

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

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

Journal:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

Only

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

Sarah Forster^{1,2}, Gemma Housley³, Tricia M McKeever⁴, 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.

1
2
3 **Outcome Measures** We examined both the actual (for local EWS) and projected workload (for
4 NEWS) created by the scores, and the sensitivity and specificity of the scores in predicting mortality
5 based on outcome within 24 hours of a score generated by vital signs observations.
6
7

8 **Results** 8812 individual patient episodes occurred during the study period. Overall mortality was
9 5.9%. Applying NEWS retrospectively (versus local EWS) generated a significant eight fold increase in
10 mandatory escalations, but had a higher sensitivity for predicting mortality at the cut points applied
11 by the protocol.
12
13

14 **Conclusions** This study highlights the issues surrounding the use of EWS in patients with respiratory
15 disease. The higher sensitivity and lower specificity of NEWS means that it acts like a d-dimer; a low
16 score is useful in ruling out deterioration, a high score is less helpful in predicting mortality. Further
17 work on the significance of changes in vital signs in patients with complex underlying comorbidity
18 and greater understanding of the pathophysiology involved is needed.
19
20
21
22
23

24 **Strengths and Limitations of this Study**

- 25 ▪ Data were obtained from a large clinical vital signs database with clear identification of
26 specialty allowing for subgroup analysis.
- 27 ▪ Granularity of data collection in the database allowed for reliable identification of patients
28 meeting the exclusion criteria.
- 29 ▪ Only 0.2% of the observations recorded during the study period were identified as being
30 incomplete.
- 31 ▪ The retrospective nature of study precludes conclusions relating to impact of introducing
32 NEWS on mortality.
33
34
35
36
37
38
39
40

41 **Background**

42
43
44 In 2012 the Royal College of Physicians published the NEWS protocol in an attempt to standardise
45 processes for identifying patients at risk of imminent deterioration [1]. EWS protocols guide
46 decisions around patient care by mandating when a patient with evidence of pathophysiology, in the
47 form of deranged vital signs, should be reviewed by a clinical member of staff, and therefore
48 influence overall clinical workload and resource allocation for all in-patients. Patients with
49 respiratory disease make up a large proportion of a hospital's in-patient population, however a
50 previous small study found that chronic physiological disturbance caused by COPD may render NEWS
51
52
53
54
55
56
57
58
59
60

1
2
3 less discriminative than in an unselected medical population [2]; consequently attempts have been
4 made to improve the score in this population [3].
5

6
7 Nottingham University Hospitals NHS Trust employs an electronic observations system with
8 mandatory escalation based on an adapted EWS. We compared the potential impact in terms of
9 workload and mortality prediction of using a locally designed EWS versus NEWS (see Figure 1) in
10 patients with respiratory disease.
11
12

13
14 **Figure 1- Escalation protocol for NEWS and Nottingham University Hospitals Early**
15 **Warning Score**

16 **Methods**

17
18 We performed a single centre retrospective analysis of all patients admitted to the respiratory
19 department at Nottingham University Hospitals NHS Trust between 01/04/2015 and 31/03/2017.
20 This is a tertiary referral centre for respiratory medicine, with one specialist admissions ward and 3
21 inpatient wards. Data from the integrated electronic observation and communication system
22 comprising respiratory rate, oxygen saturations, heart rate, blood pressure, temperature, conscious
23 level (AVPU score), and urine output were analysed. The same system also automatically generates
24 mandated escalation and referral at set scoring thresholds via a pre-determined protocol. Scores
25 from the local EWS were linked to demographics and mortality outcomes. NEWS criteria were
26 applied retrospectively to determine how many patients would have been escalated if the NEWS
27 system were followed. Results were analysed using STATA 15. The entire data set was analysed for
28 measurement of escalation patterns, analysis of workload and sensitivity and specificity in predicting
29 death within 24 hours of an observation [4]. Observations coded as end of life care following clinical
30 decision were excluded from mortality analysis (see Figure 2).
31
32
33
34
35
36
37
38

39 **Figure 2. Cohort flow diagram of exclusion criteria**

40 **Results**

41
42 236,840 observation sets were recorded during 8812 inpatient episodes (53.1% female- see Table1)
43 involving 6091 individuals. In-hospital mortality for respiratory patients was 5.9% (n=521) and
44 median length of stay was 4 days (range 0-175).
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1
2
3 Figure 3 plots sensitivity in predicting mortality, against median number of mandated clinician alerts
4 per day for both EWS types. It demonstrates that for a sensitivity of 0.7, NEWS generates a higher
5 number of mandated escalations. At both extremes of sensitivity (0 and 1) the number of escalations
6 is the same: i.e. Mandating an escalation at a NEWS or EWS of 0 would mean all patients were
7 escalated, and each score would have 100% sensitivity for predicating mortality (as everyone who
8 died would have been reviewed). Likewise only escalating patients with a maximum EWS or NEWS
9 score would lead to very few patients being escalated.
10
11
12
13

14
15 **Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS**

16 **Discussion**

17
18 In this study we examined the effect of two different EWS systems in patients admitted with
19 respiratory disease. We analysed the number of mandatory escalations generated and the
20 sensitivity and specificity of both scores in predicting imminent in-hospital mortality. Our data shows
21 that at the scores' cut points for escalation, NEWS would have generated a significantly higher
22 workload due to a lower specificity, with a higher sensitivity for predicting imminent deterioration,
23 when compared with the locally used EWS.
24
25
26

27
28 Although previous work suggested that NEWS was worse at predicting deterioration in patients with
29 respiratory disease, compared to a population of unselected medical admissions [2], NEWS has not
30 been studied in large numbers of respiratory patients across an entire admission.
31
32

33
34 Our study faced similar limitations to others, namely the low prevalence of mortality in the patient
35 population and the difficulty of studying patients in real time. However, our observed findings of an
36 increase workload generated are both novel and important as, when used as part of a system which
37 employs automatic escalation of threshold scores, NEWS leads to a significant impact on work load
38 in a resource pressured environment, with little evidence of improved clinical outcome.
39
40
41

42
43 Although there is a difference in the workload generated by both scoring systems, this relates to the
44 cut points for escalation mandated by the protocols, rather than the scores themselves;
45 unsurprisingly overall both scores perform similarly (they are based on similar clinical observations)
46 however the mandated cut points differ. The difference in protocol design relates to the way in
47 which the scores are used clinically, and can be explained as follows:
48
49

50
51 The first approach is seen in the scoring thresholds dictated by NEWS. Its cut points for each layer of
52 clinical intervention, i.e. escalation to nurse, clinician or registrar, have a higher sensitivity which acts
53 to rule out imminent clinical deterioration in those patients whose vital signs do not meet scoring
54 thresholds, meaning clinicians can be confident that patients with a low score are very unlikely to be
55
56
57
58
59
60

1
2
3 at imminent risk. This is akin to a d-dimer where a low value in an individual with low clinical
4 suspicion effectively excludes a venous thromboembolism. This approach works well in a setting
5 with less highly trained staff delivering the first layer of monitoring. However, if this approach is
6 applied in an unfiltered manner, such as can exist with systems employing electronic monitoring
7 with automatic mandatory escalation, the workload generated by escalations from patients who
8 never go on to deteriorate has significant resource and operational implications, as well as increased
9 likelihood of unnecessary intervention for patients. Again this is analogous to using a d-dimer which
10 has been demonstrated to have a positive predictive value of approximately 20% in going on to find
11 radiological evidence of a venous thromboembolism [5].
12
13
14
15
16

17
18 The second approach, used by the local EWS, is one of high specificity in the cut points for
19 escalation, with a relatively lower sensitivity. This approach acts to highlight potential imminent
20 clinical deterioration in those meeting the escalation criteria, but does not always rule out
21 deterioration in those who score under the cut point. This may seem a less preferable approach,
22 however a recent study of rapid response systems indicated that staff clinical concern in the absence
23 of a qualifying score was responsible for escalation in 47% of calls [6], highlighting the role of staff
24 education and empowerment, over EWS protocols. It may also mean that in resource limited
25 environments (such as during out of hours care) if patients are highlighted as needing intervention,
26 staff are more likely to be available to intervene as they are not being used to review patients who
27 are triggering a review but clinically stable.
28
29
30
31
32
33

34 Despite the mandated and widespread uptake of EWS, there has been minimal prospective
35 validation of their use. Efforts to improve precision in predicting outcome through scrutiny of large
36 datasets has largely employed analyses utilising area under the receiver operating characteristic
37 curves which is limited by the low prevalence of mortality in the population [7]. Before and after
38 studies have largely, but not universally [8-10] highlighted the efficacy of EWS, however no
39 randomised controlled trials have been performed. Consequently evidence of the scores' real impact
40 on clinical outcomes, such as mortality, transfer to higher level of care or length of stay, or on
41 workforce outcomes such as workload through excessive task generation and alarm fatigue, has only
42 been obtained from observational studies. These are all limited by significant confounders.
43
44
45
46
47
48

49 This evidence gap around the clinical and workforce implications of EWS systems will become
50 increasingly important as hospitals move towards automated systems with mandated referral of
51 patients who reach a threshold score. Continuing integration of more data into digital healthcare
52 systems via continuous monitoring, dynamic measures of fitness, and electronic health records will
53 further highlight this gap, as without an understanding of how these data can be applied it will be
54
55
56
57
58
59
60

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	92-95	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
0	Min 12 hourly	Continue routine NEWS monitoring
1	Min 4-6 hourly	Inform RN who must assess patient RN to decide re. increased monitoring or escalation
2		
3		
4	Min 1 hourly	RN to urgently inform medical team Urgent assessment by clinician with core competencies Clinical care in environment with monitoring facilities
5		
6	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
>7		

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

121x155mm (96 x 96 DPI)

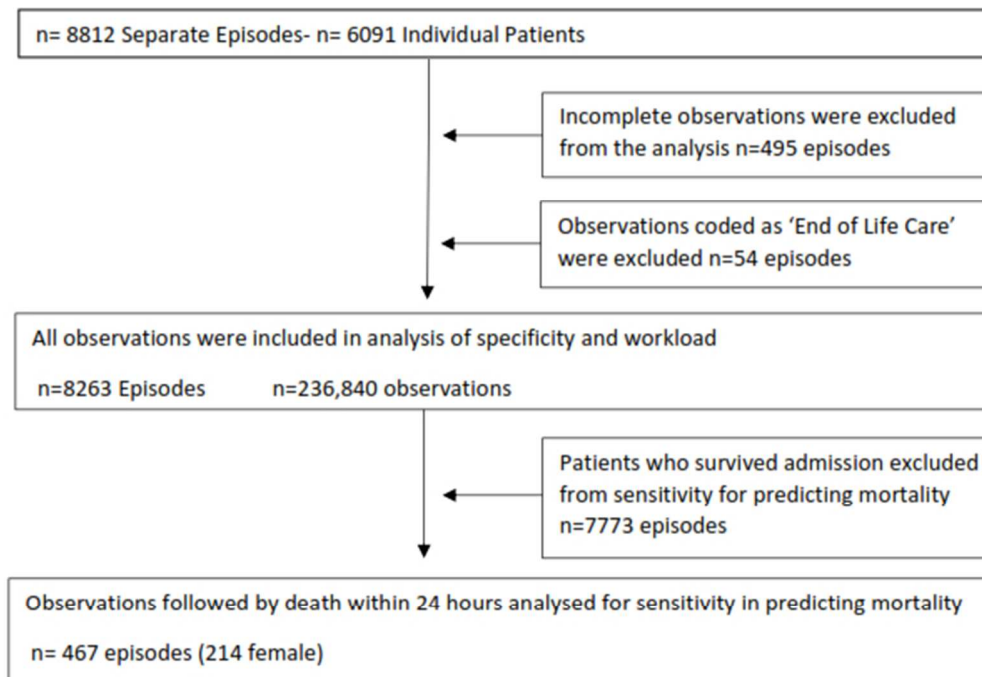


Figure 2. Cohort flow diagram of exclusion criteria

158x112mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

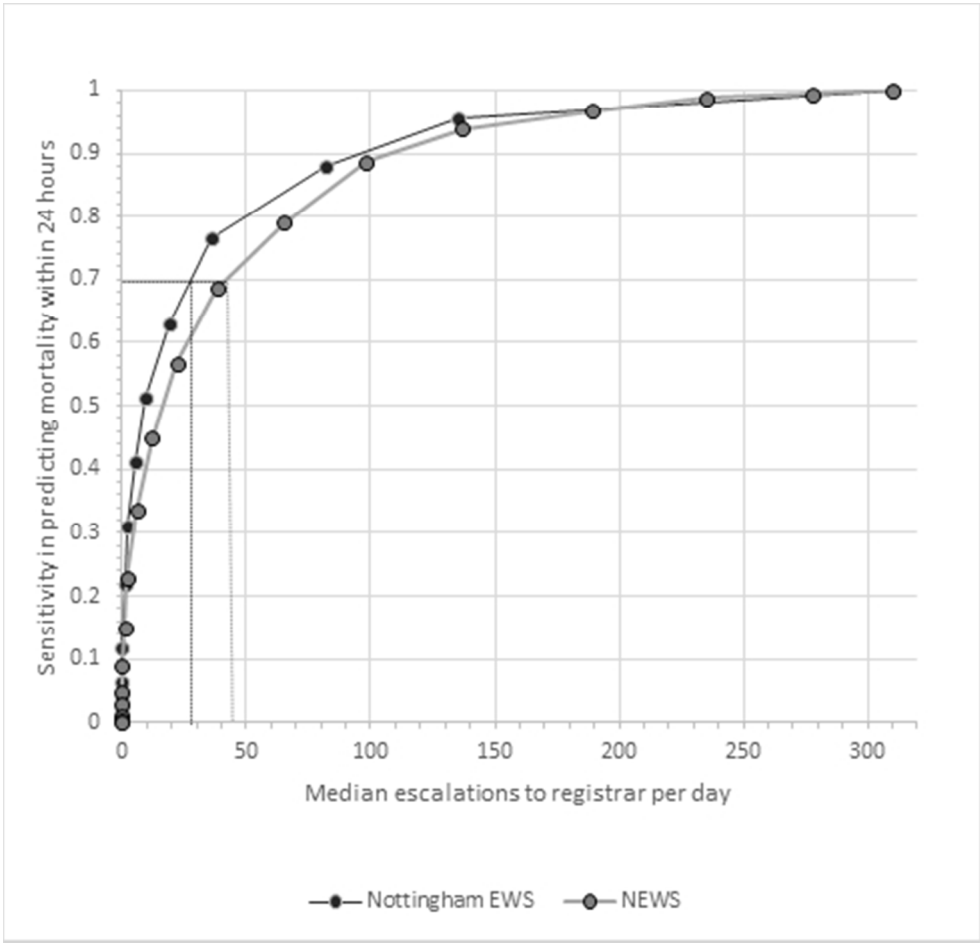


Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

129x124mm (96 x 96 DPI)

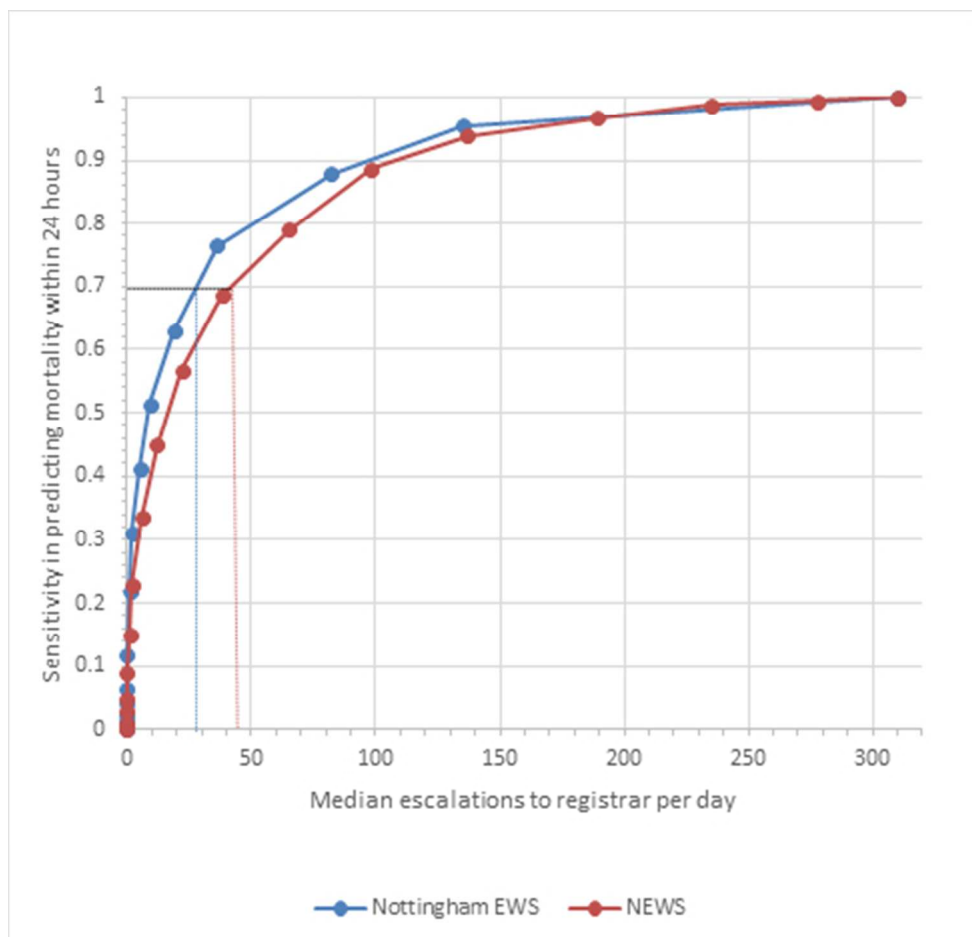


Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4
5 **Investigating the discriminative value of Early Warning Scores in**
6 **patients with respiratory disease using a retrospective cohort**
7 **analysis of admissions to a tertiary respiratory referrals centre over**
8 **a 2 year period.**
9
10
11
12
13

14 Sarah Forster^{1,2}, Gemma Housley³, Tricia M McKeever⁴, Dominick E Shaw^{2,3}
15
16
17
18

19 **Author Affiliations**
20

21 ¹ NIHR Academic Clinical Fellow, University of Nottingham, Nottingham, UK
22

23 ² Respiratory Research Unit, Division of Respiratory Medicine, University of Nottingham, Nottingham, UK
24

25 ³ Medical Informatics, East Midlands Academic Health Sciences Network, Nottingham, UK
26

27 ⁴ Division of Epidemiology, University of Nottingham, Nottingham, UK
28
29

30 Correspondence to Dr Sarah Forster, sarahforster@nhs.net
31
32
33

34 Word count: 2976
35
36
37
38

39 **Abstract**
40
41
42

43 **Objective**
44

45 Early Warning Scores (EWS) are used to monitor patients for signs of imminent deterioration.
46 Although used in respiratory disease, EWS have not been well studied in this population, despite the
47 underlying cardiopulmonary pathophysiology often present. We examined the performance of two
48 scoring systems in patients with respiratory disease.
49
50
51

52 **Design**
53
54
55
56
57

1
2
3 Retrospective cohort analysis of vital signs observations of all patients admitted to a respiratory unit
4 over a 2 year period. Scores were linked to outcome data to establish the performance of the
5 National EWS (NEWS) compared results to a locally adapted EWS.
6
7

8 **Setting**

9
10 Nottingham University Hospitals NHS Trust respiratory wards. Data were collected from an
11 integrated electronic observation and task allocation system employing a local EWS, also generating
12 mandatory referrals to clinical staff at set scoring thresholds.
13
14
15

16 **Outcome Measures**

17
18 Projected workload, and sensitivity and specificity of the scores in predicting mortality based on
19 outcome within 24 hours of a score being recorded.
20
21

22 **Results**

23
24 8812 individual patient episodes occurred during the study period. Overall mortality was 5.9%.
25 Applying NEWS retrospectively (versus local EWS) generated an eight fold increase in mandatory
26 escalations, but had higher sensitivity in predicting mortality at the protocol cut points.
27
28
29

30 **Conclusions**

31
32 This study highlights issues surrounding use of scoring systems in patients with respiratory disease.
33 NEWS demonstrated higher sensitivity for predicting death within 24 hours, offset by reduced
34 specificity. The consequent workload generated may compromise the ability of the clinical team to
35 respond to patients needing immediate input. The locally adapted EWS has higher specificity but
36 lower sensitivity. Statistical evaluation suggests this may lead to missed opportunities for
37 intervention, however this does not account for clinical concern independent of the scores, nor
38 ability to respond to alerts based on workload. Further research into the role of warning scores and
39 the impact of chronic pathophysiology is urgently needed.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 discriminative when compared to an unselected medical population [2]. This has significant
4 implications for patients, in terms of increased observations and interventions, and to clinical staff in
5 terms of workload and potential for alert fatigue. Consequently attempts have been made to
6 improve the score in this population [3].
7
8

9
10 Nottingham University Hospitals NHS Trust employs an electronic observations system with
11 mandatory escalation based on an adapted EWS. The Nottingham EWS, unlike NEWS, does not
12 score oxygen saturations and has a graduated approach to weighting for both oxygen delivery and
13 level of consciousness. As a more general marker of morbidity it also employs urine output. We
14 compared the sensitivity and specificity of the two scores in predicting mortality within 24 hours of a
15 set of observations being recorded at the clinical cut points determined by the associated protocols
16 and examined the potential impact in terms of workload of using the locally designed EWS versus
17 NEWS (see Figure 1) in patients with respiratory disease based on analysis of the vital signs
18 observations and outcomes of patients admitted to the respiratory department in Nottingham over
19 a 2 year period. We then went on to answer the same questions in a subgroup of patients who were
20 admitted with a diagnosis of COPD to examine the performance of the two scores in this cohort.
21
22
23
24
25
26
27
28
29

30 **Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham**
31 **University Hospitals Early Warning Score**
32

33 **Methods**

34
35 We performed a single centre retrospective analysis of all patients admitted to the respiratory
36 department at Nottingham University Hospitals NHS Trust between 01/04/2015 and 31/03/2017.
37 This is a tertiary referral centre for respiratory medicine, with one specialist admissions ward and 3
38 inpatient wards. The analysis included all adults admitted with respiratory disease not transferred to
39 a higher level of care, i.e. high dependency or intensive care, greater than 24 hours before death as
40 these areas are not currently employing electronic observations, long term ventilator dependent
41 patients were also excluded as hospital policy dictates that these patients are always admitted to the
42 high dependency unit. Following approval from the NHS Information Governance Lead, and in line
43 with existing permissions within the East Midlands Academic Health Sciences network, data from the
44 integrated electronic observation and communication system comprising respiratory rate, oxygen
45 saturations, heart rate, blood pressure, temperature, conscious level (AVPU score), and urine output
46 were anonymised by an NHS data analyst prior to extraction from the clinical server. The same
47 system also automatically generates mandated escalation and referral at set scoring thresholds via a
48 pre-determined protocol. Scores from the local EWS were linked to demographics and mortality
49
50
51
52
53
54
55
56
57
58

1
2
3 outcomes prior to extraction. NEWS criteria were applied retrospectively to determine how many
4 patients would have been escalated if the NEWS system were followed. Results were analysed using
5 STATA 15. The entire data set was analysed for measurement of escalation patterns, analysis of
6 workload and sensitivity and specificity in predicting death within 24 hours of an observation [4]. A
7 chi-square analysis was performed to demonstrate whether the difference in escalations was
8 significant. The statistical analysis involved the use of all vital signs observations recorded
9 throughout admission, which were linked to outcome to determine whether they were followed by
10 death within 24 hours of the observation timestamp created by the input devices at the bedside.
11 Observations coded as end of life care following clinical decision were excluded from mortality
12 analysis (see Figure 2). A further subgroup analysis was then performed on patients coded as having
13 chronic obstructive pulmonary disease at any point in their admission as per ICD10 codes (see Figure
14 3) in order to further assess the statistical performance of the two scores in the presence of chronic
15 pathophysiology.
16
17
18
19
20
21
22
23
24
25

26 **Figure 2. Cohort flow diagram of exclusion criteria**

30 **Results**

31
32 236,840 observation sets were recorded during 8812 inpatient episodes (53.1% female- see Table1)
33 involving 6091 individuals. In-hospital mortality for respiratory patients was 5.9% (n=521) and
34 median length of stay was 4 days (range 0-175).
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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	
Heart Rate (beats per minute)	Mean (+/- SD)
	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-

1
2
3 158) scores generating automatic referral to the registrar ($p < 0.001$ for difference between scores)
4 per day.
5

6
7 Sensitivity and specificity for predicting in-hospital mortality based on death within 24 hours of a set
8 of vital signs observations point is shown in Table 2. At each clinically equivalent band, the sensitivity
9 and specificity in predicting mortality of all patients scoring at and above that cut point are shown.
10 At each cut point, NEWS would have had a higher sensitivity than the local EWS (i.e. a higher
11 percentage of patients who went on to die were flagged as requiring escalation), but a lower
12 specificity.
13
14
15

16
17 Figure 3 plots sensitivity in predicting mortality, against median number of mandated clinician alerts
18 per day for both EWS types. It demonstrates that for a sensitivity of 0.7, NEWS generates a higher
19 number of mandated escalations. At both extremes of sensitivity (0 and 1) the number of escalations
20 is the same: i.e. mandating an escalation at a NEWS or EWS of 0 would mean all patients were
21 escalated, and each score would have 100% sensitivity for predicting mortality (as everyone who
22 died would have been reviewed). Likewise only escalating patients with a maximum EWS or NEWS
23 score would lead to very few patients being escalated.
24
25
26
27
28
29
30
31
32

33 **Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS**
34

35
36 Further subgroup analysis was performed on admissions with an ICD10 code for COPD at any point.
37 This yielded 56,345 observations from 2207 episodes by 1365 individual patients. Using the local
38 EWS protocol led to median of 0 (range 0-19) escalations to the registrar, while applying NEWS
39 would have generated a median of 6 (0-47) scores being escalated to the registrar each day. As in
40 the unselected respiratory cohort, NEWS was more sensitive in predicting imminent mortality than
41 the local EWS but with a significantly inferior specificity at each clinical cut point applied (see table
42 3).
43
44
45
46

47 Discussion

48
49
50

51 In this study we examined the effect of two different early warning score systems in patients
52 admitted with respiratory disease. We analysed the number of mandatory escalations generated
53 and the sensitivity and specificity of both scores in predicting imminent in-hospital mortality in an
54 unselected respiratory population and in a subgroup analysis of patients with COPD. Our data shows
55
56
57

1
2
3 that at the scores' cut points for escalation, NEWS would have generated a significantly higher
4 workload due to a lower specificity, with a higher sensitivity for predicting imminent deterioration,
5 when compared with the locally used EWS. This was accentuated in patients with COPD, an
6 observation we believe is due to chronic changes in the underlying physiology which influences the
7 way in which these patients respond to acute pathological processes.
8
9

10
11 Although previous work suggested that NEWS was less discriminative in predicting deterioration in
12 patients with respiratory disease, compared to a population of unselected medical admissions [2],
13 NEWS has not been studied in large numbers of respiratory patients across an entire admission.
14
15

16
17 Our study faced similar limitations to others, namely the low prevalence of mortality in the patient
18 population and the difficulty of studying patients in real time. However, our observed findings of an
19 increase workload generated are both novel and important as, when used as part of a system which
20 employs automatic escalation of threshold scores, NEWS leads to a significant impact on work load
21 in a resource pressured environment, with little evidence of improved clinical outcome.
22
23

24
25 Although there is a difference in the workload generated when comparing the scoring systems both
26 in a general respiratory population and in patients with COPD, this relates to the cut points for
27 escalation mandated by the protocols, rather than the scores themselves; unsurprisingly overall
28 both scores perform similarly when the individual scores are plotted (they are based on similar
29 clinical observations) however the mandated cut points differ. The difference created by the
30 protocol design relates to the way in which the scores are used clinically, and can be explained as
31 follows:
32
33
34
35

36
37 The first approach is seen in the scoring thresholds dictated by NEWS. Its cut points for each layer of
38 clinical intervention, i.e. escalation to nurse, clinician or registrar, have a higher sensitivity which acts
39 to rule out imminent clinical deterioration in those patients whose vital signs do not meet scoring
40 thresholds, meaning clinicians can be confident that patients with a low score are very unlikely to be
41 at imminent risk. This is akin to a d-dimer where a low value in an individual with low clinical
42 suspicion effectively excludes a venous thromboembolism [5, 6] . This high sensitivity approach
43 works well in a setting with less highly trained staff delivering the first layer of monitoring. However,
44 if this approach is applied in an unfiltered and automated manner, the workload generated by
45 escalations from patients who never go on to deteriorate will have significant resource and
46 operational implications, as well as increasing the likelihood of unnecessary intervention for
47 patients.
48
49
50
51
52
53
54
55
56
57

1
2
3 The second approach, used by the local EWS, is one of high specificity in the cut points for
4 escalation, with a relatively lower sensitivity. This approach acts to highlight potential imminent
5 clinical deterioration in those meeting the escalation criteria, but does not always rule out
6 deterioration in those who score under the cut point. This may seem a less preferable approach.
7 However a recent study of rapid response systems indicated that staff clinical concern in the
8 absence of a qualifying score was responsible for escalation in 47% of calls [7], highlighting the role
9 of staff education and empowerment, over and above EWS protocols The variability in physiological
10 normal baselines created by patient specific factors such as comorbidity or fitness means that using
11 vital signs observations alone as the basis for a score leading to mandatory escalation will always
12 require a trade-off between sensitivity in accurately identifying patients potentially at risk of
13 deterioration and staff alarm fatigue generated by patients who do not go onto deteriorate. This is
14 particularly pertinent in resource limited environments (such as during out of hours care),

15
16
17
18
19
20
21
22 Despite the mandated and widespread uptake of EWS, there has been minimal prospective
23 validation of their use. Efforts to improve precision in predicting outcome through scrutiny of large
24 datasets has largely employed analyses utilising area under the receiver operating characteristic
25 curves which are limited by the low prevalence of mortality in the population [8]. Before and after
26 studies have largely, but not universally [9-11] highlighted the efficacy of EWS, however no
27 randomised controlled trials have been performed. Consequently evidence of the scores' real impact
28 on clinical outcomes, such as mortality, transfer to higher level of care or length of stay, or on
29 workforce outcomes such as workload from excessive task generation and alarm fatigue, has only
30 been obtained from observational studies. These are all limited by significant confounders.

31
32
33
34
35
36
37
38 This evidence gap around the clinical and workforce implications of EWS systems will become
39 increasingly important as hospitals move towards automated systems with mandated referral of
40 patients who reach a threshold score. Continuing integration of more data into digital healthcare
41 systems via continuous monitoring, dynamic measures of fitness, and electronic health records will
42 further highlight this gap, as without an understanding of how these data can be applied it will be
43 difficult to differentiate the signal from the noise. Given the growing complexity of the inpatient
44 population more work is urgently required to understand the wider impact of EWS on outcomes
45 such as mortality and length of stay, task burden, working patterns and cost. There is also need to
46 reconsider the role of clinical concern in monitoring patients and how this can be further promoted
47 to prevent future systems depending purely on scores rather than integrating staff skills and
48 intuition into the decision making process. Early warning scores should not be developed in isolation
49 based on statistical performance as this fails to recognise that they are a component within the
50 complex clinical environment and therefore need to be designed to enhance, not complicate, the

1
2
3 clinical decision making process. This is particularly important in patients with respiratory disease
4 where physiology is often chronically deranged and less responsive to intervention and a greater
5 understanding of the contributory clinical factors and more individualised approach is required.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

For peer review only

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			Hourly for minimum 2 hours Consider fluid balance monitoring	RN to recheck Inform NIC Manual BP if deranged
3				
4	Min 1 hourly	RN to urgently inform medical team Urgent assessment by clinician with core competencies Clinical care in environment with monitoring facilities	Hourly with fluid balance monitoring	RN to check EWS F1/SHO to discuss management with SpR or consultant
5				
6				
>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	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

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)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

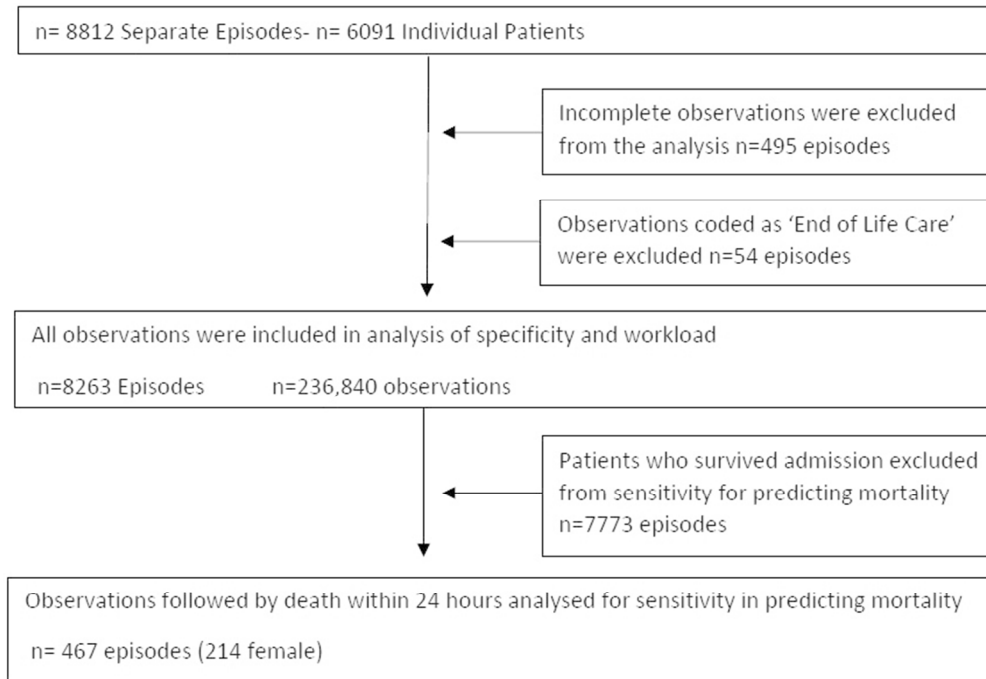


Figure 2. Cohort flow diagram of exclusion criteria

72x49mm (300 x 300 DPI)

ew only

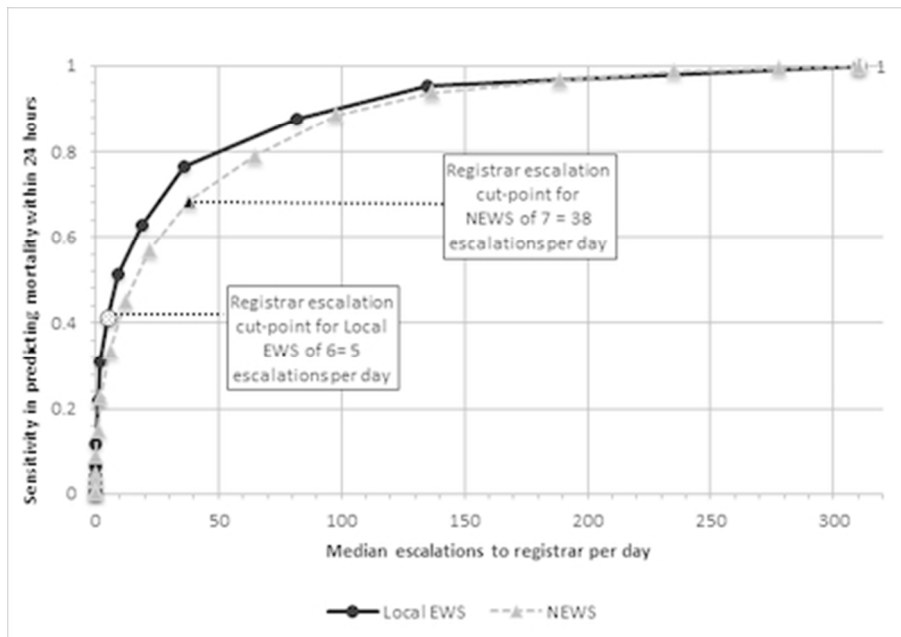


Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

38x29mm (300 x 300 DPI)

view only

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

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 Housley³, Tricia M McKeever⁴, 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

1
2
3 Retrospective cohort analysis of vital signs observations of all patients admitted to a respiratory unit
4 over a 2 year period. Scores were linked to outcome data to establish the performance of the
5 National EWS (NEWS) compared results to a locally adapted EWS.
6
7

8 **Setting**

9
10 Nottingham University Hospitals NHS Trust respiratory wards. Data were collected from an
11 integrated electronic observation and task allocation system employing a local EWS, also generating
12 mandatory referrals to clinical staff at set scoring thresholds.
13
14
15

16 **Outcome Measures**

17
18 Projected workload, and sensitivity and specificity of the scores in predicting mortality based on
19 outcome within 24 hours of a score being recorded.
20
21

22 **Results**

23
24 8812 individual patient episodes occurred during the study period. Overall mortality was 5.9%.
25 Applying NEWS retrospectively (versus local EWS) generated an eight fold increase in mandatory
26 escalations, but had higher sensitivity in predicting mortality at the protocol cut points.
27
28
29

30 **Conclusions**

31
32 This study highlights issues surrounding use of scoring systems in patients with respiratory disease.
33 NEWS demonstrated higher sensitivity for predicting death within 24 hours, offset by reduced
34 specificity. The consequent workload generated may compromise the ability of the clinical team to
35 respond to patients needing immediate input. The locally adapted EWS has higher specificity but
36 lower sensitivity. Statistical evaluation suggests this may lead to missed opportunities for
37 intervention, however this does not account for clinical concern independent of the scores, nor
38 ability to respond to alerts based on workload. Further research into the role of warning scores and
39 the impact of chronic pathophysiology is urgently needed.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

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

1
2
3 discriminative when compared to an unselected medical population [2]. This has significant
4 implications for patients, in terms of increased observations and interventions, and to clinical staff in
5 terms of workload and potential for alert fatigue. Consequently attempts have been made to
6 improve the score in this population [3].
7
8

9
10 Nottingham University Hospitals NHS Trust employs an electronic observations system with
11 mandatory escalation based on an adapted EWS. The Nottingham EWS, unlike NEWS, does not
12 score oxygen saturations and has a graduated approach to weighting for both oxygen delivery and
13 level of consciousness. As a more general marker of morbidity it also employs urine output. We
14 compared the sensitivity and specificity of the two scores in predicting mortality within 24 hours of a
15 set of observations being recorded at the clinical cut points determined by the associated protocols
16 and examined the potential impact in terms of workload of using the locally designed EWS versus
17 NEWS (see Figure 1) in patients with respiratory disease based on analysis of the vital signs
18 observations and outcomes of patients admitted to the respiratory department in Nottingham over
19 a 2 year period. We then went on to answer the same questions in a subgroup of patients who were
20 admitted with a diagnosis of COPD to examine the performance of the two scores in this cohort.
21
22
23
24
25
26
27
28
29

30 **Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham**
31 **University Hospitals Early Warning Score**
32

33 **Methods**

34
35 We performed a single centre retrospective analysis of all patients admitted to the respiratory
36 department at Nottingham University Hospitals NHS Trust between 01/04/2015 and 31/03/2017.
37 This is a tertiary referral centre for respiratory medicine, with one specialist admissions ward and 3
38 inpatient wards. The analysis included all adults admitted with respiratory disease not transferred to
39 a higher level of care, i.e. high dependency or intensive care, greater than 24 hours before death as
40 these areas are not currently employing electronic observations, long term ventilator dependent
41 patients were also excluded as hospital policy dictates that these patients are always admitted to the
42 high dependency unit. Following approval from the NHS Information Governance Lead, and in line
43 with existing permissions within the East Midlands Academic Health Sciences network, data from the
44 integrated electronic observation and communication system comprising respiratory rate, oxygen
45 saturations, heart rate, blood pressure, temperature, conscious level (AVPU score), and urine output
46 were anonymised by an NHS data analyst prior to extraction from the clinical server. The same
47 system also automatically generates mandated escalation and referral at set scoring thresholds via a
48 pre-determined protocol. Scores from the local EWS were linked to demographics and mortality
49
50
51
52
53
54
55
56
57

1
2
3 outcomes prior to extraction. NEWS criteria were applied retrospectively to determine how many
4 patients would have been escalated if the NEWS system were followed. Results were analysed using
5 STATA 15. The entire data set was analysed for measurement of escalation patterns, analysis of
6 workload and sensitivity and specificity in predicting death within 24 hours of an observation [4]. A
7 chi-square analysis was performed to demonstrate whether the difference in escalations was
8 significant. The statistical analysis involved the use of all vital signs observations recorded
9 throughout admission, which were linked to outcome to determine whether they were followed by
10 death within 24 hours of the observation timestamp created by the input devices at the bedside.
11 Observations coded as end of life care following clinical decision were excluded from mortality
12 analysis (see Figure 2). A further subgroup analysis was then performed on patients coded as having
13 chronic obstructive pulmonary disease at any point in their admission as per ICD10 codes in order to
14 further assess the statistical performance of the two scores in the presence of chronic
15 pathophysiology.
16
17
18
19
20
21
22
23
24
25

26 **Figure 2. Cohort flow diagram of exclusion criteria**

27 **Patient and public involvement**

28
29
30
31 Prior to carrying out this work, a questionnaire was performed amongst stakeholders, in this case 26
32 medical registrars working in the East Midlands region. All worked in acute trusts that employed
33 either NEWS or the Nottingham EWS as part of a system to highlight patients felt to be at risk of
34 deterioration. Of the stakeholder responders, 70% believed that using EWS failed to highlight all
35 patients who went on to deteriorate and 88% felt that use of an EWS led to unnecessary reviews. All
36 responders felt there were issues in the setting of chronic disease with some chronic patients scoring
37 even at baseline, and 76% felt that alert fatigue due to high EWS was an issue. These findings guided
38 the interrogation of the data in creating the study detailed in this paper. It is also worth noting that
39 similar work presented to patients with recent inpatient experience at NUHT highlighted the belief
40 that sleep was too often interrupted by observations or reviews.
41
42
43
44
45
46
47

48 **Results**

49
50 236,840 observation sets were recorded during 8812 inpatient episodes (53.1% female- see Table1)
51 involving 6091 individuals. In-hospital mortality for respiratory patients was 5.9% (n=521) and
52 median length of stay was 4 days (range 0-175).
53
54
55
56
57

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	
Heart Rate (beats per minute)	Mean (+/- SD)
Respiratory rate (breaths per minute)	87 (16)
Systolic BP (mmHg)	19 (3)
Temperature (°C)	130 (22)
Oxygen saturations (%)	36.6 (1)
	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-

1
2
3 158) scores generating automatic referral to the registrar ($p < 0.001$ for difference between scores)
4 per day.
5

6
7 Sensitivity and specificity for predicting in-hospital mortality based on death within 24 hours of a set
8 of vital signs observations point is shown in Table 2. At each clinically equivalent band, the sensitivity
9 and specificity in predicting mortality of all patients scoring at and above that cut point are shown.
10 At each cut point, NEWS would have had a higher sensitivity than the local EWS (i.e. a higher
11 percentage of patients who went on to die were flagged as requiring escalation), but a lower
12 specificity.
13
14
15

16
17 Figure 3 plots sensitivity in predicting mortality, against median number of mandated clinician alerts
18 per day for both EWS types. It demonstrates that for a sensitivity of 0.7, NEWS generates a higher
19 number of mandated escalations. At both extremes of sensitivity (0 and 1) the number of escalations
20 is the same: i.e. mandating an escalation at a NEWS or EWS of 0 would mean all patients were
21 escalated, and each score would have 100% sensitivity for predicting mortality (as everyone who
22 died would have been reviewed). Likewise only escalating patients with a maximum EWS or NEWS
23 score would lead to very few patients being escalated.
24
25
26
27
28
29
30
31
32

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

34
35 Further subgroup analysis was performed on admissions with an ICD10 code for COPD at any point.
36 This yielded 56,345 observations from 2207 episodes by 1365 individual patients. Using the local
37 EWS protocol led to median of 0 (range 0-19) escalations to the registrar, while applying NEWS
38 would have generated a median of 6 (0-47) scores being escalated to the registrar each day. As in
39 the unselected respiratory cohort, NEWS was more sensitive in predicting imminent mortality than
40 the local EWS but with a significantly inferior specificity at each clinical cut point applied (see table
41 3).
42
43
44
45
46

47 Discussion

48
49
50
51 In this study we examined the effect of two different early warning score systems in patients
52 admitted with respiratory disease to a tertiary referrals centre. The respiratory department at NUHT
53 manages patients in line with national guidelines and has outcomes comparable with other similar
54
55
56
57

1
2
3 units; consequently linking of raw observations to outcomes prior to analysis enables conclusions
4 which are applicable to other centres.
5

6
7 We analysed the number of mandatory escalations generated and the sensitivity and specificity of
8 both of the scores in predicting imminent in-hospital mortality in an unselected respiratory
9 population and in a subgroup analysis of patients with COPD. Our data shows that at the scores' cut
10 points for escalation, NEWS would have generated a significantly higher workload due to a lower
11 specificity, with a higher sensitivity for predicting imminent deterioration, when compared with the
12 locally used EWS. This was accentuated in patients with COPD, an observation we believe is due to
13 chronic changes in the underlying physiology which influences the way in which these patients
14 respond to acute pathological processes.
15
16
17
18

19
20 Although NEWS may become less relevant with the publication of NEWS2 in December 2017, our
21 study remains relevant. Firstly, it highlights the wider impact of the different approaches to
22 designing a scoring system and the paucity of evidence in relation to how this is evaluated. Secondly,
23 as it is currently unclear how widely NEWS2 has been adopted by hospitals across the NHS and what
24 the likely roll out will be, NEWS remains a current clinical tool in many trusts.
25
26
27

28
29 Previous work has suggested that NEWS was less discriminative in predicting deterioration in
30 patients with respiratory disease, compared to a population of unselected medical admissions [2],
31 however NEWS has not previously been studied in large numbers of respiratory patients across an
32 entire admission. Our study faced similar limitations to others, namely the low prevalence of
33 mortality in the patient population and the difficulty of studying patients in real time. However, our
34 observed findings of an increase workload generated are both novel and important as, when used as
35 part of a system which employs automatic escalation of threshold scores, NEWS leads to a significant
36 impact on work load in a resource pressured environment, with little evidence of improved clinical
37 outcome. This While there is a difference in the workload generated when comparing the scoring
38 systems both in a general respiratory population and in patients with COPD, this relates to the cut
39 points for escalation mandated by the protocols, rather than the scores themselves; unsurprisingly
40 overall both scores perform similarly when the individual scores are plotted (they are based on
41 similar clinical observations) however the mandated cut points differ. The difference created by the
42 protocol design relates to the way in which the scores are used clinically, and can be explained as
43 follows:
44
45
46
47
48
49
50
51
52

53 The first approach is seen in the scoring thresholds dictated by NEWS. Its cut points for each layer of
54 clinical intervention, i.e. escalation to nurse, clinician or registrar, have a higher sensitivity which acts
55 to rule out imminent clinical deterioration in those patients whose vital signs do not meet scoring
56
57

58 8
59

1
2
3 thresholds, meaning clinicians can be confident that patients with a low score are very unlikely to be
4 at imminent risk. This is akin to a d-dimer where a low value in an individual with low clinical
5 suspicion effectively excludes a venous thromboembolism [5, 6] . This high sensitivity approach
6 works well in a setting with less highly trained staff delivering the first layer of monitoring. However,
7 if this approach is applied in an unfiltered and automated manner, the workload generated by
8 escalations from patients who never go on to deteriorate will have significant resource and
9 operational implications, as well as increasing the likelihood of unnecessary intervention for
10 patients.
11
12
13
14
15

16 The second approach, used by the local EWS, is one of high specificity in the cut points for
17 escalation, with a relatively lower sensitivity. This approach acts to highlight potential imminent
18 clinical deterioration in those meeting the escalation criteria, but does not always rule out
19 deterioration in those who score under the cut point. This may seem a less preferable approach.
20 However a recent study of rapid response systems indicated that staff clinical concern in the
21 absence of a qualifying score was responsible for escalation in 47% of calls [7], highlighting the role
22 of staff education and empowerment, over and above EWS protocols The variability in physiological
23 normal baselines created by patient specific factors such as comorbidity or fitness means that using
24 vital signs observations alone as the basis for a score leading to mandatory escalation will always
25 require a trade-off between sensitivity in accurately identifying patients potentially at risk of
26 deterioration and staff alarm fatigue generated by patients who do not go onto deteriorate. This is
27 particularly pertinent in resource limited environments (such as during out of hours care),
28
29
30
31
32
33
34
35

36 Despite the mandated and widespread uptake of EWS, there has been minimal prospective
37 validation of their use. Efforts to improve precision in predicting outcome through scrutiny of large
38 datasets has largely employed analyses utilising area under the receiver operating characteristic
39 curves which are limited by the low prevalence of mortality in the population [8]. Before and after
40 studies have largely, but not universally [9-11] highlighted the efficacy of EWS, however no
41 randomised controlled trials have been performed. Consequently evidence of the scores' real impact
42 on clinical outcomes, such as mortality, transfer to higher level of care or length of stay, or on
43 workforce outcomes such as workload from excessive task generation and alarm fatigue, has only
44 been obtained from observational studies. These are all limited by significant confounders.
45
46
47
48
49
50

51 This evidence gap around the clinical and workforce implications of EWS systems will become
52 increasingly important as hospitals move towards automated systems with mandated referral of
53 patients who reach a threshold score. Continuing integration of more data into digital healthcare
54 systems via continuous monitoring, dynamic measures of fitness, and electronic health records will
55
56
57

1
2
3 further highlight this gap, as without an understanding of how these data can be applied it will be
4 difficult to differentiate the signal from the noise. Given the growing complexity of the inpatient
5 population more work is urgently required to understand the wider impact of EWS on outcomes
6 such as mortality and length of stay, task burden, working patterns and cost. There is also need to
7 reconsider the role of clinical concern in monitoring patients and how this can be further promoted
8 to prevent future systems depending purely on scores rather than integrating staff skills and
9 intuition into the decision making process. Early warning scores should not be developed in isolation
10 based on statistical performance as this fails to recognise that they are a component within the
11 complex clinical environment and therefore need to be designed to enhance, not complicate, the
12 clinical decision making process. This is particularly important in patients with respiratory disease
13 where physiology is often chronically deranged and less responsive to intervention and a greater
14 understanding of the contributory clinical factors and more individualised approach is required.
15 Although NEWS2 has been developed to address concerns regarding the altered physiology of
16 patients with respiratory disease, the new score was not based on any significant development in
17 the evidence base. Therefore the same questions currently remain regarding the real terms impact
18 of introducing any EWS, including NEWS2, and the associated software platforms on the patients
19 being monitored, the staff and resources required to deploy it and react to it, and the associated
20 opportunity cost.

21
22
23
24
25
26
27
28
29
30
31
32 Healthcare is becoming increasingly individualised, with significant amounts of digital healthcare
33 data collected. In recognition of this, a possible future direction would be to create scores which,
34 rather than being based solely on observations, integrate other more patient specific factors such as
35 comorbidity, premorbid fitness and age to apply specific weighting to observations. For example,
36 through applying a lower score to a high respiratory rate in someone who had chronic respiratory
37 disease and could mobilise 5 metres as a baseline as opposed to a young marathon runner, it would
38 be possible to maintain the same scoring thresholds at which a response was triggered, while
39 making those thresholds more meaningful through an evidence-based application of risk of
40 deterioration based on what a clinical observation represents in a particular individual.

41
42
43
44
45
46
47 Analysis of big data is the first stage to making this possible. However, the ability to demonstrate the
48 significance of changing either scoring thresholds or the scores themselves on patient and system
49 outcomes, driven by an attempt to compensate for changes to existing baseline physiology, will
50 require considerable numbers, novel prospective study design and collaboration across multiple
51 sites and research disciplines.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

For peer review only

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 band	EWS	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			Hourly for minimum 2 hours Consider fluid balance monitoring	RN to recheck Inform NIC Manual BP if deranged
3				Hourly with fluid balance monitoring
4	Min 1 hourly	RN to urgently inform medical team Urgent assessment by clinician with core competencies Clinical care in environment with monitoring facilities	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
5				
6	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		
>7				

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)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

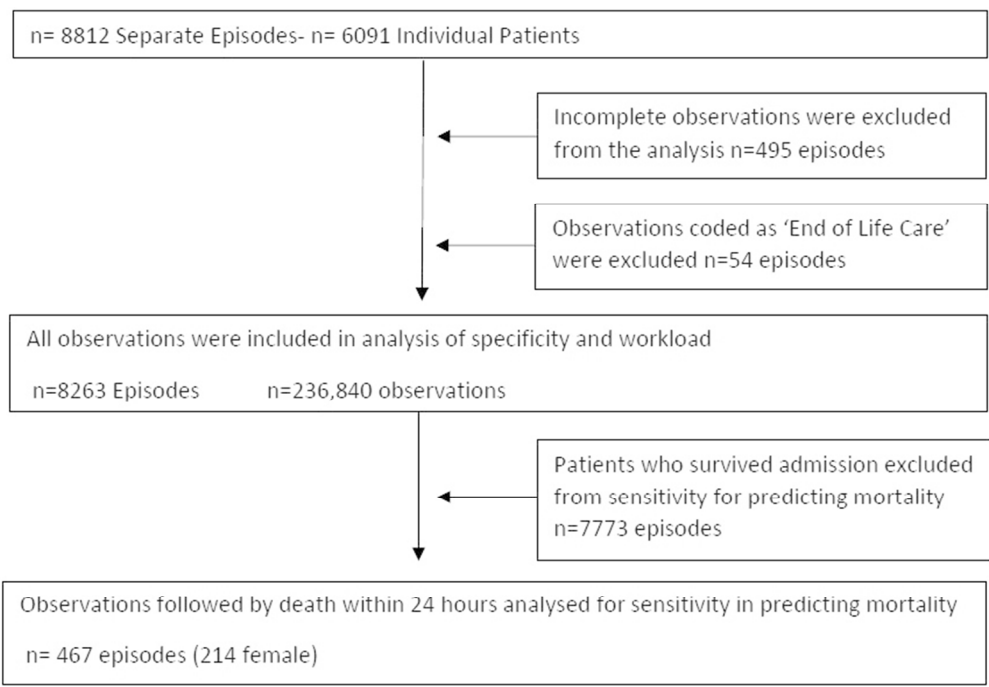


Figure 2. Cohort flow diagram of exclusion criteria

72x49mm (300 x 300 DPI)

ew only

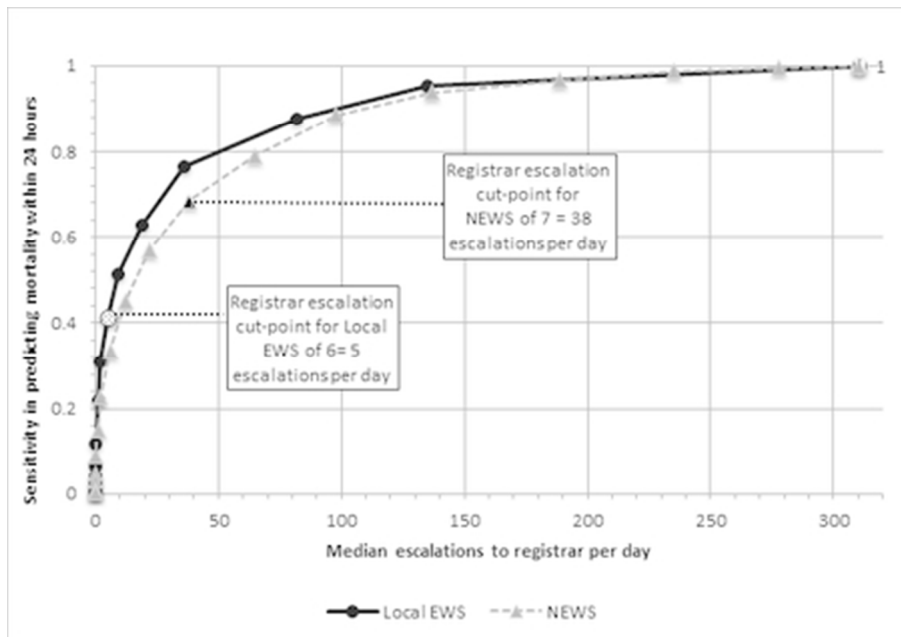


Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

38x29mm (300 x 300 DPI)

View only

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4
5 **Investigating the discriminative value of Early Warning Scores in**
6 **patients with respiratory disease using a retrospective cohort**
7 **analysis of admissions to Nottingham University Hospitals Trust over**
8 **a 2 year period.**
9
10
11
12
13

14 Sarah Forster^{1,2}, Gemma Housley³, Tricia M McKeever⁴, Dominick E Shaw^{2,3}
15
16
17
18

19 **Author Affiliations**
20

21 ¹ NIHR Academic Clinical Fellow, University of Nottingham, Nottingham, UK
22

23 ² Respiratory Research Unit, Division of Respiratory Medicine, University of Nottingham, Nottingham, UK
24

25 ³ Medical Informatics, East Midlands Academic Health Sciences Network, Nottingham, UK
26

27 ⁴ Division of Epidemiology, University of Nottingham, Nottingham, UK
28
29

30 Correspondence to Dr Sarah Forster, sarahforster@nhs.net
31
32
33

34 Word count: 3217
35
36
37
38

39 **Abstract**
40
41
42

43 **Objective**
44

45 Early Warning Scores (EWS) are used to monitor patients for signs of imminent deterioration.
46 Although used in respiratory disease, EWS have not been well studied in this population, despite the
47 underlying cardiopulmonary pathophysiology often present. We examined the performance of two
48 scoring systems in patients with respiratory disease.
49
50
51

52 **Design**
53
54
55
56
57

1
2
3 Retrospective cohort analysis of vital signs observations of all patients admitted to a respiratory unit
4 over a 2 year period. Scores were linked to outcome data to establish the performance of the
5 National EWS (NEWS) compared results to a locally adapted EWS.
6
7

8 **Setting**

9
10 Nottingham University Hospitals NHS Trust respiratory wards. Data were collected from an
11 integrated electronic observation and task allocation system employing a local EWS, also generating
12 mandatory referrals to clinical staff at set scoring thresholds.
13
14
15

16 **Outcome Measures**

17
18 Projected workload, and sensitivity and specificity of the scores in predicting mortality based on
19 outcome within 24 hours of a score being recorded.
20
21

22 **Results**

23
24 8812 individual patient episodes occurred during the study period. Overall mortality was 5.9%.
25 Applying NEWS retrospectively (versus local EWS) generated an eight fold increase in mandatory
26 escalations, but had higher sensitivity in predicting mortality at the protocol cut points.
27
28
29

30 **Conclusions**

31
32 This study highlights issues surrounding use of scoring systems in patients with respiratory disease.
33 NEWS demonstrated higher sensitivity for predicting death within 24 hours, offset by reduced
34 specificity. The consequent workload generated may compromise the ability of the clinical team to
35 respond to patients needing immediate input. The locally adapted EWS has higher specificity but
36 lower sensitivity. Statistical evaluation suggests this may lead to missed opportunities for
37 intervention, however this does not account for clinical concern independent of the scores, nor
38 ability to respond to alerts based on workload. Further research into the role of warning scores and
39 the impact of chronic pathophysiology is urgently needed.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

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

1
2
3 discriminative when compared to an unselected medical population [2]. This has significant
4 implications for patients, in terms of increased observations and interventions, and to clinical staff in
5 terms of workload and potential for alert fatigue. Consequently attempts have been made to
6 improve the score in this population [3].
7
8

9
10 Nottingham University Hospitals NHS Trust employs an electronic observations system with
11 mandatory escalation based on an adapted EWS. The Nottingham EWS, unlike NEWS, does not
12 score oxygen saturations and has a graduated approach to weighting for both oxygen delivery and
13 level of consciousness. As a more general marker of morbidity it also employs urine output. We
14 compared the sensitivity and specificity of the two scores in predicting mortality within 24 hours of a
15 set of observations being recorded at the clinical cut points determined by the associated protocols
16 and examined the potential impact in terms of workload of using the locally designed EWS versus
17 NEWS (see Figure 1) in patients with respiratory disease based on analysis of the vital signs
18 observations and outcomes of patients admitted to the respiratory department in Nottingham over
19 a 2 year period. We then went on to answer the same questions in a subgroup of patients who were
20 admitted with a diagnosis of COPD to examine the performance of the two scores in this cohort.
21
22
23
24
25
26

27
28 **Figure 1- Vital signs weighting and escalation protocol for NEWS and Nottingham**
29 **University Hospitals Early Warning Score**
30

31 **Methods**

32
33 We performed a single centre retrospective analysis of all patients admitted to the respiratory
34 department at Nottingham University Hospitals NHS Trust between 01/04/2015 and 31/03/2017.
35 This is a tertiary referral centre for respiratory medicine, with one specialist admissions ward and 3
36 inpatient wards. The analysis included all adults admitted with respiratory disease not transferred to
37 a higher level of care, i.e. high dependency or intensive care, greater than 24 hours before death as
38 these areas are not currently employing electronic observations, long term ventilator dependent
39 patients were also excluded as hospital policy dictates that these patients are always admitted to the
40 high dependency unit. Following approval from the NHS Information Governance Lead, and in line
41 with existing permissions within the East Midlands Academic Health Sciences network, data from the
42 integrated electronic observation and communication system comprising respiratory rate, oxygen
43 saturations, heart rate, blood pressure, temperature, conscious level (AVPU score), and urine output
44 were anonymised by an NHS data analyst prior to extraction from the clinical server. The same
45 system also automatically generates mandated escalation and referral at set scoring thresholds via a
46 pre-determined protocol. Scores from the local EWS were linked to demographics and mortality
47 outcomes prior to extraction. NEWS criteria were applied retrospectively to determine how many
48 patients would have been escalated if the NEWS system were followed. Results were analysed using
49
50
51
52
53
54
55
56
57
58
59

1
2
3 STATA 15. The entire data set was analysed for measurement of escalation patterns, analysis of
4 workload and sensitivity and specificity in predicting death within 24 hours of an observation [4]. A
5 chi-square analysis was performed to demonstrate whether the difference in escalations was
6 significant. The statistical analysis involved the use of all vital signs observations recorded
7 throughout admission, which were linked to outcome to determine whether they were followed by
8 death within 24 hours of the observation timestamp created by the input devices at the bedside.
9 Observations coded as end of life care following clinical decision were excluded from mortality
10 analysis (see Figure 2). A further subgroup analysis was then performed on patients coded as having
11 chronic obstructive pulmonary disease at any point in their admission as per ICD10 codes in order to
12 further assess the statistical performance of the two scores in the presence of chronic
13 pathophysiology.
14
15
16
17
18
19

20
21 **Figure 2. Cohort flow diagram of exclusion criteria**
22
23
24

25 **Patient and public involvement**

26
27 Prior to carrying out this work, a questionnaire was performed amongst stakeholders, in this case 26
28 medical registrars working in the East Midlands region. All worked in acute trusts that employed
29 either NEWS or the Nottingham EWS as part of a system to highlight patients felt to be at risk of
30 deterioration. Of the stakeholder responders, 70% believed that using EWS failed to highlight all
31 patients who went on to deteriorate and 88% felt that use of an EWS led to unnecessary reviews. All
32 responders felt there were issues in the setting of chronic disease with some chronic patients scoring
33 even at baseline, and 76% felt that alert fatigue due to high EWS was an issue. These findings guided
34 the interrogation of the data in creating the study detailed in this paper. It is also worth noting that
35 similar work presented to patients with recent inpatient experience at NUHT highlighted the belief
36 that sleep was too often interrupted by observations or reviews. However patients were not
37 involved directly in the design of this study.
38
39
40
41
42
43
44

45 **Results**

46
47 236,840 observation sets were recorded during 8812 inpatient episodes (53.1% female- see Table1)
48 involving 6091 individuals. In-hospital mortality for respiratory patients was 5.9% (n=521) and
49 median length of stay was 4 days (range 0-175).
50
51
52
53
54
55
56
57

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-

1
2
3 158) scores generating automatic referral to the registrar ($p < 0.001$ for difference between scores)
4 per day.
5

6
7 Sensitivity and specificity for predicting in-hospital mortality based on death within 24 hours of a set
8 of vital signs observations point is shown in Table 2. At each clinically equivalent band, the sensitivity
9 and specificity in predicting mortality of all patients scoring at and above that cut point are shown.
10 At each cut point, NEWS would have had a higher sensitivity than the local EWS (i.e. a higher
11 percentage of patients who went on to die were flagged as requiring escalation), but a lower
12 specificity.
13
14
15

16
17 Figure 3 plots sensitivity in predicting mortality, against median number of mandated clinician alerts
18 per day for both EWS types. It demonstrates that for a sensitivity of 0.7, NEWS generates a higher
19 number of mandated escalations. At both extremes of sensitivity (0 and 1) the number of escalations
20 is the same: i.e. mandating an escalation at a NEWS or EWS of 0 would mean all patients were
21 escalated, and each score would have 100% sensitivity for predicting mortality (as everyone who
22 died would have been reviewed). Likewise only escalating patients with a maximum EWS or NEWS
23 score would lead to very few patients being escalated.
24
25
26
27

28
29 **Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS**
30

31 Further subgroup analysis was performed on admissions with an ICD10 code for COPD at any point.
32 This yielded 56,345 observations from 2207 episodes by 1365 individual patients. Using the local
33 EWS protocol led to median of 0 (range 0-19) escalations to the registrar, while applying NEWS
34 would have generated a median of 6 (0-47) scores being escalated to the registrar each day. As in
35 the unselected respiratory cohort, NEWS was more sensitive in predicting imminent mortality than
36 the local EWS but with a significantly inferior specificity at each clinical cut point applied (see table
37 3).
38
39
40
41

42 **Discussion**

43
44
45
46

47 In this study we examined the effect of two different early warning score systems in patients
48 admitted with respiratory disease to a tertiary referrals centre. The respiratory department at NUHT
49 manages patients in line with national guidelines and has outcomes comparable with other similar
50 units; consequently linking of raw observations to outcomes prior to analysis enables conclusions
51 which are applicable to other centres.
52
53
54
55
56
57

1
2
3 We analysed the number of mandatory escalations generated and the sensitivity and specificity of
4 both of the scores in predicting imminent in-hospital mortality in an unselected respiratory
5 population and in a subgroup analysis of patients with COPD. Our data shows that at the scores' cut
6 points for escalation, NEWS would have generated a significantly higher workload due to a lower
7 specificity, with a higher sensitivity for predicting imminent deterioration, when compared with the
8 locally used EWS. This was accentuated in patients with COPD, an observation we believe is due to
9 chronic changes in the underlying physiology which influences the way in which these patients
10 respond to acute pathological processes.
11
12
13
14
15

16 Although NEWS may become less relevant with the publication of NEWS2 in December 2017, our
17 study remains relevant. Firstly, it highlights the wider impact of the different approaches to
18 designing a scoring system and the paucity of evidence in relation to how this is evaluated. Secondly,
19 as it is currently unclear how widely NEWS2 has been adopted by hospitals across the NHS and what
20 the likely roll out will be, NEWS remains a current clinical tool in many trusts.
21
22
23
24

25 Previous work has suggested that NEWS was less discriminative in predicting deterioration in
26 patients with respiratory disease, compared to a population of unselected medical admissions [2],
27 however NEWS has not previously been studied in large numbers of respiratory patients across an
28 entire admission.
29
30
31
32
33

34 Our study faced similar limitations to others published in this area. These include retrospective
35 study design preventing analysis of the real terms impact of introducing different scores into the
36 study environment on outcomes including length of stay, cardiac arrest rate and mortality; the low
37 prevalence of mortality in the patient population and the subsequent impact on observed effect
38 size; and the difficulty in recording accurate time of death in a general ward setting for use in
39 mortality analysis. .
40
41
42
43
44
45
46

47 However, our observed findings of an increase workload generated are both novel and important as,
48 when used as part of a system which employs automatic escalation of threshold scores, NEWS leads
49 to a significant impact on work load in a resource pressured environment, with little evidence of
50 improved clinical outcome. While there is a difference in the workload generated when comparing
51 the scoring systems both in a general respiratory population and in patients with COPD, this relates
52 to the cut points for escalation mandated by the protocols, rather than the scores themselves;
53 unsurprisingly overall both scores perform similarly when the individual scores are plotted (they are
54
55
56
57
58
59
60

1
2
3 based on similar clinical observations) however the mandated cut points differ. The difference
4 created by the protocol design relates to the way in which the scores are used clinically, and can be
5 explained as follows:
6

7
8 The first approach is seen in the scoring thresholds dictated by NEWS. Its cut points for each layer of
9 clinical intervention, i.e. escalation to nurse, clinician or registrar, have a higher sensitivity which acts
10 to rule out imminent clinical deterioration in those patients whose vital signs do not meet scoring
11 thresholds, meaning clinicians can be confident that patients with a low score are very unlikely to be
12 at imminent risk. This is akin to a d-dimer where a low value in an individual with low clinical
13 suspicion effectively excludes a venous thromboembolism [5, 6] . This high sensitivity approach
14 works well in a setting with less highly trained staff delivering the first layer of monitoring. However,
15 if this approach is applied in an unfiltered and automated manner, the workload generated by
16 escalations from patients who never go on to deteriorate will have significant resource and
17 operational implications, as well as increasing the likelihood of unnecessary intervention for
18 patients.
19

20
21 The second approach, used by the local EWS, is one of high specificity in the cut points for
22 escalation, with a relatively lower sensitivity. This approach acts to highlight potential imminent
23 clinical deterioration in those meeting the escalation criteria, but does not always rule out
24 deterioration in those who score under the cut point. This may seem a less preferable approach.
25 However a recent study of rapid response systems indicated that staff clinical concern in the
26 absence of a qualifying score was responsible for escalation in 47% of calls [7], highlighting the role
27 of staff education and empowerment, over and above EWS protocols The variability in physiological
28 normal baselines created by patient specific factors such as comorbidity or fitness means that using
29 vital signs observations alone as the basis for a score leading to mandatory escalation will always
30 require a trade-off between sensitivity in accurately identifying patients potentially at risk of
31 deterioration and staff alarm fatigue generated by patients who do not go onto deteriorate. This is
32 particularly pertinent in resource limited environments (such as during out of hours care),
33

34
35 Despite the mandated and widespread uptake of EWS, there has been minimal prospective
36 validation of their use. Efforts to improve precision in predicting outcome through scrutiny of large
37 datasets has largely employed analyses utilising area under the receiver operating characteristic
38 curves which are limited by the low prevalence of mortality in the population [8]. Before and after
39 studies have largely, but not universally [9-11] highlighted the efficacy of EWS, however no
40 randomised controlled trials have been performed. Consequently evidence of the scores' real impact
41 on clinical outcomes, such as mortality, transfer to higher level of care or length of stay, or on
42
43
44
45

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.

1
2
3 Analysis of big data is the first stage to making this possible. However, the ability to demonstrate the
4 significance of changing either scoring thresholds or the scores themselves on patient and system
5 outcomes, driven by an attempt to compensate for changes to existing baseline physiology, will
6 require considerable numbers, novel prospective study design and collaboration across multiple
7 sites and research disciplines.
8
9

10
11 However challenging, these points need to be addressed before any meaningful advances can be
12 made in this area and to ensure the most effective use of resources in the pursuit of improving the
13 safety and efficiency of patient care.
14
15

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

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			Hourly for minimum 2 hours Consider fluid balance monitoring	RN to recheck Inform NIC Manual BP if deranged
3				
4	Min 1 hourly	RN to urgently inform medical team Urgent assessment by clinician with core competencies Clinical care in environment with monitoring facilities	Hourly with fluid balance monitoring	RN to check EWS F1/SHO to discuss management with SpR or consultant
5				
6				
>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	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

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)

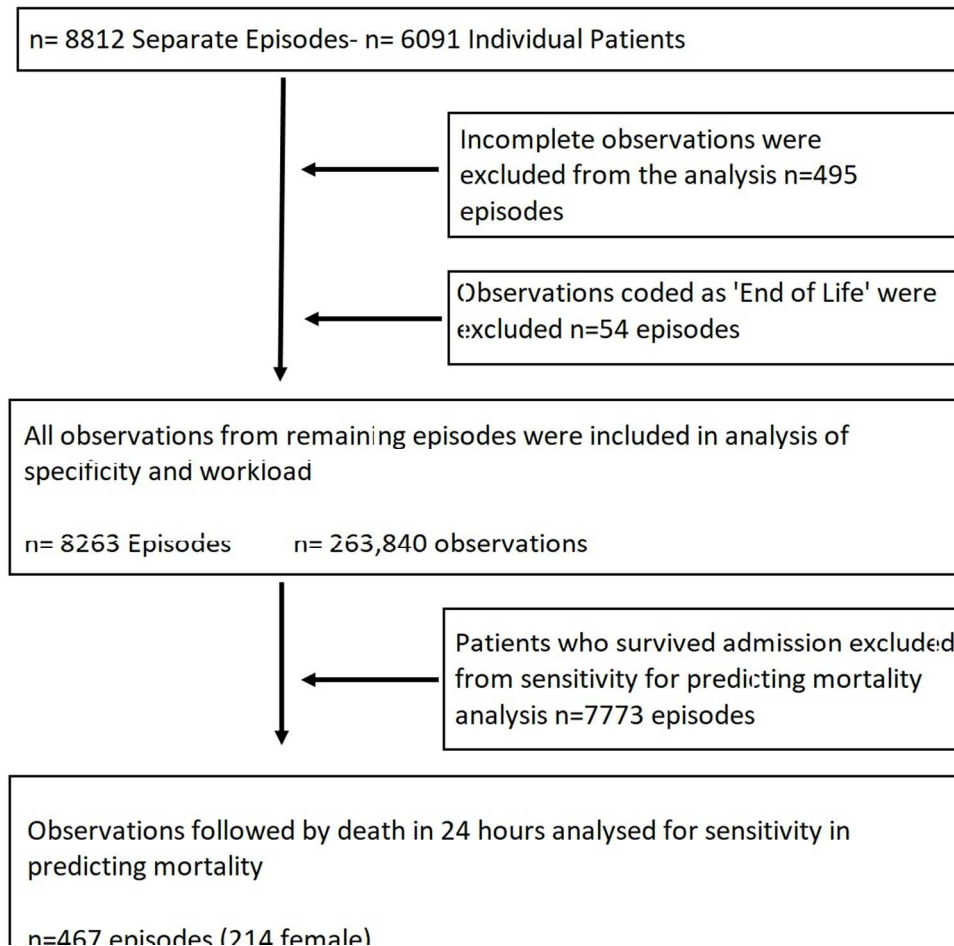


Figure 2. Cohort flow diagram of exclusion criteria

105x104mm (300 x 300 DPI)



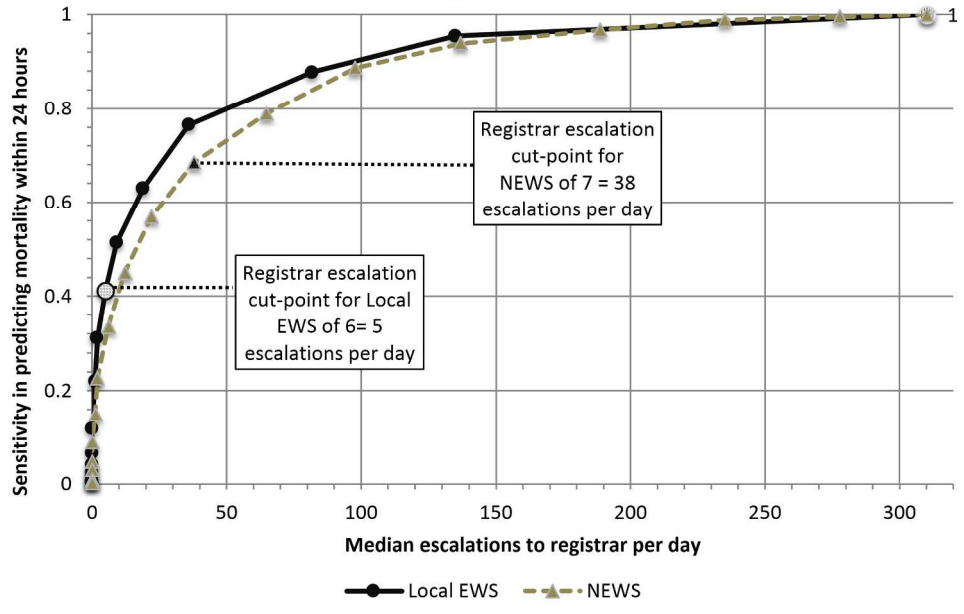


Figure 3. Graph of sensitivity versus alerts created for NEWS and local EWS

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.