine output. We compared the sensitivity and specificity of the two scores in predicting mortality within 24 hours of a set of observations being recorded at the clinical cut points determined by the associated protocols and examined the potential impact in terms of workload of using the locally designed EWS versus NEWS (see Figure 1) in patients with respiratory disease based on analysis of the vital signs observations and outcomes of patients admitted to the respiratory department in Nottingham over a 2 year period. We then went on to answer the same questions in a subgroup of patients who were admitted with a diagnosis of COPD to examine the performance of the two scores in this cohort.

Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham University Hospitals Early Warning Score

Methods

We performed a single centre retrospective analysis of all patients admitted to the respiratory department at Nottingham University Hospitals NHS Trust between 01/04/2015 and 31/03/2017. This is a tertiary referral centre for respiratory medicine, with one specialist admissions ward and 3 inpatient wards. The analysis included all adults admitted with respiratory disease not transferred to a higher level of care, i.e. high dependency or intensive care, greater than 24 hours before death as these areas are not currently employing electronic observations, long term ventilator dependent patients were also excluded as hospital policy dictates that these patients are always admitted to the high dependency unit. Following approval from the NHS Information Governance Lead, and in line with existing permissions within the East Midlands Academic Health Sciences network, data from the integrated electronic observation and communication system comprising respiratory rate, oxygen saturations, heart rate, blood pressure, temperature, conscious level (AVPU score), and urine output were anonymised by an NHS data analyst prior to extraction from the clinical server. The same system also automatically generates mandated escalation and referral at set scoring thresholds via a pre-determined protocol. Scores from the local EWS were linked to demographics and mortality

outcomes prior to extraction. NEWS criteria were applied retrospectively to determine how many patients would have been escalated if the NEWS system were followed. Results were analysed using STATA 15. The entire data set was analysed for measurement of escalation patterns, analysis of workload and sensitivity and specificity in predicting death within 24 hours of an observation [4]. A chi-square analysis was performed to demonstrate whether the difference in escalations was significant. The statistical analysis involved the use of all vital signs observations recorded throughout admission, which were linked to outcome to determine whether they were followed by death within 24 hours of the observation timestamp created by the input devices at the bedside. Observations coded as end of life care following clinical decision were excluded from mortality analysis (see Figure 2). A further subgroup analysis was then performed on patients coded as having chronic obstructive pulmonary disease at any point in their admission as per ICD10 codes in order to further assess the statistical performance of the two scores in the presence of chronic pathophysiology.

Figure 2. Cohort flow diagram of exclusion criteria

Patient and public involvement

Prior to carrying out this work, a questionnaire was performed amongst stakeholders, in this case 26 medical registrars working in the East Midlands region. All worked in acute trusts that employed either NEWS or the Nottingham EWS as part of a system to highlight patients felt to be at risk of deterioration. Of the stakeholder responders, 70% believed that using EWS failed to highlight all patients who went on to deteriorate and 88% felt that use of an EWS led to unnecessary reviews. All responders felt there were issues in the setting of chronic disease with some chronic patients scoring even at baseline, and 76% felt that alert fatigue due to high EWS was an issue. These findings guided the interrogation of the data in creating the study detailed in this paper. It is also worth noting that similar work presented to patients with recent inpatient experience at NUHT highlighted the belief that sleep was too often interrupted by observations or reviews.

Results

236,840 observation sets were recorded during 8812 inpatient episodes (53.1% female- see Table1) involving 6091 individuals. In-hospital mortality for respiratory patients was 5.9% (n=521) and median length of stay was 4 days (range 0-175).

Characteristics of the patients in this study	
	Numbers (%)
Male	3824 (47)
Female	4438 (53)
Total	8812
Mean age in years	
Male	63.7
Female	62.7
Total	63.1
Vital Signs	Mean (+/- SD)
Heart Rate (beats per minute)	87 (16)
Respiratory rate (breaths per minute)	19 (3)
Systolic BP (mmHg)	130 (22)
Temperature (°C)	36.6 (1)
Oxygen saturations (%)	94 (6)

Table 1. Population characteristics of study cohort

59,434 (25.1%) observations sets were recorded between the hours of 0900-1700 Monday to Friday (excluding bank holidays). 177,406 (74.9%) were recorded outside of these hours. The local EWS and escalation protocol led to a median of 36 (range 1-148, calculated from the raw data of scores between 3 and 5 each day) scores per day that triggered a medical review (Table 2). This included a median of 5 (range 0-41) automated referrals to the resident on call senior clinician (medical registrar) every day. Direct comparison of workload generated to other members of the clinical team was not possible as the escalation protocol for both scores is only directly comparable at registrar level, however the workload generated at each of the clinically applied cut points can be seen in tables 2 and 3.

If NEWS criteria were applied to the same population it would have generated a median of 98 (range 12-270) escalations to a doctor per day ($p<0.001$ for difference between scores), with 38 (range 2-

158) scores generating automatic referral to the registrar (p<0.001 for difference between scores) per day.

Sensitivity and specificity for predicting in-hospital mortality based on death within 24 hours of a set of vital signs observations point is shown in Table 2. At each clinically equivalent band, the sensitivity and specificity in predicting mortality of all patients scoring at and above that cut point are shown. At each cut point, NEWS would have had a higher sensitivity than the local EWS (i.e. a higher percentage of patients who went on to die were flagged as requiring escalation), but a lower specificity.

Figure 3 plots sensitivity in predicting mortality, against median number of mandated clinician alerts per day for both EWS types. It demonstrates that for a sensitivity of 0.7, NEWS generates a higher number of mandated escalations. At both extremes of sensitivity (0 and 1) the number of escalations is the same: i.e. mandating an escalation at a NEWS or EWS of 0 would mean all patients were escalated, and each score would have 100% sensitivity for predicting mortality (as everyone who died would have been reviewed). Likewise only escalating patients with a maximum EWS or NEWS score would lead to very few patients being escalated.

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

Further subgroup analysis was performed on admissions with an ICD10 code for COPD at any point. This yielded 56,345 observations from 2207 episodes by 1365 individual patients. Using the local EWS protocol led to median of 0 (range 0-19) escalations to the registrar, while applying NEWS would have generated a median of 6 (0-47) scores being escalated to the registrar each day. As in the unselected respiratory cohort, NEWS was more sensitive in predicting imminent mortality than the local EWS but with a significantly inferior specificity at each clinical cut point applied (see table 3).

Discussion

In this study we examined the effect of two different early warning score systems in patients admitted with respiratory disease to a tertiary referrals centre. The respiratory department at NUHT manages patients in line with national guidelines and has outcomes comparable with other similar

units; consequently linking of raw observations to outcomes prior to analysis enables conclusions which are applicable to other centres.

We analysed the number of mandatory escalations generated and the sensitivity and specificity of both of the scores in predicting imminent in-hospital mortality in an unselected respiratory population and in a subgroup analysis of patients with COPD. Our data shows that at the scores' cut points for escalation, NEWS would have generated a significantly higher workload due to a lower specificity, with a higher sensitivity for predicting imminent deterioration, when compared with the locally used EWS. This was accentuated in patients with COPD, an observation we believe is due to chronic changes in the underlying physiology which influences the way in which these patients respond to acute pathological processes.

Although NEWS may become less relevant with the publication of NEWS2 in December 2017, our study remains relevant. Firstly, it highlights the wider impact of the different approaches to designing a scoring system and the paucity of evidence in relation to how this is evaluated. Secondly, as it is currently unclear how widely NEWS2 has been adopted by hospitals across the NHS and what the likely roll out will be, NEWS remains a current clinical tool in many trusts.

Previous work has suggested that NEWS was less discriminative in predicting deterioration in patients with respiratory disease, compared to a population of unselected medical admissions [2], however NEWS has not previously been studied in large numbers of respiratory patients across an entire admission. Our study faced similar limitations to others, namely the low prevalence of mortality in the patient population and the difficulty of studying patients in real time. However, our observed findings of an increase workload generated are both novel and important as, when used as part of a system which employs automatic escalation of threshold scores, NEWS leads to a significant impact on work load in a resource pressured environment, with little evidence of improved clinical outcome. This While there is a difference in the workload generated when comparing the scoring systems both in a general respiratory population and in patients with COPD, this relates to the cut points for escalation mandated by the protocols, rather than the scores themselves; unsurprisingly overall both scores perform similarly when the individual scores are plotted (they are based on similar clinical observations) however the mandated cut points differ. The difference created by the protocol design relates to the way in which the scores are used clinically, and can be explained as follows:

The first approach is seen in the scoring thresholds dictated by NEWS. Its cut points for each layer of clinical intervention, i.e. escalation to nurse, clinician or registrar, have a higher sensitivity which acts to rule out imminent clinical deterioration in those patients whose vital signs do not meet scoring

thresholds, meaning clinicians can be confident that patients with a low score are very unlikely to be at imminent risk. This is akin to a d-dimer where a low value in an individual with low clinical suspicion effectively excludes a venous thromboembolism [5, 6] . This high sensitivity approach works well in a setting with less highly trained staff delivering the first layer of monitoring. However, if this approach is applied in an unfiltered and automated manner, the workload generated by escalations from patients who never go on to deteriorate will have significant resource and operational implications, as well as increasing the likelihood of unnecessary intervention for patients.

The second approach, used by the local EWS, is one of high specificity in the cut points for escalation, with a relatively lower sensitivity. This approach acts to highlight potential imminent clinical deterioration in those meeting the escalation criteria, but does not always rule out deterioration in those who score under the cut point. This may seem a less preferable approach. However a recent study of rapid response systems indicated that staff clinical concern in the absence of a qualifying score was responsible for escalation in 47% of calls [7], highlighting the role of staff education and empowerment, over and above EWS protocols The variability in physiological normal baselines created by patient specific factors such as comorbidity or fitness means that using vital signs observations alone as the basis for a score leading to mandatory escalation will always require a trade-off between sensitivity in accurately identifying patients potentially at risk of deterioration and staff alarm fatigue generated by patients who do not go onto deteriorate. This is particularly pertinent in resource limited environments (such as during out of hours care),

Despite the mandated and widespread uptake of EWS, there has been minimal prospective validation of their use. Efforts to improve precision in predicting outcome through scrutiny of large datasets has largely employed analyses utilising area under the receiver operating characteristic curves which are limited by the low prevalence of mortality in the population [8]. Before and after studies have largely, but not universally [9-11] highlighted the efficacy of EWS, however no randomised controlled trials have been performed. Consequently evidence of the scores' real impact on clinical outcomes, such as mortality, transfer to higher level of care or length of stay, or on workforce outcomes such as workload from excessive task generation and alarm fatigue, has only been obtained from observational studies. These are all limited by significant confounders.

This evidence gap around the clinical and workforce implications of EWS systems will become increasingly important as hospitals move towards automated systems with mandated referral of patients who reach a threshold score. Continuing integration of more data into digital healthcare systems via continuous monitoring, dynamic measures of fitness, and electronic health records will

further highlight this gap, as without an understanding of how these data can be applied it will be difficult to differentiate the signal from the noise. Given the growing complexity of the inpatient population more work is urgently required to understand the wider impact of EWS on outcomes such as mortality and length of stay, task burden, working patterns and cost. There is also need to reconsider the role of clinical concern in monitoring patients and how this can be further promoted to prevent future systems depending purely on scores rather than integrating staff skills and intuition into the decision making process. Early warning scores should not be developed in isolation based on statistical performance as this fails to recognise that they are a component within the complex clinical environment and therefore need to be designed to enhance, not complicate, the clinical decision making process. This is particularly important in patients with respiratory disease where physiology is often chronically deranged and less responsive to intervention and a greater understanding of the contributory clinical factors and more individualised approach is required. Although NEWS2 has been developed to address concerns regarding the altered physiology of patients with respiratory disease, the new score was not based on any significant development in the evidence base. Therefore the same questions currently remain regarding the real terms impact of introducing any EWS, including NEWS2, and the associated software platforms on the patients being monitored, the staff and resources required to deploy it and react to it, and the associated opportunity cost.

Healthcare is becoming increasingly individualised, with significant amounts of digital healthcare data collected. In recognition of this, a possible future direction would be to create scores which, rather than being based solely on observations, integrate other more patient specific factors such as comorbidity, premorbid fitness and age to apply specific weighting to observations. For example, through applying a lower score to a high respiratory rate in someone who had chronic respiratory disease and could mobilise 5 metres as a baseline as opposed to a young marathon runner, it would be possible to maintain the same scoring thresholds at which a response was triggered, while making those thresholds more meaningful through an evidence-based application of risk of deterioration based on what a clinical observation represents in a particular individual.

Analysis of big data is the first stage to making this possible. However, the ability to demonstrate the significance of changing either scoring thresholds or the scores themselves on patient and system outcomes, driven by an attempt to compensate for changes to existing baseline physiology, will require considerable numbers, novel prospective study design and collaboration across multiple sites and research disciplines.

However challenging, these points need to be addressed before any meaningful advances can be made in this area and to ensure the most effective use of resources in the pursuit of improving the safety and efficiency of patient care.

NEWS band	Mandated escalation to:	% of observations in each band	Median number per day (range)	Sensitivity for predicting death within 24 hours	Specificity for predicting death within 24 hours
0	Nil	17.86	32 (3-75)	100.00	0.00
1 to 4	Nurse	67.34	180 (21-457)	99.44	15.09
5 to 6	Doctor	8.82	60 (10-184)	88.64	74.51
7 or more	Registrar	5.97	38 (2-158)	68.53	91.16
NUH EWS band	Mandates escalation to:	% of observations in each band	Median number per day (range)	Sensitivity for predicting death within 24 hours	Specificity for predicting death within 24 hours
0	Nil	56.11	174 (20-409)	100.00	0
1 to 2	Nurse	31.83	99 (16-300)	95.49	56.32
3	Nurse/ Doctor	5.39	16 (1-116)	76.65	88.24
4 to 5	Doctor	4.74	14 (1-55)	63.33	93.58
6 or more	Registrar	1.94	5 (0-41)	41.91	98.24

Table 2- Workload predictions and sensitivity and specificity in predicting death within 24 hours for NEWS and local EWS for unselected respiratory population

NEWS band	Mandated Escalation	% of observations in each band	Median number (range)	Sensitivity for death within 24 hours	Specificity for death within 24 hours
0	Nil	7.96	5 (0-23)	100.00	0.00
1 to 4	Nurse	59.3	43 (4-112)	100.00	7.99
5 to 6	Doctor	22.2	16 (1-59)	89.85	67.47
7 or more	Registrar	10.54	6 (0-47)	71.07	89.68
NUH EWS band	Mandated Escalation	% of observations in each band	Median number (range)	Sensitivity for death within 24 hours	Specificity for death within 24 hours
0	Nil	53.89	39 (1-101)	100.00	0.00
1 to 2	Nurse	35.05	26 (4-90)	92.39	54.06
3	Nurse/Doctor	5.46	3 (0-30)	70.56	88.50
4 to 5	Doctor	4.31	2 (0-18)	58.38	94.20
6 or more	Registrar	1.28	0 (0-19)	38.07	98.75

Table 3- Workload predictions and sensitivity and specificity in predicting death within 24 hours for NEWS and local EWS for patients with COPD

Collaborators Dr Mark Simmonds was instrumental in delivering the electronic observations system at Nottingham University Hospitals Trust and has liaised with our group in relation to our work on EWS

Author Contributorship Statement

Sarah Forster was first author and performed initial data analysis, literature review and created initial document and revisions.

Gemma Housley is an NHS data analyst who extracted the data for the article and provided editorial input.

Tricia Mckeever is a University of Nottingham Statistician who provided statistical advice and oversight with editorial input.

Dominick Shaw is the supervising author. In collaboration with the first author he developed the protocol, advising on data sources and research question while providing editorial input at all stages of the manuscript's development.

Competing Interests

None declared

Data Sharing Statement

Unpublished metadata from the dataset analysed for this work is available through contacting the authors. However, due to the nature of the raw data it is not possible to make it freely available due to the limitations placed on the use of the dataset by NHS Information Governance procedures and approvals.

Funding Statement

There is no funding to declare for this project

Corresponding Author Email

Dr Sarah Forster, sarahforster@nhs.net

References

1. Physicians, R.C.o., *National Early Warning Score (NEWS): Standardising the assessment of acute illness severity in the NHS.* 2012, Royal College of Physicians.
2. Hodgson, L.E., B.D. Dimitrov, J. Congleton, 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.
3. Eccles, S.R., C. Subbe, D. Hancock, and N. Thomson, *CREWS: improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia.* Resuscitation, 2014. **85**(1): p. 109-11.
4. Jarvis, S.W., C. Kovacs, J. Briggs, et al., *Are observation selection methods important when comparing early warning score performance?* Resuscitation, 2015. **90**: p. 1-6.
5. Stein, P.D., R.D. Hull, K.C. Patel, et al., *D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism: a systematic review.* Ann Intern Med, 2004. **140**(8): p. 589-602.
6. Owaidah, T., N. AlGhasham, S. AlGhamdi, et al., *Evaluation of the usefulness of a D dimer test in combination with clinical pretest probability score in the prediction and exclusion of Venous Thromboembolism by medical residents.* Thromb J, 2014. **12**(1): p. 28.
7. McGaughey, J., P. O'Halloran, S. Porter, J. Trinder, and B. Blackwood, *Early warning systems and rapid response to the deteriorating patient in hospital: A realist evaluation.* J Adv Nurs, 2017.
8. Romero-Brufau, S., J.M. Huddleston, G.J. Escobar, and M. Liebow, *Why the C-statistic is not informative to evaluate early warning scores and what metrics to use.* Crit Care, 2015. **19**: p. 285.
9. Subbe, C.P., R.G. Davies, E. Williams, P. Rutherford, and L. Gemmell, *Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions.* Anaesthesia, 2003. **58**(8): p. 797-802.
10. Booth, C., *Effect on outcome of an early warning score in acute medical admissions.* British Journal of Anaesthesia, 2003. **90**(4): p. 1.
11. Moon, A., J.F. Cosgrove, D. Lea, A. Fairs, and D.M. Cressey, *An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR.* Resuscitation, 2011. **82**(2): p. 150-4.

	NEWS		Nottingham Early Warning Score	
	Monitoring frequency	Clinical Response	Monitoring frequency	Clinical Response
0	Min 12 hourly	Continue routine NEWS monitoring	Min 12 hourly	
1	Min 4-6 hourly	Inform RN who must assess patient RN to decide re. increased monitoring or escalation	Min 4-6 hourly	
2				
3			Hourly for minimum 2 hours Consider fluid balance monitoring	RN to recheck Inform NIC Manual BP if deranged
4			Hourly with fluid balance monitoring	RN to check EWS F1/SHO to discuss management with SpR or consultant
5	Min 1 hourly	RN to urgently inform medical team Urgent assessment by clinician with core competencies Clinical care in environment with monitoring facilities		
6			Minimum ½ hourly with hourly fluid balance. Consider GCS	SpR Review If no improvement following intervention for discussion with consultant SpR to contact critical care for advice
>7	Continuous	RN to immediately inform medical team at minimum SpR Emergency assessment by clinical team with critical care competencies Consider transfer to level 2/3 care		

Weighted score	3	2	1	0	1	2	3
Respiratory rate	<8	-	9-11	12-20	-	21-24	>25
Oxygen saturation (NEWS only)	<91	92-93	94-95	>96	-	-	-
Supplemental oxygen	-	**NEWS-**Yes	-	**NEWS-**No **Notts-** 0-9L/min	**Notts-** 10-14L/min		**Notts-** ≥15L/min
Heart Rate	<40	-	41-50	51-90	91-110	111-130	≥131
Blood Pressure	<90	91-100	101-110	111-219	-	-	≥220
Temperature	<35	-	35.1-36	36.1-38	38.1-39	>39.1	-
Urine output (Notts only)	-	Not PU'd in 12 hours or <20 ml/hour	Not PU'd 6 hours or 20-30	PU'd in last 6 hours or 31-199ml/hour	>200ml/hr	-	-
AVPU	**Both-** Unresponsive	**Notts-** Pain	**Notts-** Voice	**Both-** Alert			

Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham University Hospitals Early Warning Score

105x143mm (300 x 300 DPI)

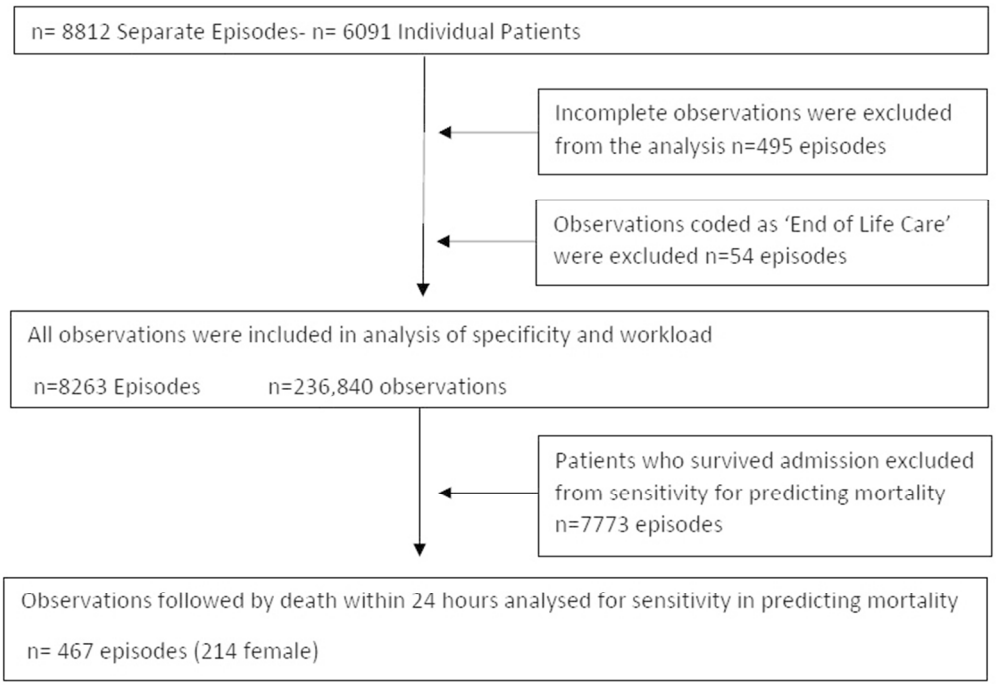

Figure 2. Cohort flow diagram of exclusion criteria

72x49mm (300 x 300 DPI)

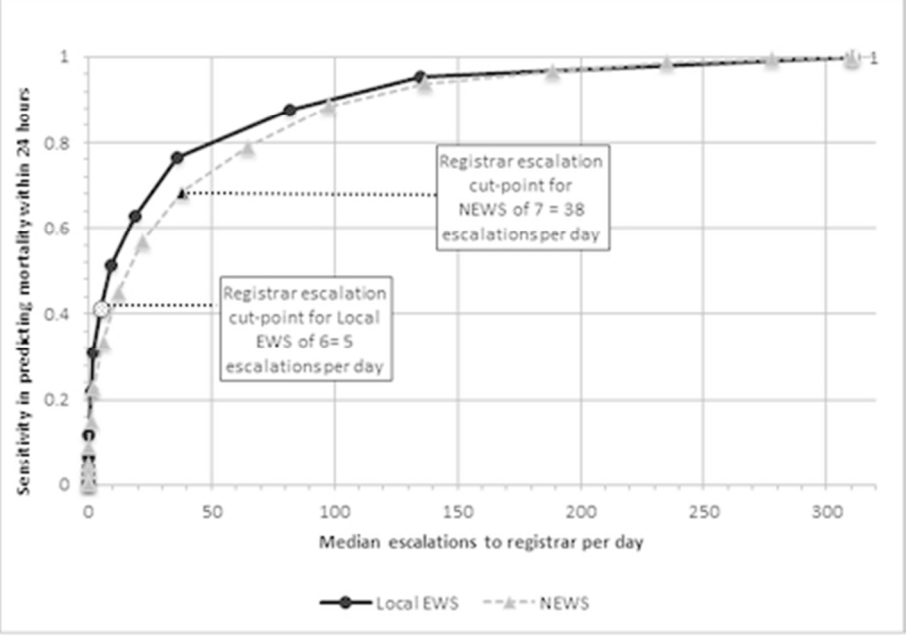

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

38x29mm (300 x 300 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **[See design section of abstract pages 1-2]**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **[See abstract pages 1-2]**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **[See abstract and Background sections- pages 1-2]**
Objectives	3	State specific objectives, including any prespecified hypotheses **[See objective section in abstract]**
Methods		
Study design	4	Present key elements of study design early in the paper **[methods in abstract – expanded in methods section of main article on page 3]**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **[Methods and Setting in abstract, Methods page 3]**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **[Methods section page 3]**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **[N/A]**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **[Background, Methods, Discussion]**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **[Methods page 3]**
Bias	9	Describe any efforts to address potential sources of bias **[N/A]**
Study size	10	Explain how the study size was arrived at **[Methods page 3]**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **[Methods page 3]**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **[Methods page 3]**
		(*b*) Describe any methods used to examine subgroups and interactions **[Methods page 3]**
		(*c*) Explain how missing data were addressed **[Cohort Diagram- Figure 2, Methods page 3]**
		(*d*) If applicable, explain how loss to follow-up was addressed **[n/a]**
		(*e*) Describe any sensitivity analyses **[Methods page 3]**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **[Results page 4]**
		(b) Give reasons for non-participation at each stage **[N/A]**
		(c) Consider use of a flow diagram **[Cohort flow diagram in methods section]**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **[Table 1- characteristics of**

		study population]
		(b) Indicate number of participants with missing data for each variable of interest **[See cohort flow diagram- figure 2 in methods page 3]**
		(c) Summarise follow-up time (eg, average and total amount) **[N/A]**
Outcome data	15*	Report numbers of outcome events or summary measures over time **[Abstract and results]**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **[N/A]**
		(*b*) Report category boundaries when continuous variables were categorized **[See figure 1 + Table 2]**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **[n/a]**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **[Methods and results page 3-4]**
Discussion		
Key results	18	Summarise key results with reference to study objectives **[Discussion page 5 onwards and abstract]**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **[Strengths and weaknesses page 1; Discussion page 5 onwards]**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **[Background page 2 and Discussion page 5 onwards]**
Generalisability	21	Discuss the generalisability (external validity) of the study results **[Discussion page 5]**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **[Included after main body of text before references]**

*Give information separately for exposed and unexposed groups.

BMJ Open

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.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020269.R3
Article Type:	Research
Date Submitted by the Author:	24-Apr-2018
Complete List of Authors:	Forster, Sarah; University of Nottingham School of Medicine, Respiratory Medicine Housley, Gemma; Nottingham University Hospitals NHS Trust McKeever, Tricia; University of Nottingham, Division of Epidemiology and Public Health Shaw, Dominick; University of Nottingham, Respiratory Medicine
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Evidence based practice
Keywords:	Thoracic medicine < INTERNAL MEDICINE, Adult intensive & critical care < INTENSIVE & CRITICAL CARE, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

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.

Sarah Forster[1,2], Gemma Housley[3], Tricia M McKeever[4], Dominick E Shaw[2,3]

Author Affiliations

[1] NIHR Academic Clinical Fellow, University of Nottingham, Nottingham, UK

[2] Respiratory Research Unit, Division of Respiratory Medicine, University of Nottingham, Nottingham, UK

[3] Medical Informatics, East Midlands Academic Health Sciences Network, Nottingham, UK

[4] Division of Epidemiology, University of Nottingham, Nottingham, UK

Correspondence to Dr Sarah Forster, sarahforster@nhs.net

Word count: 3217

Abstract

Objective

Early Warning Scores (EWS) are used to monitor patients for signs of imminent deterioration. Although used in respiratory disease, EWS have not been well studied in this population, despite the underlying cardiopulmonary pathophysiology often present. We examined the performance of two scoring systems in patients with respiratory disease.

Design

Retrospective cohort analysis of vital signs observations of all patients admitted to a respiratory unit over a 2 year period. Scores were linked to outcome data to establish the performance of the National EWS (NEWS) compared results to a locally adapted EWS.

Setting

Nottingham University Hospitals NHS Trust respiratory wards. Data were collected from an integrated electronic observation and task allocation system employing a local EWS, also generating mandatory referrals to clinical staff at set scoring thresholds.

Outcome Measures

Projected workload, and sensitivity and specificity of the scores in predicting mortality based on outcome within 24 hours of a score being recorded.

Results

8812 individual patient episodes occurred during the study period. Overall mortality was 5.9%. Applying NEWS retrospectively (versus local EWS) generated an eight fold increase in mandatory escalations, but had higher sensitivity in predicting mortality at the protocol cut points.

Conclusions

This study highlights issues surrounding use of scoring systems in patients with respiratory disease. NEWS demonstrated higher sensitivity for predicting death within 24 hours, offset by reduced specificity. The consequent workload generated may compromise the ability of the clinical team to respond to patients needing immediate input. The locally adapted EWS has higher specificity but lower sensitivity. Statistical evaluation suggests this may lead to missed opportunities for intervention, however this does not account for clinical concern independent of the scores, nor ability to respond to alerts based on workload. Further research into the role of warning scores and the impact of chronic pathophysiology is urgently needed.

Strengths and Limitations of this Study

- Data were obtained from a large clinical vital signs database with clear identification of specialty allowing for subgroup analysis. All observations were included in the analysis, regardless of whether there had previously been a high score which may have resulted in a change of management by the clinical team
- Granularity of data collection in the database allowed for reliable identification of patients meeting the exclusion criteria. Only 0.2% of the observations recorded during the study period were identified as being incomplete.
- The retrospective nature of study precludes conclusions relating to impact of introducing NEWS on mortality.
- DNACPR decisions were not linked as part of the analysis.
- Inherent inaccuracy in recording time of death in hospital records means 24 hour cut off may not be always be exact

Background

Early Warning Scores combine vital sign measures into a composite score in order to identify patients at risk of clinical deterioration, guide early intervention and reduce avoidable mortality. Scores have evolved over the last 30 years following the recognition that patients experiencing a serious adverse event, such as unplanned transfer to intensive care, in hospital cardiac arrest or death, showed evidence of pathophysiology in their vital signs observations in the hours leading up to overt deterioration. Initially this information was captured in the form of single parameter scores where significant derangement in a single vital sign or clinical concern triggered a set clinical response. In the UK this led to the development of aggregate weighted scores, whereby each vital sign is given a weighting depending on how far outside the predetermined normal range it falls; the sum of these scores is then used to guide response.

In 2012 the Royal College of Physicians published the NEWS protocol in an attempt to standardise processes for identifying patients at risk of imminent deterioration [1]. EWS protocols guide decisions around patient care by mandating when a patient with evidence of pathophysiology, in the form of deranged vital signs, should be reviewed by a clinical member of staff, and therefore influence overall clinical workload and resource allocation for all in-patients. Patients with respiratory disease make up a large proportion of a hospital's in-patient population, however it is recognised that chronic physiological disturbance caused by COPD may render NEWS less

discriminative when compared to an unselected medical population [2]. This has significant implications for patients, in terms of increased observations and interventions, and to clinical staff in terms of workload and potential for alert fatigue. Consequently attempts have been made to improve the score in this population [3].

Nottingham University Hospitals NHS Trust employs an electronic observations system with mandatory escalation based on an adapted EWS. The Nottingham EWS, unlike NEWS, does not score oxygen saturations and has a graduated approach to weighting for both oxygen delivery and level of consciousness. As a more general marker of morbidity it also employs urine output. We compared the sensitivity and specificity of the two scores in predicting mortality within 24 hours of a set of observations being recorded at the clinical cut points determined by the associated protocols and examined the potential impact in terms of workload of using the locally designed EWS versus NEWS (see Figure 1) in patients with respiratory disease based on analysis of the vital signs observations and outcomes of patients admitted to the respiratory department in Nottingham over a 2 year period. We then went on to answer the same questions in a subgroup of patients who were admitted with a diagnosis of COPD to examine the performance of the two scores in this cohort.

Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham University Hospitals Early Warning Score

Methods

We performed a single centre retrospective analysis of all patients admitted to the respiratory department at Nottingham University Hospitals NHS Trust between 01/04/2015 and 31/03/2017. This is a tertiary referral centre for respiratory medicine, with one specialist admissions ward and 3 inpatient wards. The analysis included all adults admitted with respiratory disease not transferred to a higher level of care, i.e. high dependency or intensive care, greater than 24 hours before death as these areas are not currently employing electronic observations, long term ventilator dependent patients were also excluded as hospital policy dictates that these patients are always admitted to the high dependency unit. Following approval from the NHS Information Governance Lead, and in line with existing permissions within the East Midlands Academic Health Sciences network, data from the integrated electronic observation and communication system comprising respiratory rate, oxygen saturations, heart rate, blood pressure, temperature, conscious level (AVPU score), and urine output were anonymised by an NHS data analyst prior to extraction from the clinical server. The same system also automatically generates mandated escalation and referral at set scoring thresholds via a pre-determined protocol. Scores from the local EWS were linked to demographics and mortality outcomes prior to extraction. NEWS criteria were applied retrospectively to determine how many patients would have been escalated if the NEWS system were followed. Results were analysed using

STATA 15. The entire data set was analysed for measurement of escalation patterns, analysis of workload and sensitivity and specificity in predicting death within 24 hours of an observation [4]. A chi-square analysis was performed to demonstrate whether the difference in escalations was significant. The statistical analysis involved the use of all vital signs observations recorded throughout admission, which were linked to outcome to determine whether they were followed by death within 24 hours of the observation timestamp created by the input devices at the bedside. Observations coded as end of life care following clinical decision were excluded from mortality analysis (see Figure 2). A further subgroup analysis was then performed on patients coded as having chronic obstructive pulmonary disease at any point in their admission as per ICD10 codes in order to further assess the statistical performance of the two scores in the presence of chronic pathophysiology.

Figure 2. Cohort flow diagram of exclusion criteria

Patient and public involvement

Prior to carrying out this work, a questionnaire was performed amongst stakeholders, in this case 26 medical registrars working in the East Midlands region. All worked in acute trusts that employed either NEWS or the Nottingham EWS as part of a system to highlight patients felt to be at risk of deterioration. Of the stakeholder responders, 70% believed that using EWS failed to highlight all patients who went on to deteriorate and 88% felt that use of an EWS led to unnecessary reviews. All responders felt there were issues in the setting of chronic disease with some chronic patients scoring even at baseline, and 76% felt that alert fatigue due to high EWS was an issue. These findings guided the interrogation of the data in creating the study detailed in this paper. It is also worth noting that similar work presented to patients with recent inpatient experience at NUHT highlighted the belief that sleep was too often interrupted by observations or reviews. However patients were not involved directly in the design of this study.

Results

236,840 observation sets were recorded during 8812 inpatient episodes (53.1% female- see Table1) involving 6091 individuals. In-hospital mortality for respiratory patients was 5.9% (n=521) and median length of stay was 4 days (range 0-175).

Characteristics of the patients in this study	
	Numbers (%)
Male	3824 (47)
Female	4438 (53)
Total	8812
Mean age in years	
Male	63.7
Female	62.7
Total	63.1
Vital Signs	Mean (+/- SD)
Heart Rate (beats per minute)	87 (16)
Respiratory rate (breaths per minute)	19 (3)
Systolic BP (mmHg)	130 (22)
Temperature (°C)	36.6 (1)
Oxygen saturations (%)	94 (6)

Table 1. Population characteristics of study cohort

59,434 (25.1%) observations sets were recorded between the hours of 0900-1700 Monday to Friday (excluding bank holidays). 177,406 (74.9%) were recorded outside of these hours. The local EWS and escalation protocol led to a median of 36 (range 1-148, calculated from the raw data of scores between 3 and 5 each day) scores per day that triggered a medical review (Table 2). This included a median of 5 (range 0-41) automated referrals to the resident on call senior clinician (medical registrar) every day. Direct comparison of workload generated to other members of the clinical team was not possible as the escalation protocol for both scores is only directly comparable at registrar level, however the workload generated at each of the clinically applied cut points can be seen in tables 2 and 3.

If NEWS criteria were applied to the same population it would have generated a median of 98 (range 12-270) escalations to a doctor per day ($p<0.001$ for difference between scores), with 38 (range 2-

158) scores generating automatic referral to the registrar (p<0.001 for difference between scores) per day.

Sensitivity and specificity for predicting in-hospital mortality based on death within 24 hours of a set of vital signs observations point is shown in Table 2. At each clinically equivalent band, the sensitivity and specificity in predicting mortality of all patients scoring at and above that cut point are shown. At each cut point, NEWS would have had a higher sensitivity than the local EWS (i.e. a higher percentage of patients who went on to die were flagged as requiring escalation), but a lower specificity.

Figure 3 plots sensitivity in predicting mortality, against median number of mandated clinician alerts per day for both EWS types. It demonstrates that for a sensitivity of 0.7, NEWS generates a higher number of mandated escalations. At both extremes of sensitivity (0 and 1) the number of escalations is the same: i.e. mandating an escalation at a NEWS or EWS of 0 would mean all patients were escalated, and each score would have 100% sensitivity for predicting mortality (as everyone who died would have been reviewed). Likewise only escalating patients with a maximum EWS or NEWS score would lead to very few patients being escalated.

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

Further subgroup analysis was performed on admissions with an ICD10 code for COPD at any point. This yielded 56,345 observations from 2207 episodes by 1365 individual patients. Using the local EWS protocol led to median of 0 (range 0-19) escalations to the registrar, while applying NEWS would have generated a median of 6 (0-47) scores being escalated to the registrar each day. As in the unselected respiratory cohort, NEWS was more sensitive in predicting imminent mortality than the local EWS but with a significantly inferior specificity at each clinical cut point applied (see table 3).

Discussion

In this study we examined the effect of two different early warning score systems in patients admitted with respiratory disease to a tertiary referrals centre. The respiratory department at NUHT manages patients in line with national guidelines and has outcomes comparable with other similar units; consequently linking of raw observations to outcomes prior to analysis enables conclusions which are applicable to other centres.

We analysed the number of mandatory escalations generated and the sensitivity and specificity of both of the scores in predicting imminent in-hospital mortality in an unselected respiratory population and in a subgroup analysis of patients with COPD. Our data shows that at the scores' cut points for escalation, NEWS would have generated a significantly higher workload due to a lower specificity, with a higher sensitivity for predicting imminent deterioration, when compared with the locally used EWS. This was accentuated in patients with COPD, an observation we believe is due to chronic changes in the underlying physiology which influences the way in which these patients respond to acute pathological processes.

Although NEWS may become less relevant with the publication of NEWS2 in December 2017, our study remains relevant. Firstly, it highlights the wider impact of the different approaches to designing a scoring system and the paucity of evidence in relation to how this is evaluated. Secondly, as it is currently unclear how widely NEWS2 has been adopted by hospitals across the NHS and what the likely roll out will be, NEWS remains a current clinical tool in many trusts.

Previous work has suggested that NEWS was less discriminative in predicting deterioration in patients with respiratory disease, compared to a population of unselected medical admissions [2], however NEWS has not previously been studied in large numbers of respiratory patients across an entire admission.

Our study faced similar limitations to others published in this area. These include retrospective study design preventing analysis of the real terms impact of introducing different scores into the study environement on outcomes including length of stay, cardiac arrest rate and mortality; the low prevalence of mortality in the patient population and the subsequent impact on observed effect size; and the difficulty in recording accurate time of death in a general ward setting for use in mortality analysis. .

However, our observed findings of an increase workload generated are both novel and important as, when used as part of a system which employs automatic escalation of threshold scores, NEWS leads to a significant impact on work load in a resource pressured environment, with little evidence of improved clinical outcome. While there is a difference in the workload generated when comparing the scoring systems both in a general respiratory population and in patients with COPD, this relates to the cut points for escalation mandated by the protocols, rather than the scores themselves; unsurprisingly overall both scores perform similarly when the individual scores are plotted (they are

based on similar clinical observations) however the mandated cut points differ. The difference created by the protocol design relates to the way in which the scores are used clinically, and can be explained as follows:

The first approach is seen in the scoring thresholds dictated by NEWS. Its cut points for each layer of clinical intervention, i.e. escalation to nurse, clinician or registrar, have a higher sensitivity which acts to rule out imminent clinical deterioration in those patients whose vital signs do not meet scoring thresholds, meaning clinicians can be confident that patients with a low score are very unlikely to be at imminent risk. This is akin to a d-dimer where a low value in an individual with low clinical suspicion effectively excludes a venous thromboembolism [5, 6] . This high sensitivity approach works well in a setting with less highly trained staff delivering the first layer of monitoring. However, if this approach is applied in an unfiltered and automated manner, the workload generated by escalations from patients who never go on to deteriorate will have significant resource and operational implications, as well as increasing the likelihood of unnecessary intervention for patients.

The second approach, used by the local EWS, is one of high specificity in the cut points for escalation, with a relatively lower sensitivity. This approach acts to highlight potential imminent clinical deterioration in those meeting the escalation criteria, but does not always rule out deterioration in those who score under the cut point. This may seem a less preferable approach. However a recent study of rapid response systems indicated that staff clinical concern in the absence of a qualifying score was responsible for escalation in 47% of calls [7], highlighting the role of staff education and empowerment, over and above EWS protocols The variability in physiological normal baselines created by patient specific factors such as comorbidity or fitness means that using vital signs observations alone as the basis for a score leading to mandatory escalation will always require a trade-off between sensitivity in accurately identifying patients potentially at risk of deterioration and staff alarm fatigue generated by patients who do not go onto deteriorate. This is particularly pertinent in resource limited environments (such as during out of hours care),

Despite the mandated and widespread uptake of EWS, there has been minimal prospective validation of their use. Efforts to improve precision in predicting outcome through scrutiny of large datasets has largely employed analyses utilising area under the receiver operating characteristic curves which are limited by the low prevalence of mortality in the population [8]. Before and after studies have largely, but not universally [9-11] highlighted the efficacy of EWS, however no randomised controlled trials have been performed. Consequently evidence of the scores' real impact on clinical outcomes, such as mortality, transfer to higher level of care or length of stay, or on

workforce outcomes such as workload from excessive task generation and alarm fatigue, has only been obtained from observational studies. These are all limited by significant confounders.

This evidence gap around the clinical and workforce implications of EWS systems will become increasingly important as hospitals move towards automated systems with mandated referral of patients who reach a threshold score. Continuing integration of more data into digital healthcare systems via continuous monitoring, dynamic measures of fitness, and electronic health records will further highlight this gap, as without an understanding of how these data can be applied it will be difficult to differentiate the signal from the noise. Given the growing complexity of the inpatient population more work is urgently required to understand the wider impact of EWS on outcomes such as mortality and length of stay, task burden, working patterns and cost. There is also need to reconsider the role of clinical concern in monitoring patients and how this can be further promoted to prevent future systems depending purely on scores rather than integrating staff skills and intuition into the decision making process. Early warning scores should not be developed in isolation based on statistical performance as this fails to recognise that they are a component within the complex clinical environment and therefore need to be designed to enhance, not complicate, the clinical decision making process. This is particularly important in patients with respiratory disease where physiology is often chronically deranged and less responsive to intervention and a greater understanding of the contributory clinical factors and more individualised approach is required. Although NEWS2 has been developed to address concerns regarding the altered physiology of patients with respiratory disease, the new score was not based on any significant development in the evidence base. Therefore the same questions currently remain regarding the real terms impact of introducing any EWS, including NEWS2, and the associated software platforms on the patients being monitored, the staff and resources required to deploy it and react to it, and the associated opportunity cost.

Healthcare is becoming increasingly individualised, with significant amounts of digital healthcare data collected. In recognition of this, a possible future direction would be to create scores which, rather than being based solely on observations, integrate other more patient specific factors such as comorbidity, premorbid fitness and age to apply specific weighting to observations. For example, through applying a lower score to a high respiratory rate in someone who had chronic respiratory disease and could mobilise 5 metres as a baseline as opposed to a young marathon runner, it would be possible to maintain the same scoring thresholds at which a response was triggered, while making those thresholds more meaningful through an evidence-based application of risk of deterioration based on what a clinical observation represents in a particular individual.

Analysis of big data is the first stage to making this possible. However, the ability to demonstrate the significance of changing either scoring thresholds or the scores themselves on patient and system outcomes, driven by an attempt to compensate for changes to existing baseline physiology, will require considerable numbers, novel prospective study design and collaboration across multiple sites and research disciplines.

However challenging, these points need to be addressed before any meaningful advances can be made in this area and to ensure the most effective use of resources in the pursuit of improving the safety and efficiency of patient care.

NEWS band	Mandated escalation to:	% of observations in each band	Median number per day (range)	Sensitivity for predicting death within 24 hours	Specificity for predicting death within 24 hours
0	Nil	17.86	32 (3-75)	100.00	0.00
1 to 4	Nurse	67.34	180 (21-457)	99.44	15.09
5 to 6	Doctor	8.82	60 (10-184)	88.64	74.51
7 or more	Registrar	5.97	38 (2-158)	68.53	91.16
NUH EWS band	Mandates escalation to:	% of observations in each band	Median number per day (range)	Sensitivity for predicting death within 24 hours	Specificity for predicting death within 24 hours
0	Nil	56.11	174 (20-409)	100.00	0
1 to 2	Nurse	31.83	99 (16-300)	95.49	56.32
3	Nurse/ Doctor	5.39	16 (1-116)	76.65	88.24
4 to 5	Doctor	4.74	14 (1-55)	63.33	93.58
6 or more	Registrar	1.94	5 (0-41)	41.91	98.24

Table 2- Workload predictions and sensitivity and specificity in predicting death within 24 hours for NEWS and local EWS for unselected respiratory population

NEWS band	Mandated Escalation	% of observations in each band	Median number (range)	Sensitivity for death within 24 hours	Specificity for death within 24 hours
0	Nil	7.96	5 (0-23)	100.00	0.00
1 to 4	Nurse	59.3	43 (4-112)	100.00	7.99
5 to 6	Doctor	22.2	16 (1-59)	89.85	67.47
7 or more	Registrar	10.54	6 (0-47)	71.07	89.68
NUH EWS band	Mandated Escalation	% of observations in each band	Median number (range)	Sensitivity for death within 24 hours	Specificity for death within 24 hours
0	Nil	53.89	39 (1-101)	100.00	0.00
1 to 2	Nurse	35.05	26 (4-90)	92.39	54.06
3	Nurse/Doctor	5.46	3 (0-30)	70.56	88.50

4 to 5	Doctor	4.31	2 (0-18)	58.38	94.20
6 or more	Registrar	1.28	0 (0-19)	38.07	98.75

Table 3- Workload predictions and sensitivity and specificity in predicting death within 24 hours for NEWS and local EWS for patients with COPD

Collaborators Dr Mark Simmonds was instrumental in delivering the electronic observations system at Nottingham University Hospitals Trust and has liaised with our group in relation to our work on EWS

Author Contributorship Statement

Sarah Forster was first author and performed initial data analysis, literature review and created initial document and revisions.

Gemma Housley is an NHS data analyst who extracted the data for the article and provided editorial input.

Tricia Mckeever is a University of Nottingham Statistician who provided statistical advice and oversight with editorial input.

Dominick Shaw is the supervising author. In collaboration with the first author he developed the protocol, advising on data sources and research question while providing editorial input at all stages of the manuscript's development.

Competing Interests None declared

Data Sharing Statement

Unpublished metadata from the dataset analysed for this work is available through contacting the authors. However, due to the nature of the raw data it is not possible to make it freely available due to the limitations placed on the use of the dataset by NHS Information Governance procedures and approvals.

Funding Statement There is no funding to declare for this project

Corresponding Author Email

Dr Sarah Forster, sarahforster@nhs.net

References

1. Physicians, R.C.o., *National Early Warning Score (NEWS): Standardising the assessment of acute illness severity in the NHS.* 2012, Royal College of Physicians.
2. Hodgson, L.E., B.D. Dimitrov, J. Congleton, 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.
3. Eccles, S.R., C. Subbe, D. Hancock, and N. Thomson, *CREWS: improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia.* Resuscitation, 2014. **85**(1): p. 109-11.
4. Jarvis, S.W., C. Kovacs, J. Briggs, et al., *Are observation selection methods important when comparing early warning score performance?* Resuscitation, 2015. **90**: p. 1-6.
5. Stein, P.D., R.D. Hull, K.C. Patel, et al., *D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism: a systematic review.* Ann Intern Med, 2004. **140**(8): p. 589-602.
6. Owaidah, T., N. AlGhasham, S. AlGhamdi, et al., *Evaluation of the usefulness of a D dimer test in combination with clinical pretest probability score in the prediction and exclusion of Venous Thromboembolism by medical residents.* Thromb J, 2014. **12**(1): p. 28.
7. McGaughey, J., P. O'Halloran, S. Porter, J. Trinder, and B. Blackwood, *Early warning systems and rapid response to the deteriorating patient in hospital: A realist evaluation.* J Adv Nurs, 2017.
8. Romero-Brufau, S., J.M. Huddleston, G.J. Escobar, and M. Liebow, *Why the C-statistic is not informative to evaluate early warning scores and what metrics to use.* Crit Care, 2015. **19**: p. 285.
9. Subbe, C.P., R.G. Davies, E. Williams, P. Rutherford, and L. Gemmell, *Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions.* Anaesthesia, 2003. **58**(8): p. 797-802.
10. Booth, C., *Effect on outcome of an early warning score in acute medical admissions.* British Journal of Anaesthesia, 2003. **90**(4): p. 1.
11. Moon, A., J.F. Cosgrove, D. Lea, A. Fairs, and D.M. Cressey, *An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR.* Resuscitation, 2011. **82**(2): p. 150-4.

	NEWS		Nottingham Early Warning Score	
	Monitoring frequency	Clinical Response	Monitoring frequency	Clinical Response
0	Min 12 hourly	Continue routine NEWS monitoring	Min 12 hourly	
1	Min 4-6 hourly	Inform RN who must assess patient RN to decide re. increased monitoring or escalation	Min 4-6 hourly	
2				
3			Hourly for minimum 2 hours Consider fluid balance monitoring	RN to recheck Inform NIC Manual BP if deranged
4			Hourly with fluid balance monitoring	RN to check EWS F1/SHO to discuss management with SpR or consultant
5	Min 1 hourly	RN to urgently inform medical team Urgent assessment by clinician with core competencies Clinical care in environment with monitoring facilities		
6			Minimum ½ hourly with hourly fluid balance. Consider GCS	SpR Review If no improvement following intervention for discussion with consultant SpR to contact critical care for advice
>7	Continuous	RN to immediately inform medical team at minimum SpR Emergency assessment by clinical team with critical care competencies Consider transfer to level 2/3 care		

Weighted score	3	2	1	0	1	2	3
Respiratory rate	<8	-	9-11	12-20	-	21-24	>25
Oxygen saturation (NEWS only)	<91	92-93	94-95	>96	-	-	-
Supplemental oxygen	-	**NEWS-**Yes	-	**NEWS-**No **Notts-** 0-9L/min	**Notts-** 10-14L/min		**Notts-** ≥15L/min
Heart Rate	<40	-	41-50	51-90	91-110	111-130	≥131
Blood Pressure	<90	91-100	101-110	111-219	-	-	≥220
Temperature	<35	-	35.1-36	36.1-38	38.1-39	>39.1	-
Urine output (Notts only)	-	Not PU'd in 12 hours or <20 ml/hour	Not PU'd 6 hours or 20-30	PU'd in last 6 hours or 31-199ml/hour	>200ml/hr	-	-
AVPU	**Both-** Unresponsive	**Notts-** Pain	**Notts-** Voice	**Both-** Alert			

Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham University Hospitals Early Warning Score

105x143mm (300 x 300 DPI)

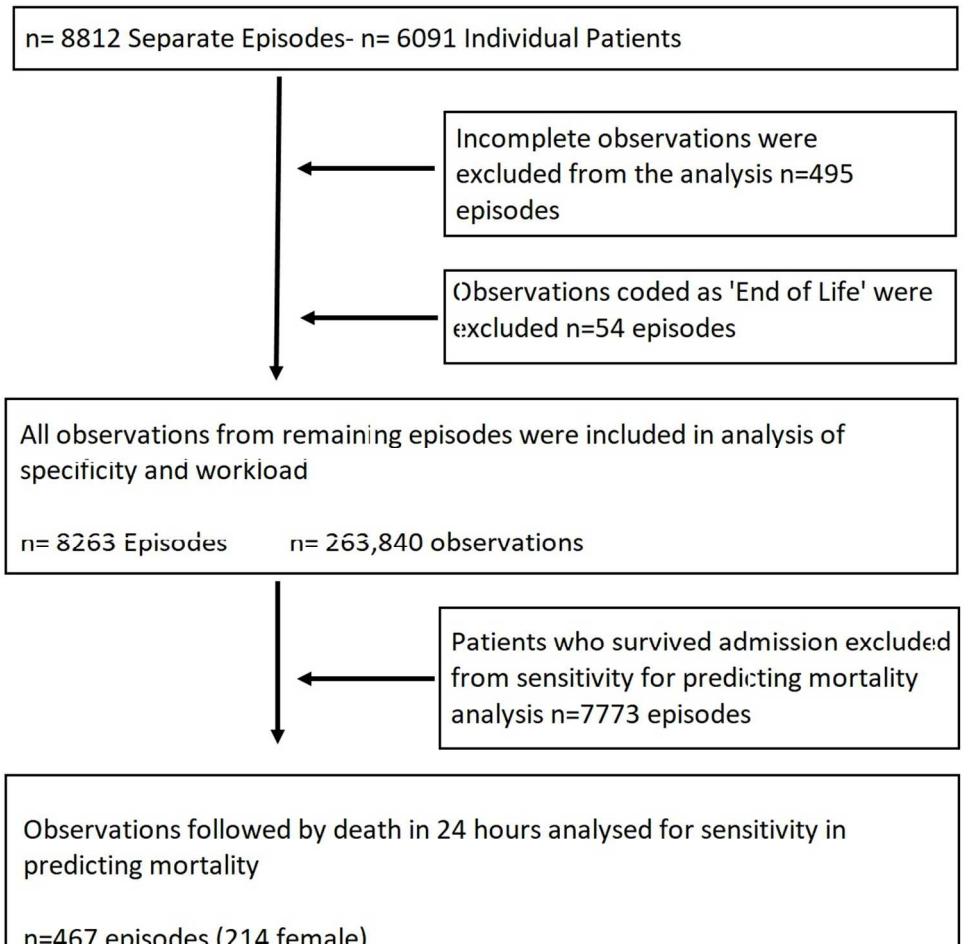

Figure 2. Cohort flow diagram of exclusion criteria

105x104mm (300 x 300 DPI)

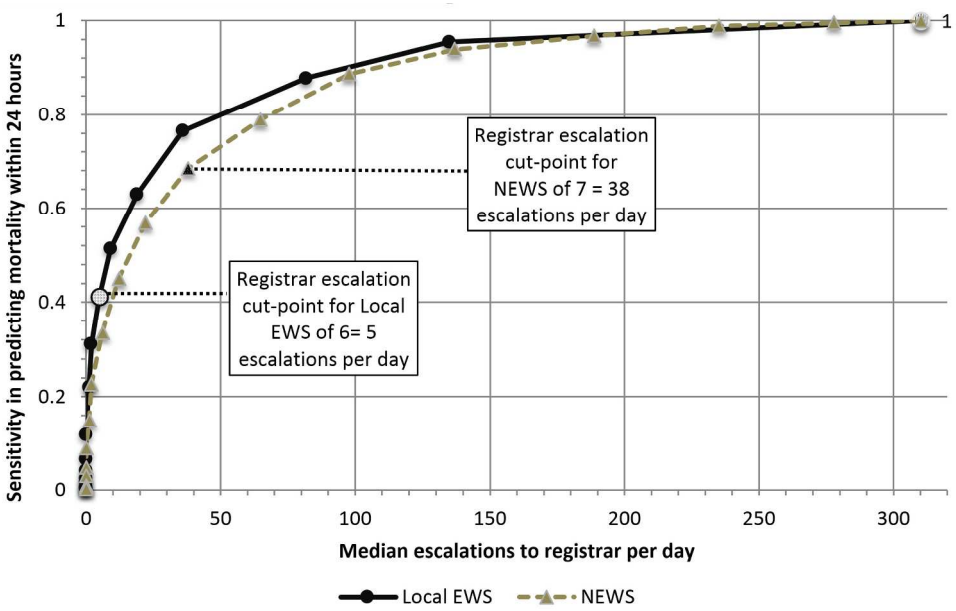

Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

206x130mm (300 x 300 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **[See title and design section of abstract pages 1-2]**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **[See abstract pages 1-2]**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **[See abstract and Background sections- pages 1-2]**
Objectives	3	State specific objectives, including any prespecified hypotheses **[See objective section in abstract]**
Methods		
Study design	4	Present key elements of study design early in the paper **[methods in abstract – expanded in methods section of main article on page 4]**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **[Methods and Setting in abstract, Methods page 4]**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **[Methods section page 4]**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **[N/A]**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **[Background, Methods, Discussion]**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **[Methods page 4]**
Bias	9	Describe any efforts to address potential sources of bias **[N/A]**
Study size	10	Explain how the study size was arrived at **[Methods page 4]**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **[Methods page 4]**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **[Methods page 4]**
		(*b*) Describe any methods used to examine subgroups and interactions **[Methods page 4]**
		(*c*) Explain how missing data were addressed **[Cohort Diagram- Figure 2, Methods page 4]**
		(*d*) If applicable, explain how loss to follow-up was addressed **[n/a]**
		(*e*) Describe any sensitivity analyses **[Methods page 4]**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **[Results page 5]**
		(b) Give reasons for non-participation at each stage **[N/A]**
		(c) Consider use of a flow diagram **[Cohort flow diagram in methods section]**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **[Table 1- characteristics of**

		study population]
		(b) Indicate number of participants with missing data for each variable of interest **[See cohort flow diagram- figure 2 in methods page 4]**
		(c) Summarise follow-up time (eg, average and total amount) **[N/A]**
Outcome data	15*	Report numbers of outcome events or summary measures over time **[Abstract and results]**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **[N/A]**
		(*b*) Report category boundaries when continuous variables were categorized **[See figure 1 + Table 2]**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **[n/a]**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **[Methods and results page 4-5]**
Discussion		
Key results	18	Summarise key results with reference to study objectives **[Discussion page 7 onwards and abstract]**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **[Strengths and weaknesses page 1; Discussion page 7 onwards]**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **[Background page 2 and Discussion page 7 onwards]**
Generalisability	21	Discuss the generalisability (external validity) of the study results **[Discussion page 7]**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **[Included after main body of text before references]**

*Give information separately for exposed and unexposed groups.