Computer-aided National Early Warning Score to predict the risk of sepsis following emergency medical admission to hospital: a model development and external validation study.

Authors

Muhammad Faisal, PhD
Senior Research Fellow in Medical Statistics
Faculty of Health Studies, University of Bradford, Bradford, UK
Bradford Institute for Health Research
E-mail: M.Faisal1@bradford.ac.uk

Donald Richardson, MBChB
Deputy Medical Director
Renal Physician
Department of Renal Medicine, York Teaching Hospital NHS Foundation Trust Hospital
E-mail: drichardson@doctors.org.uk

Andrew J Scally, MSc
Senior Lecturer
School of Clinical Therapies
University College Cork, Ireland
E-mail: andrew.scally@ucc.ie

Robin Howes, MSc
Digital Research and Development Manager
Department of Strategy & Planning
Northern Lincolnshire and Goole Hospitals
E-mail: robin.howes@nhs.net

Kevin Beatson, MSc
Development Manager
York Teaching Hospital NHS Foundation Trust
E-mail: Kevin.Beatson@York.NHS.uk

Kevin Speed, MBBS, MSc
Consultant Haematologist
Northern Lincolnshire and Goole Hospitals
E-mail: kevin.speed@nhs.net

Mohammed A Mohammed, PhD
Professor of Healthcare Quality & Effectiveness
Faculty of Health Studies, University of Bradford, Bradford, UK
The Strategy Unit, NHS Midlands and Lancashire Commissioning Support Unit
E-mail: M.A.Mohammed5@Bradford.ac.uk

Correspondence to Mohammed A Mohammed

Contributorship

DR & MAM had the original idea for this work. RH and KB extracted the necessary data frames. DR gave a clinical perspective. MF, AS & MAM undertook the statistical analyses. MF and DR wrote the first draft of this paper and all authors subsequently assisted in redrafting and have approved the final version. MAM will act as guarantor.

Funding

This research was supported by the Health Foundation. The Health Foundation is an independent charity working to improve the quality of health care in the UK.

This research was supported by the National Institute for Health Research (NIHR) Yorkshire and Humberside Patient Safety Translational Research Centre (NIHR YHPSTRC). The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Competing Interests

The authors declare no conflicts of interest.

Abstract

Background: In English hospitals, the patient's vital signs are monitored and summarised into a National Early Warning Score (NEWS). NEWS is more accurate than the quick sepsis related organ failure assessment (qSOFA) score at identifying patients with sepsis. We investigate the extent to which the accuracy of the NEWS is enhanced by developing computer-aided NEWS (cNEWS) models. We compared three cNEWS models (M0=NEWS alone; M1=M0 + age + sex; M2=M1 + subcomponents of NEWS + diastolic blood pressure) to predict the risk of sepsis.

Methods: All adult emergency medical admissions discharged over 24-months from two acute hospitals (YH–York Hospital for model development; NH–Northern Lincolnshire and Goole Hospital for external model validation). We used a validated Canadian method for defining sepsis from administrative hospital data.

Findings: The prevalence of sepsis was lower in YH (4.5%=1596/35807) than NH (8.5%=2983/35161). The c-statistic increased across models (YH: M0: 0.705, M1:0.763, M2:0.777; NH:M0: 0.708, M1:0.777, M2:0.791). At NEWS 5+, sensitivity increased (YH: 47.24% vs 50.56% vs 52.69%; NH: 37.91% vs 43.35% vs 48.07%)., the positive likelihood ratio increased (YH: 2.77 vs 2.99 vs 3.06; NH: 3.18 vs 3.32 vs 3.45) and the positive predictive value increased (YH: 11.44% vs 12.24% vs 12.49%; NH: 22.75% vs 23.55% vs 24.21%).

Interpretation: From the three cNEWS models, Model M2 is the most accurate. Since it places no additional data collection burden on clinicians and can be automated, it may now be carefully introduced and evaluated in hospitals with sufficient informatics infrastructure.

Key words: vital signs, national early warning score, emergency admission, sepsis, computer aided national early warning score

Introduction

Sepsis is a major cause of mortality in hospitals. Survival is dependent on early recognition and treatment. Although each hour of delay is associated with a 7% reduction in survival (1,2), studies have found that treatment delays are not uncommon in hospitals (3). Several risk scores have been devised to aid the early detection of sepsis (4). Two widely used scores are the Systemic Inflammatory Response Syndrome (SIRS) (5) and the quick sepsis related organ failure assessment (qSOFA) scores (4). Although studies have found SIRS to be more accurate than qSOFA for diagnosis of sepsis (6), a recent study found that the National Early Warning Score (NEWS) compares favourably with qSOFA (7,8).

The NEWS was introduced in 2012, by the Royal College of Physicians of London to identify acutely ill patients, including those with sepsis (9). The NEW score is used to identify those at risk of death and increased morbidity in all patient diagnostic groups (with some noted exceptions – i.e., head injury). This score has been widely adopted in National Health Service (NHS) hospitals in England and other countries (9). NEWS is derived from seven physiological variables or vital signs - respiration rate, oxygen saturations, any supplemental oxygen, temperature, systolic blood pressure, heart rate and level of consciousness (Alert, Voice, Pain, Unresponsive) – which are routinely collected by nursing staff as an integral part of the process of care, usually within 30 minutes for most patients and subsequently repeated at a frequency dependent on local hospital protocols. NEWS points are allocated according to basic clinical observations and the higher the NEWS the more likely it is that the patient is developing a critical illness (see appendix for further details of the NEWS). The clinical rationale for NEWS is that early recognition of deterioration in the vital signs of a patient can provide opportunities for earlier, more effective intervention. Furthermore, studies have shown that electronically collected NEWS (10) are highly reliable and accurate when compared with paper based methods (11–13) and about two thirds of NHS hospitals now report the use of electronic NEWS (eNEWS) (14).

A NEWS of 5+ has been recommended as a trigger point to screen for sepsis by the National Institute for Health and Care Excellence in England (15,16) and has been widely adopted in NHS Hospitals. However, given the widespread use of eNEWS and its potential to support real-time computer aided screening for sepsis, we investigate the extent to which the accuracy of NEWS for predicting sepsis, could be enhanced by developing computer-aided NEWS (cNEWS) models. We examine the accuracy of cNEWS models which include age, sex and the subcomponents of NEWS versus a reference model that uses NEWS only using a validated method for defining sepsis developed by Jolley et al in Canada (17). An important feature of our cNEWS models is that they are not designed for paper based systems and do not place any additional burden of data collection and/or calculation on the clinicians because cNEWS relies on data which are (a) routinely collected as part of the process of care, (b) already stored in the patient's electronic health record and (c) accessible in real-time thus offering the prospects of real-time risk predictions without hindering clinical workflows.

Methods

Setting & data

Our cohorts of emergency medical admissions are from three acute hospitals which are approximately 100 kilometres apart in the Yorkshire & Humberside region of England– the Diana, Princess of Wales Hospital (n~400 beds) and Scunthorpe General Hospital (n~400 beds) managed by the Northern Lincolnshire and Goole NHS Foundation Trust (NLAG), and York Hospital (YH) (n~700 beds), managed by York Teaching Hospitals NHS Foundation Trust. For the purposes of this study, the two acute hospitals in NLAG are combined into a single dataset and collectively referred to as NLAG Hospitals (NH). NH and YH have been exclusively using electronic NEWS scoring since at least 2013 as part of their in-house electronic patient record systems. We selected these hospitals because they had electronic NEWS which are collected as part of the patient's process of care and were agreeable to the study.

We considered all adult (age≥16 years) emergency medical admissions, discharged during a 24-month period (1 January 2014 to 31 December 2015), with eNEWS. For each emergency admission, we obtained a pseudo-anonymised patient identifier, patient's age (years), sex (male/female), discharge status (alive/dead), admission and discharge date and time, and eNEWS (including its subcomponents respiratory rate, temperature, systolic blood pressure, pulse rate, oxygen saturation, oxygen supplementation, and alertness). NEWS does not include diastolic blood pressure, but we incorporate it in our statistical models because this data item is routinely collected (see later). NEWS ranges from 0 (indicating the lowest severity of illness) to a maximum of 20. (see Appendix for further details). The admission/discharge date and eNEWS were date and time stamped and the index eNEWS was defined as the first score electronically recorded within ±24 hours of the admission time. We excluded records where the index eNEWS was not within ±24 hours or was not recorded at all (see Table S1 & S2 – supplemental digital content).

We define sepsis (with at least one organ failure or septic shock) (4) based on 84 selected ICD-10 codes identified by an optimised validated method reported by Jolley et al in Canada (17) (which we adapted to our study by excluding six Canadian specific ICD-10 codes and three procedure codes). We used this optimised approach for identifying sepsis using ICD-10 codes because other methods are known to underestimate sepsis from administrative data (17,18). We reported the statistical differences in characteristics of our two hospitals using two independent sample t-test (for continuous data) and chi square proportion test (for categorical data).

Statistical Modelling

We began with exploratory analyses including scatter plots and box plots that showed the relationship between covariates and risk of sepsis in our hospitals. We developed three logistic regression models for the risk of sepsis. The models (M0,M1,M2) use the index or first recorded eNEWS within ±24hours of admission. Model M0 uses eNEWS alone; Model M1 extends M0 with age and sex and Model M2 extends M1 with all the subcomponents of NEWS plus diastolic blood pressure. We used likelihood ratio tests to determine the extent to which progressing from models M0 to M2 improved the goodness of fit.

We used the *qladder* function (Stata (19)), which displays the quantiles of transformed variable against the quantiles of a normal distribution according to the ladder powers $\left(x^3, x^2, x^1, x, \sqrt{x}, \log(x), x^{-1}, x^{-2}, x^{-3}\right)$ for each continuous covariate and chose the following

transformations:- $\log_e$ (respiratory rate), $\log_e$ (pulse rate), $\log_e$ (systolic blood pressure), and $\log_e$ (diastolic blood pressure).

All models were developed to predict the risk of sepsis following emergency medical admission using data from only YH (development dataset). We then externally validated these models using data from another hospital NH (external validation dataset). We report discrimination and calibration statistics as performance measures for these models (20).

Discrimination relates to how well a model can separate, (or discriminate) between cases with and without sepsis and is given by the area under the Receiver Operating Characteristics (ROC) curve (AUC) or c-statistic after adjusting for differences in the baseline (21) risk of sepsis in our two hospitals. The 95% confidence interval for the c-statistic was derived using DeLong's method as implemented in the pROC library (22) in R (23).

Calibration is the relationship between the observed and predicted risk of sepsis (24) and can be readily seen on a scatter plot (y-axis observed risk, x-axis predicted risk). Perfect predictions should be on the 45° line. The intercept (a) and slope (b) of this line gives an assessment of 'calibration-in-the-large'. At model development, a = 0 and b = 1, but at external validation, calibration-in-the-large problems are indicated if a is not 0 and if b is more/less than 1 as this reflects problems of under/over prediction.

The cut-off of NEWS 5+ is the recommended threshold for screening sepsis (15,16). We determined the sensitivity, specificity, positive and negative predictive values and likelihood ratios for these models (M0, M1, M2) at eNEWS 4+,5+,and 6+ thresholds (25). For the best performing model (M2) we further analysed its performance across a range of risks of sepsis (5% to 15%) to highlight the performance characteristics of this model which may inform choice of thresholds in routine clinical practice.

Analyses were carried using R (23), using the ROCR (25) library and Stata (19).

Supplementary Materials

Tables and figures prefixed by "S" and shown as – supplemental digital content.

Ethical Approval

We obtained ethical approval for the main research project of which this is a sub study from Yorkshire & The Humber - Leeds West Research Ethics Committee (reference number 15/YH/0348).

Results

Description of development and validation datasets

The number (YH: n=36751; NH: n=37100) of emergency medical admissions over a 24-month period was similar in our two hospitals. We excluded 2.6% (944/36751) of admissions in YH and 5.2% (1939/37100) in NH because the index eNEWS was not recorded within 24 hours of the admission or there was no eNEWS recorded at all (see Table S1 & S2 – supplemental digital content).

The characteristics of the admissions included in our study are shown in Table 1. Emergency admission in YH were older than those in NH (67.8 years vs 66.4 years), less likely to be male (47.3% vs 49.8%), had higher index eNEWS (2.5 vs 2.1), much lower prevalence of sepsis (4.5% vs 8.5%) but similar in-hospital mortality (5.8% vs 5.4%). The prevalence of oxygen supplementation was lower in the YH compared to NH (11.3% vs 19.2%). See accompanying scatter and boxplots in figure S1 to S4 – supplemental digital content. Figure 1 shows the relationship between the index eNEWS with sepsis in each hospital. As the index eNEWS increases so too does the risk of sepsis.

Characteristic	**Development dataset (YH)**	**Validation dataset (NH)**	**p-value**
	N=35807	**N=35161**	**-**
Mean Age [years] (SD)	67.8 (19.5)	66.4 (19.5)	<0.001
Male (%)	16936 (47.3)	17498 (49.8)	<0.001
Sepsis [outcome] (%)	1596 (4.5)	2983 (8.5)	<0.001
In-Hospital Mortality (%)	2080 (5.8)	1900 (5.4)	0.32
Mean index eNEWS (SD)	2.5 (2.6)	2.1 (2.3)	<0.001
Alertness			<0.001
Alert (%)	34769 (97.1)	34503 (98.1)	
Pain (%)	243 (0.7)	126 (0.4)	
Voice (%)	607 (1.7)	435 (1.2)	
Unconscious (%)	188 (0.5)	97 (0.3)	
Mean Respiratory rate [breaths per minute] (SD)	18.6 (4.8)	18.1 (3.6)	<0.001
Mean Temperature [^{o}C] (SD)	36.3 (0.8)	36.5 (0.7)	<0.001
Mean Systolic pressure [mmHg] (SD)	136 (27.3)	129.4 (23)	<0.001
Mean Diastolic pressure [mmHg] (SD)	75.4 (15.5)	74.9 (14.9)	<0.001
Mean Pulse rate [beats per minute] (SD)	85.6 (21.1)	81.2 (17.8)	<0.001
Oxygen supplementation (%)	4053 (11.3)	6750 (19.2)	<0.001
Mean % Oxygen saturation (SD)	96.3 (2.9)	95.9 (3.0)	<0.001

Table 1 Characteristics of emergency medical admissions in the development and validation datasets.

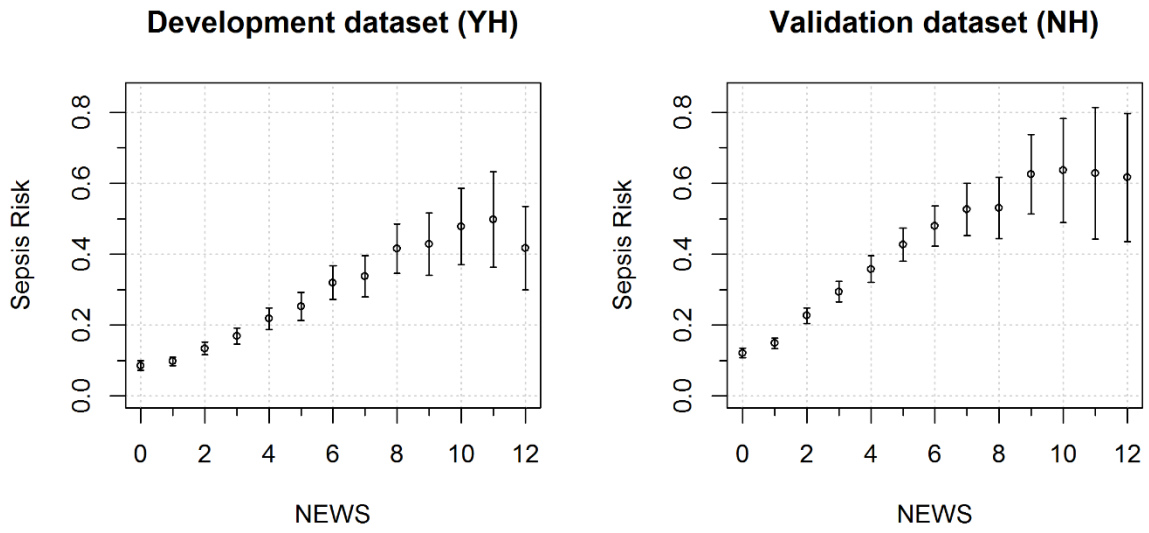

Figure 1: Observed different type of sepsis risk versus index electronic NEWS in YH and NH hospitals. Vertical bars are exact binomial 95% confidence intervals.
Note: for visualisation purposes, we capped NEWS to 12.

Statistical Modelling

We compared three cNEWS models (M0=NEWS alone; M1=M0 plus age and sex; M2=M1 plus the subcomponents of NEWS plus diastolic blood pressure) to predict the risk of sepsis. The models were developed to predict the risk of sepsis on data from YH and externally validated using data from NH.

The likelihood ratio test showed statistically significant improvement in model goodness of fits (M0 vs M1: $\chi 2=416.8$ (df=2) $p<0.001$; M1 vs M2: $\chi 2=161.8$ (df=10) $p<0.001$). The ROC plots for each model are shown in figure 2 with their accompanying discrimination and calibration statistics shown in table 2. Model M0 had the lowest c-statistic in the development (0.705) and validation datasets (0.708). Models M1 and M2 had higher c-statistics (see table 2).

The external validation slope reduced from 1.18 (M0) to 1.15 (M2). The internal and external validation plots are shown in Figure S5 - supplemental digital content.

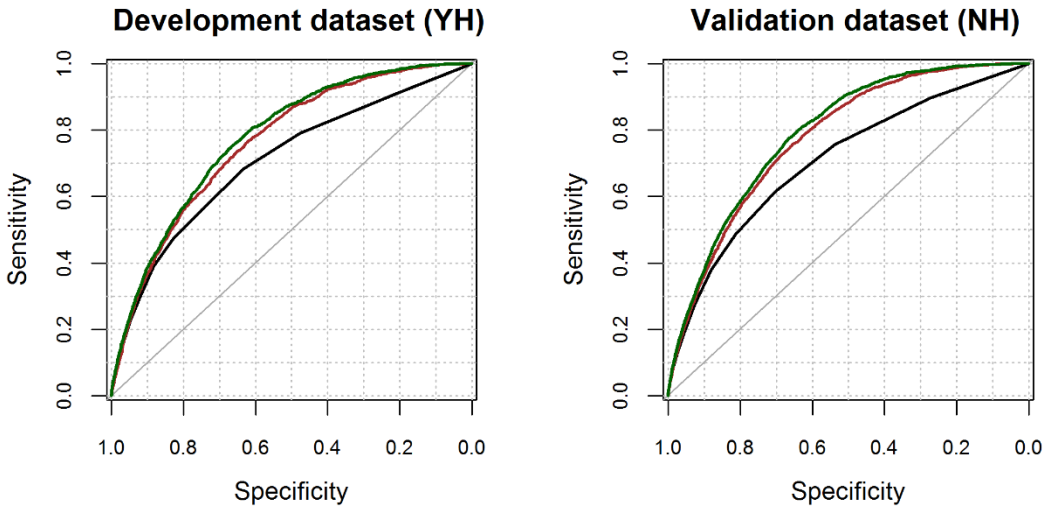

Figure 2 Receiver Operating Characteristic curve for four models (M0,M1,M2) in predicting the risk of sepsis in the YH and NH hospitals. Black line is M0; Red line is M1 and Green line is M2.

Model	Discrimination c-statistic: Development dataset YH (95%CI)	Discrimination c-statistic: Validation dataset NH (95%CI)	Calibration slope validation dataset NH (95%CI)
M0=eNEWS only	0.705 (0.692 to 0.719)	0.708 (0.698 to 0.718)	1.18 (1.12 to 1.23)
M1=M0+age+sex	0.763 (0.752 to 0.774)	0.777 (0.769 to 0.784)	1.18 (1.13 to 1.23)
M2= M1+ subcomponents of NEWS + diastolic blood pressure	0.777 (0.766 to 0.787)	0.791 (0.783 to 0.798)	1.15 (1.11 to 1.18)

Table 2 Performance of three cNEWS models in predicting the risk of sepsis in the development and validation datasets.

Table 3 shows the sensitivity, specificity, positive predictive value and negative predictive value at eNEWS 4+,5+, and 6+ for models M0, M1 and M2 for predicting the risk of sepsis.

At the current recommended threshold of a NEWS of 5+ for screening for sepsis, sensitivity increased across models M0, M1 and M2 in the development dataset (47.24% vs 50.56% vs 52.69%) and the external validation dataset (37.91% vs 43.35% vs 48.07%). Specificity changed little in the development dataset (82.94% vs 83.09% vs 82.77%) and external validation data set (88.07 vs 86.96 vs 86.05). The positive likelihood ratio increased across models in the development dataset (2.77 vs 2.99 vs 3.06) and the external validation dataset (3.18 vs 3.32 vs 3.45). The negative likelihood ratio reduced across models in the development dataset (0.64 vs 0.59 vs 0.57) and the external validation dataset (0.70 vs 0.65 vs 0.60). The positive predictive value increased in the development dataset (11.44 vs 12.24 vs 12.49) and the external validation dataset (22.75 vs 23.55 vs 24.21). The negative predictive value increased slightly in the development dataset from (97.12 vs 97.30 vs 97.40) and the external validation dataset (93.87 vs 94.30 vs 94.70).

We further explored the behaviour of the best performing cNEWS model (M2) across a range of cut-off probabilities (5% to 15%) – see Table 3. For YH a cut-off of 0.06 appears to offer reasonable performance, likewise, 0.12 appears to be a reasonable cut-off for NH.

		Development dataset (YH)						External Validation dataset (NH)					
NEWS threshold	**Models**	**Sensitivity%**	**Specificity%**	**PPV**	**NPV**	**LR+**	**LR-**	**Sensitivity%**	**Specificity%**	**PPV**	**NPV**	**LR+**	**LR-**
4+ (equivalent predicted probability YH=0.051 & NH = 0.11)	M0	55.95 (53.48 to 58.41)	74.91 (74.45 to 75.37)	9.42 (8.84 to 10.03)	97.33 (97.13 to 97.52)	2.23 (2.13 to 2.34)	0.59 (0.56 to 0.62)	48.58 (46.77 to 50.39)	81.29 (80.85 to 81.71)	19.39 (18.5 to 20.31)	94.46 (94.18 to 94.73)	2.60 (2.49 to 2.71)	0.63 (0.61 to 0.66)
	M1	60.96 (58.52 to 63.37)	75.53 (75.07 to 75.98)	10.41 (9.8 to 11.05)	97.65 (97.46 to 97.83)	2.49 (2.39 to 2.6)	0.52 (0.49 to 0.55)	57.09 (55.29 to 58.88)	79.77 (79.32 to 80.2)	20.73 (19.86 to 21.63)	95.25 (94.99 to 95.5)	2.82 (2.72 to 2.93)	0.54 (0.52 to 0.56)
	M2	63.22 (60.8 to 65.59)	75.55 (75.1 to 76.01)	10.77 (10.15 to 11.41)	97.78 (97.59 to 97.95)	2.59 (2.48 to 2.7)	0.49 (0.46 to 0.52)	59.5 (57.72 to 61.27)	79.38 (78.94 to 79.82)	21.11 (20.24 to 22)	95.48 (95.23 to 95.73)	2.89 (2.78 to 2.99)	0.51 (0.49 to 0.53)
5+ (equivalent predicted probability YH=0.063 & NH = 0.138)	M0	47.24 (44.77 to 49.73)	82.94 (82.54 to 83.34)	11.44 (10.68 to 12.23)	97.12 (96.92 to 97.31)	2.77 (2.62 to 2.93)	0.64 (0.61 to 0.67)	37.91 (36.17 to 39.68)	88.07 (87.71 to 88.42)	22.75 (21.59 to 23.94)	93.87 (93.59 to 94.13)	3.18 (3.01 to 3.36)	0.70 (0.69 to 0.73)
	M1	50.56 (48.08 to 53.05)	83.09 (82.69 to 83.49)	12.24 (11.46 to 13.06)	97.30 (97.11 to 97.48)	2.99 (2.83 to 3.16)	0.59 (0.57 to 0.63)	43.35 (41.56 to 45.15)	86.96 (86.58 to 87.32)	23.55 (22.43 to 24.7)	94.30 (94.03 to 94.57)	3.32 (3.16 to 3.49)	0.65 (0.63 to 0.67)
	M2	52.69 (50.21 to 55.17)	82.77 (82.37 to 83.17)	12.49 (11.71 to 13.3)	97.40 (97.21 to 97.58)	3.06 (2.9 to 3.22)	0.57 (0.54 to 0.6)	48.07 (46.27 to 49.88)	86.05 (85.67 to 86.43)	24.21 (23.12 to 25.32)	94.70 (94.44 to 94.96)	3.45 (3.29 to 3.61)	0.60 (0.58 to 0.62)
6+ (equivalent predicted probability YH=0.079 & NH = 0.169)	M0	39.22 (36.82 to 41.67)	88.1 (87.75 to 88.44)	13.33 (12.37 to 14.33)	96.88 (96.68 to 97.07)	3.30 (3.08 to 3.53)	0.69 (0.66 to 0.72)	27.49 (25.89 to 29.13)	92.65 (92.36 to 92.93)	25.74 (24.23 to 27.29)	93.24 (92.95 to 93.51)	3.74 (3.49 to 4.01)	0.78 (0.77 to 0.8)
	M1	40.23 (37.81 to 42.68)	88.22 (87.87 to 88.56)	13.74 (12.77 to 14.76)	96.94 (96.74 to 97.12)	3.41 (3.2 to 3.65)	0.68 (0.65 to 0.71)	31.38 (29.71 to 33.08)	91.91 (91.61 to 92.21)	26.46 (25.01 to 27.94)	93.53 (93.25 to 93.8)	3.88 (3.64 to 4.14)	0.75 (0.73 to 0.77)
	M2	42.04 (39.61 to 44.51)	88.13 (87.78 to 88.47)	14.18 (13.2 to 15.21)	97.02 (96.83 to 97.21)	3.54 (3.32 to 3.78)	0.66 (0.63 to 0.69)	35.67 (33.95 to 37.42)	90.8 (90.48 to 91.11)	26.44 (25.08 to 27.83)	93.84 (93.56 to 94.1)	3.88 (3.65 to 4.11)	0.71 (0.69 to 0.73)

Table 3 Sensitivity, specific and related analyses of M0, M1 and M2 to predict the risk of sepsis in the development and external validation datasets at selected eNEWS thresholds.

PPV=Positive Predictive Value; NPV= Negative Predictive Value; LR+=Positive Likelihood Ratio; LR-=Negative Likelihood Ratio

Predicted probabilities cut-offs		Development dataset (YH)						External Validation dataset (NH)						
	N+	Sensitivity%	Specificity%	PPV	NPV	LR+	LR-	N+	Sensitivity%	Specificity%	PPV	NPV	LR+	LR-
0.05	9558	63.85 (61.43 to 66.21)	75.04 (74.58 to 75.5)	10.66 (10.05 to 11.3)	97.8 (97.62 to 97.98)	2.56 (2.45 to 2.67)	0.48 (0.45 to 0.51)	20864	93.66 (92.73 to 94.51)	43.84 (43.3 to 44.39)	13.39 (12.93 to 13.86)	98.68 (98.48 to 98.86)	1.67 (1.65 to 1.69)	0.14 (0.13 to 0.17)
0.06	**7337**	**55.2 (52.72 to 57.66)**	**81.13 (80.71 to 81.54)**	**12.01 (11.27 to 12.77)**	**97.49 (97.3 to 97.67)**	**2.93 (2.78 to 3.07)**	**0.55 (0.52 to 0.58)**	18160	89.78 (88.63 to 90.84)	51.89 (51.34 to 52.43)	14.75 (14.23 to 15.27)	98.21 (98 to 98.4)	1.87 (1.84 to 1.9)	0.2 (0.18 to 0.22)
0.07	5753	47.24 (44.77 to 49.73)	85.39 (85.01 to 85.76)	13.11 (12.24 to 14.01)	97.2 (97.01 to 97.38)	3.23 (3.05 to 3.43)	0.62 (0.59 to 0.65)	15471	83.04 (81.64 to 84.37)	59.62 (59.08 to 60.16)	16.01 (15.44 to 16.6)	97.43 (97.2 to 97.65)	2.06 (2.01 to 2.1)	0.28 (0.26 to 0.31)
0.08	4639	41.79 (39.36 to 44.26)	88.39 (88.05 to 88.73)	14.38 (13.38 to 15.42)	97.02 (96.82 to 97.21)	3.6 (3.37 to 3.84)	0.66 (0.63 to 0.69)	13331	77.98 (76.44 to 79.45)	65.8 (65.28 to 66.32)	17.45 (16.81 to 18.1)	96.99 (96.76 to 97.21)	2.28 (2.23 to 2.34)	0.33 (0.31 to 0.36)
0.09	3834	36.97 (34.59 to 39.39)	90.52 (90.2 to 90.83)	15.39 (14.26 to 16.57)	96.85 (96.66 to 97.04)	3.9 (3.63 to 4.19)	0.7 (0.67 to 0.72)	11477	71.61 (69.95 to 73.22)	70.97 (70.47 to 71.47)	18.61 (17.9 to 19.34)	96.42 (96.18 to 96.66)	2.47 (2.4 to 2.54)	0.4 (0.38 to 0.42)
0.10	3243	32.39 (30.1 to 34.75)	92.03 (91.74 to 92.32)	15.94 (14.7 to 17.25)	96.69 (96.49 to 96.88)	4.07 (3.75 to 4.4)	0.73 (0.71 to 0.76)	9956	65.87 (64.14 to 67.58)	75.17 (74.69 to 75.64)	19.74 (18.96 to 20.53)	95.96 (95.71 to 96.2)	2.65 (2.57 to 2.74)	0.45 (0.43 to 0.48)
0.11	2677	28.13 (25.94 to 30.41)	93.49 (93.22 to 93.75)	16.77 (15.38 to 18.24)	96.54 (96.34 to 96.73)	4.32 (3.96 to 4.72)	0.77 (0.75 to 0.79)	8644	60.41 (58.63 to 62.17)	78.74 (78.29 to 79.18)	20.85 (19.99 to 21.72)	95.55 (95.29 to 95.79)	2.84 (2.74 to 2.94)	0.5 (0.48 to 0.53)
0.12	2265	24.87 (22.77 to 27.07)	94.54 (94.29 to 94.78)	17.53 (15.98 to 19.16)	96.43 (96.22 to 96.62)	4.56 (4.14 to 5.01)	0.79 (0.77 to 0.82)	**7540**	**55.88 (54.08 to 57.68)**	**81.75 (81.32 to 82.17)**	**22.11 (21.18 to 23.06)**	**95.24 (94.98 to 95.48)**	**3.06 (2.94 to 3.18)**	**0.54 (0.52 to 0.56)**
0.13	1913	21.68 (19.68 to 23.78)	95.42 (95.19 to 95.64)	18.09 (16.39 to 19.89)	96.31 (96.11 to 96.51)	4.73 (4.26 to 5.26)	0.82 (0.8 to 0.84)	6556	51.39 (49.58 to 53.2)	84.39 (83.99 to 84.78)	23.38 (22.36 to 24.43)	94.93 (94.67 to 95.18)	3.29 (3.15 to 3.44)	0.58 (0.55 to 0.6)
0.14	1644	19.24 (17.33 to 21.26)	96.09 (95.88 to 96.29)	18.67 (16.82 to 20.64)	96.23 (96.02 to 96.43)	4.92 (4.39 to 5.51)	0.84 (0.82 to 0.86)	5776	47.07 (45.26 to 48.88)	86.41 (86.03 to 86.79)	24.31 (23.21 to 25.44)	94.63 (94.36 to 94.88)	3.46 (3.31 to 3.63)	0.61 (0.59 to 0.63)
0.15	1407	16.92 (15.11 to 18.85)	96.68 (96.48 to 96.86)	19.19 (17.16 to 21.35)	96.15 (95.94 to 96.35)	5.09 (4.5 to 5.76)	0.86 (0.84 to 0.88)	5086	42.74 (40.96 to 44.54)	88.16 (87.8 to 88.51)	25.07 (23.88 to 26.28)	94.32 (94.05 to 94.58)	3.61 (3.43 to 3.8)	0.65 (0.63 to 0.67)

Table 4 Sensitivity, specific and related analyses of M2 to predict the risk of sepsis in the development and external validation datasets at range of predicted probabilities cut-offs

N+= the number of positive cases; PPV=Positive Predictive Value; NPV= Negative Predictive Value; LR+=Positive Likelihood Ratio; LR-=Negative Likelihood Ratio

Bold underline tentatively indicates a reasonable cut-off choice.

Discussion

In this study we developed three computer-aided versions of eNEWS models which incorporated progressively more information. Model M0 uses eNEWS alone; Model M1 extends M0 with age and sex and Model M2 extends M1 with all the subcomponents of NEWS plus diastolic blood pressure. We found that M2 was the best model. This is unsurprising because it incorporates additional information about the patient's age and other vital signs. The main advantages of these computer-aided eNEWS models is that they are designed to incorporate data which are already available in the patient's electronic health record and so place no additional data collection or computational burden on clinicians and they are readily automated. Nonetheless, we note that computer aided risk scores are not designed to replace clinical judgement. They are intended and designed to support, not undermine, clinical decision making and can be overridden by clinical concern (9,26). The hypothesis for our computer aided eNEWS scores is that they may enhance situational awareness of sepsis by processing information already available without impeding the workflow of clinical staff.

Although our previously published computer aided risk of sepsis (CARS) score (27), which is based on physiological variables and blood results, offers more accuracy than cNEWS, it has two main disadvantages. (1) Up to a quarter of emergency medical admissions do not have a blood test undertaken within 24 hours of admission and (2) it takes some time, usually an hour or so, for blood results to be reported. These disadvantages would delay automated assessment of sepsis risk. The advantage cNEWS is that it can trigger for sepsis screening as soon as the first set of physiological observations have been electronically recorded – which is usually within 30 minutes of admission for most patients.

There are a number of important limitations in relation to our study. We identified sepsis based on a validated optimised algorithm using ICD-10 codes (17). Nonetheless the extent to which differences between this approach to identifying sepsis and more recent consensus clinical definitions of sepsis (4) uphold or undermine the evaluation of cNEWS merits further study (7,28). Although our cNEWS models performed well in external validation, the 95%CI of the external calibration slope, despite adjusting for baseline differences in prevalence of sepsis between our two hospitals, did not include the ideal value of 1 – which therefore indicates some differences between observed and predicted risk of sepsis in the external dataset. Further work is required to understand why this is the case (eg it may be attributable to different ways of recording sepsis between our two hospitals). We used the index eNEWS data in our cNEWS models, which reflect the "on-admission" risk of sepsis of the patient. Nonetheless, eNEWS is repeatedly updated for each patient according to local hospital protocols, and the extent to which changes in eNEWS over time reflect changes in sepsis risk that need to be incorporated in our cNEWS models needs further study. Since cNEWS is based on NEWS and escalation protocols in hospitals are based on NEWS, work is required to determine how to successfully blend the risk estimates into existing escalation policies. For example, we could start by using a cut-off that is similar to the current threshold of NEWS 5+ (15,16) but note that this will increase the number of patients that will trigger for screening. An updated version of NEWS, known as NEWS2, has now been released (29) which includes a second oxygen scale for patients with proven type 2 respiratory failure and the inclusion of confusion (C) in the AVPU scale for conscious level ("alert, verbal, pain, unrespon-sive" scale becomes ACVPU). The extent to which these changes to NEWS enhance our cNEWS models requires further investiga-tion. So, a crucial next phase of this work is to field test cNEWS by carefully engineering it to build upon the current use of NEWS in routine clinical practice (30,31) to see if it does support the earlier detection and treatment of sepsis in emergency medical patients without unintended adverse consequences.

Conclusions

From the three cNEWS models, Model M2 is the most accurate. Since it places no additional data collection burden on clinicians and can be automated, it may now be carefully introduced and evaluated in hospitals with sufficient informatics infrastructure.

References

1. Ferrer R, Martin-Loeches I, Phillips G, Osborn TM, Townsend S, Dellinger RP, et al. Empiric Antibiotic Treatment Reduces Mortality in Severe Sepsis and Septic Shock From the First Hour. Crit Care Med [Internet]. 2014 Aug [cited 2018 Oct 24];42(8):1749–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24717459

2. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. Crit Care Med [Internet]. 2006 Jun [cited 2018 Oct 24];34(6):1589–96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16625125

3. Marik PE. Don't miss the diagnosis of sepsis! Crit Care [Internet]. BioMed Central; 2014 Sep 27 [cited 2019 Feb 24];18(5):529. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25675360

4. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. American Medical Association; 2016 Feb 23;315(8):801.

5. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Intensive Care Med. Springer-Verlag; 2003 Apr 28;29(4):530–8.

6. Serafim R, Gomes JA, Salluh J, Póvoa P. A Comparison of the Quick-SOFA and Systemic Inflammatory Response Syndrome Criteria for the Diagnosis of Sepsis and Prediction of Mortality. Chest [Internet]. 2018 Mar [cited 2019 Feb 24];153(3):646–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29289687

7. Churpek MM, Snyder A, Han X, Sokol S, Pettit N, Howell MD, et al. Quick Sepsis-related Organ Failure Assessment, Systemic Inflammatory Response Syndrome, and Early Warning Scores for Detecting Clinical Deterioration in Infected Patients outside the Intensive Care Unit. Am J Respir Crit Care Med. American Thoracic Society; 2017 Apr;195(7):906–11.

8. Goulden R, Hoyle M-C, Monis J, Railton D, Riley V, Martin P, et al. qSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. Emerg Med J. England; 2018 Jun;35(6):345–9.

9. Royal College of Physicians. National Early Warning Score (NEWS): Standardising the assessment of acuteillness severity in the NHS - Report of a working party. London: Royal College of Physicians; 2012.

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

11. Edwards M, McKay H, Van Leuvan C, Mitchell I. Modified Early Warning Scores: inaccurate summation or inaccurate assignment of score? Crit Care. BioMed Central; 2010;14(Suppl 1):P257.

12. Prytherch DR, Smith GB, Schmidt P, Featherstone PI, Stewart K, Knight D, et al. Calculating early warning scores--a classroom comparison of pen and paper and hand-held computer methods. Resuscitation. 2006 Aug;70(2):173–8.

13. Mohammed M, Hayton R, Clements G, Smith G, Prytherch D. Improving accuracy and efficiency of early warning scores in acute care. Br J Nurs. 18(1):18–24.

14. NHS Improvement. Patient Safety Measurement Unit. http://psmu.improvement.nhs.uk/workstreams/deterioration/news-baseline-survey.

15. National Institute for Health and Care Excellence. Sepsis Quality Standard [QS161]. Published September 2017 [Internet]. Available from: https://www.nice.org.uk/guidance/qs161/resources/sepsis-pdf-75545595402181

16. NHS England. Sepsis Guidance implementation advice for adults: NHS England, Sept 17 [Internet]. Available from: https://www.england.nhs.uk/wp-content/uploads/2017/09/sepsis-guidance-implementation-advice-for-adults.pdf

17. Jolley RJ, Quan H, Jetté N, Sawka KJ, Diep L, Goliath J, et al. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. BMJ Open. BMJ Publishing Group; 2015 Dec 23;5(12):e009487.

18. Jolley RJ, Sawka KJ, Yergens DW, Quan H, Jetté N, Doig CJ. Validity of administrative data in recording sepsis: a systematic review. Crit Care. BioMed Central; 2015 Dec 6;19(1):139.

19. StatCorp. Stata: Release 14. Statistical Software. College Station, TX: StataCorp LP. 2016.

20. Steyerberg EW. Clinical Prediction Models. A practical approach to development, validation and updating. Springer; 2008.

21. Faisal M, Howes R, Steyerberg EW, Richardson D, Mohammed MA. Using routine blood test results to predict the risk of death for emergency medical admissions to hospital: an external model validation study. QJM. Oxford University Press; 2017 Jan 1;110(1):27–31.

22. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JJ-CC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. BioMed Central; 2011;12(1):77.

23. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing http://www.r-project.org/. Vienna; 2015.

24. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010 Jan;21(1):128–38.

25. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005 Oct 15;21(20):3940–1.

26. Balamuth F, Alpern ER, Abbadessa MK, Hayes K, Schast A, Lavelle J, et al. Improving Recognition of Pediatric Severe Sepsis in the Emergency Department: Contributions of a Vital Sign-Based Electronic Alert and Bedside Clinician Identification. Ann Emerg Med. Elsevier; 2017 Dec 1;70(6):759–768.e2.

27. Faisal M, Scally A, Richardson D, Beatson K, Howes R, Speed K, et al. Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test Results. Crit Care Med. 2018 Jan 24;46(4):612–8.

28. Corfield AR, Lees F, Zealley I, Houston G, Dickie S, Ward K, et al. Utility of a single early warning score in patients with sepsis in the emergency department. 2012;

29. NHS. Royal College of Physicians: NHS England approves use of National Early Warning Score (NEWS) 2 to improve detection of acutely ill patients. [Internet]. 2017. Available from: https://www.rcplondon.ac.uk/news/nhs-england-approves-use-national-early-warning-score-news-2-improve-detection-acutely-ill

30. Escobar GJ, Dellinger RP. Early detection, prevention, and mitigation of critical illness outside intensive care settings. J Hosp Med. 2016 Nov;11:S5–10.

31. Escobar GJ, Turk BJ, Ragins A, Ha J, Hoberman B, LeVine SM, et al. Piloting electronic medical record-based early detection of inpatient deterioration in community hospitals. J Hosp Med. 2016 Nov;11:S18–24.