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

Muhammad Faisal PhD, Donald Richardson MBChB, Andrew J. Scally MSc, Robin Howes MSc, Kevin Beatson MSc, Kevin Speed MBBS MSc, Mohammed A. Mohammed PhD

ABSTRACT

BACKGROUND: In hospitals in England, patients' vital signs are monitored and summarized into the National Early Warning Score (NEWS); this score is more accurate than the Quick Sepsis-related Organ Failure Assessment (qSOFA) score at identifying patients with sepsis. We investigated the extent to which the accuracy of the NEWS is enhanced by developing and comparing 3 computer-aided NEWS (cNEWS) models (M0 = NEWS alone, M1 = M0 + age + sex, M2 = M1 + subcomponents of NEWS + diastolic blood pressure) to predict the risk of sepsis.

METHODS: We included all emergency medical admissions of patients 16 years of age and older discharged over 24 months from 2 acute care hospital centres (York Hospital [YH] for model development and a combined data set from 2 hospitals [Diana, Princess of Wales Hospital and Scunthorpe General Hospital] in the Northern Lincolnshire and Goole National Health Service Foundation Trust [NH] for external model validation). We used a validated Canadian method for defining sepsis from administrative hospital data.

RESULTS: The prevalence of sepsis was lower in YH (4.5%, 1596/35 807) than in NH (8.5%, 2983/35 161). 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). For NEWS of 5 or higher, sensitivity increased (YH: 47.24% v. 50.56% v. 52.69%; NH: 37.91% v. 43.35% v. 48.07%), the positive likelihood ratio increased (YH: 2.77 v. 2.99 v. 3.06; NH: 3.18 v. 3.32 v. 3.45) and the positive predictive value increased (YH: 11.44% v. 12.24% v. 12.49%; NH: 22.75% v. 23.55% v. 24.21%).

INTERPRETATION: From the 3 cNEWS models, model M2 is the most accurate. Given that it places no additional burden of data collection on clinicians and can be automated, it may now be carefully introduced and evaluated in hospitals with sufficient informatics infrastructure.

S epsis 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 NEWS is used to identify those at risk of death and increased morbidity in all patient diagnostic groups (with some noted exceptions, e.g., head injury). This score has been widely adopted in National Health Service (NHS) hospitals in England and other countries.[9]

The NEWS is derived from 7 physiologic variables or vital signs — respiration rate, oxygen saturations, any supplemental oxygen, temperature, systolic blood pressure, heart rate and level of consciousness as measured by the AVPU (alert, voice, pain, unresponsive) scale — that 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 1, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.181418/-/DC1, for further details on the NEWS).

The clinical rationale for the 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 or higher has been recommended as a trigger 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 the eNEWS and the potential of this system to support real-time computer-aided screening for sepsis, we investigated the extent to which the accuracy of the NEWS for predicting sepsis could be enhanced by developing computer-aided NEWS (cNEWS) models. 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 or calculation on clinicians, because cNEWS relies on data that are routinely collected as part of the process of care, are already stored in the patient's electronic health record and are accessible in real time, thus offering the prospects of real-time risk predictions without hindering clinical workflows.

We sought to examine the accuracy of 3 cNEWS models in predicting sepsis that include age, sex and the subcomponents of NEWS compared with a reference model that uses NEWS only, using a validated method for defining sepsis developed by Jolley and colleagues in Canada.[17]

Methods

Setting and data

Our cohorts comprised emergency medical admissions from 3 acute care hospitals located about 100 km apart in the Yorkshire & Humberside region of England — the Diana, Princess of Wales Hospital (about 400 beds) and Scunthorpe General Hospital (about 400 beds) managed by the Northern Lincolnshire and Goole NHS Foundation Trust (NLAG), and York Hospital (YH) (about 700 beds), managed by York Teaching Hospitals NHS Foundation Trust. For the purposes of this study, the 2 acute hospitals in NLAG are combined into a single data set and collectively referred to as NLAG Hospitals (NH). Both NH and YH use electronic NEWS scoring exclusively as part of their in-house electronic patient record systems. We selected these hospitals because these data have been collected as part of patients' process of care since at least 2013, and the hospitals were agreeable to the study.

We considered all patients 16 years of age and older with emergency medical admissions, discharged during a 24-month period (Jan. 1, 2014, to Dec. 31, 2015), with eNEWS data. For each emergency admission, we obtained a pseudoanonymized 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). The NEWS does not include diastolic blood pressure, but we incorporated it in our statistical models because this data item is routinely collected.

The NEWS scale ranges from 0 (indicating the lowest severity of illness) to a maximum of 20 (see Appendix 1 for further details). The index eNEWS was defined as the first score electronically recorded within ± 24 hours of the admission time. We excluded records in which the index eNEWS was not within ± 24 hours of recorded admission or for which eNEWS was not available (Appendix 1, Table S1 and S2).

We define sepsis (with at least 1 organ failure or septic shock)[4] based on 84 selected *International Classification of Diseases, 10th Revision* (ICD-10) codes identified by an optimized validated method reported by Jolley and colleagues in Canada[17] (which we adapted to our study by excluding 6 Canadian-specific ICD-10 codes and 3 procedure codes). We used this optimized approach for identifying sepsis using ICD-10 codes because other methods are known to underestimate sepsis from administrative data.[17,18]

Statistical analysis

We reported the statistical differences in characteristics of our 2 hospitals using 2 independent sample *t* test (for continuous data) and χ^2 proportion test (for categorical data).

We first performed exploratory analyses including scatter plots and box plots that showed the relation between covariates and risk of sepsis in our hospitals. We developed 3 logistic regression models for the risk of sepsis. The models (M0, M1 and M2) use the index or first-recorded eNEWS within ± 24 hours 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 (x^3, x^2, x^1, x, $\sqrt{x}$, $\log(x)$, x^{-1}, x^{-2}, x^{-3}) 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 data set). We then externally validated these models using data from the NH data set (external validation). 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 characteristic (ROC) curve (AUC) or C statistic after adjusting for differences in the baseline[21] risk of sepsis in our 2 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 relation 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 or less than 1, as this reflects problems of underprediction or overprediction.

The cut-off point of a NEWS of 5 or higher 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 and M2) at eNEWS thresholds of 4 or higher, 5 or higher, and 6 or higher.[25] For the best performing model (M2), we further analyzed its performance across a range of risks of sepsis (5%–15%) to highlight the performance characteristics of this model, which may inform choice of thresholds in routine clinical practice.

Analyses were carried out using the statistical language R,[23] the ROCR[25] library and Stata.[19]

Ethics approval

We obtained ethics approval for this study from Yorkshire & The Humber — Leeds West Research Ethics Committee (reference number 15/YH/0348).

Results

The number (YH: n = 36 751; NH: n = 37 100) of emergency medical admissions over a 24-month period was similar in YH and NH. We excluded 2.6% (944/36 751) of admissions in YH and 5.2% (1939/37 100) in NH because the index eNEWS was not recorded within 24 hours of the admission or there was no eNEWS recorded at all (Appendix 1, Tables S1 and S2).

The characteristics of the admissions included in our study are shown in Table 1. Patients with an emergency admission at YH were older than those at NH (67.8 yr v. 66.4 yr), less likely to be male (47.3% v. 49.8%), had higher index eNEWS (2.5 v. 2.1), and a much lower prevalence of sepsis (4.5% v. 8.5%) but similar in-hospital mortality (5.8% v. 5.4%). The prevalence of oxygen supplementation was lower at the YH than at the NH (11.3% v. 19.2%). See accompanying scatter and boxplots in Appendix 1, Figures S1 to S4. Figure 1 shows the relation between the index eNEWS with sepsis in each hospital. As the index eNEWS increases, so too does the risk of sepsis.

Model comparison

We compared the 3 cNEWS models described previously to predict the risk of sepsis, using data from YH for model development and data from NH for validation.

Table 1: Characteristics of emergency medical admissions in the development and validation data sets		
	No. (%) of admissions or mean ± SD	
Characteristic	**Development data set (YH)** **n = 35 807**	**Validation data set (NH)** **n = 35 161**
Patient age, yr*	67.8 ± 19.5	66.4 ± 19.5
Patient sex, male*	16 936 (47.3)	17 498 (49.8)
Sepsis, outcome*	1596 (4.5)	2983 (8.5)
In-hospital mortality†	2080 (5.8)	1900 (5.4)
Index eNEWS*	2.5 ± 2.6	2.1 ± 2.3
Alertness*		
Alert	34 769 (97.1)	34 503 (98.1)
Pain	243 (0.7)	126 (0.4)
Voice	607 (1.7)	435 (1.2)
Unconscious	188 (0.5)	97 (0.3)
Respiratory rate, breaths/min*	18.6 ± 4.8	18.1 ± 3.6
Temperature, °C*	36.3 ± 0.8	36.5 ± 0.7
Systolic pressure, mm Hg*	136 ± 27.3	129.4 ± 23
Diastolic pressure, mm Hg*	75.4 ± 15.5	74.9 ± 14.9
Pulse rate, beats/min*	85.6 ± 21.1	81.2 ± 17.8
Oxygen supplementation*	4053 (11.3)	6750 (19.2)
% Oxygen saturation*	96.3 ± 2.9	95.9 ± 3.0

Note: eNEWS = electronic National Early Warning Score, NH = NLAG Hospitals, NLAG = Northern Lincolnshire and Goole NHS Foundation Trust, SD = standard deviation, YH = York Hospital.
*p < 0.001.
†p = 0.3.

The likelihood ratio test showed significant improvement in model goodness of fits (M0 v. M1: $\chi^2 = 416.8$ [df = 2], $p < 0.001$; M1 v. M2: $\chi^2 = 161.8$ [df = 10], $p < 0.001$). The ROC plots for each model are shown in Figure 2, and the accompanying discrimination and calibration statistics are shown in Table 2. Model M0 had the lowest C statistic in the development (0.705) and validation data sets (0.708). Models M1 and M2 had higher C statistics (Table 2).

The external validation slope reduced from 1.18 (M0) to 1.15 (M2). The internal and external validation plots are shown in Appendix 1, Figure S5.

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

At the current recommended threshold of a NEWS of 5 or higher for screening for sepsis, sensitivity increased across models M0, M1 and M2 in the development data set (47.24% v. 50.56% v. 52.69%) and the external validation data set (37.91% v. 43.35% v. 48.07%, respectively). Specificity changed little in the development data set (82.94% v. 83.09% v. 82.77%) and external validation data set (88.07
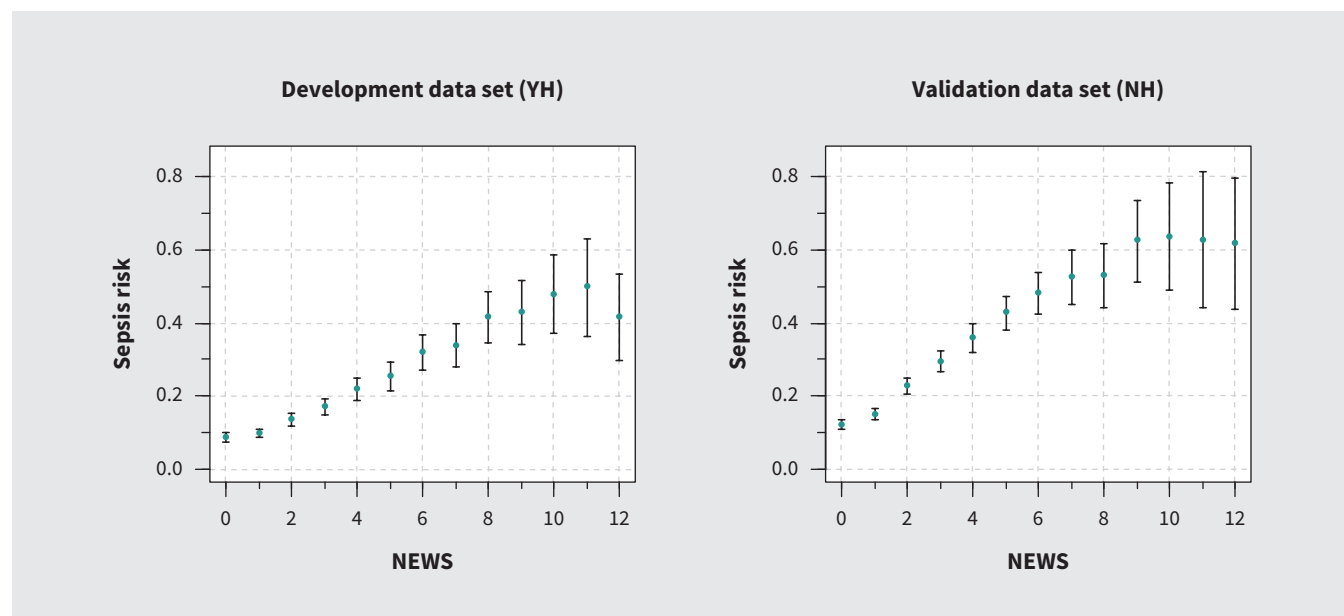
Figure 1: Observed sepsis risk versus index eNEWS in YH and NH hospitals. Vertical bars are exact binomial 95% confidence intervals. For visualization purposes, we capped NEWS at 12. Note: eNEWS = electronic National Early Warning Score, NH = NLAG Hospitals, NLAG = Northern Lincolnshire and Goole NHS Foundation Trust, YH = York Hospital.
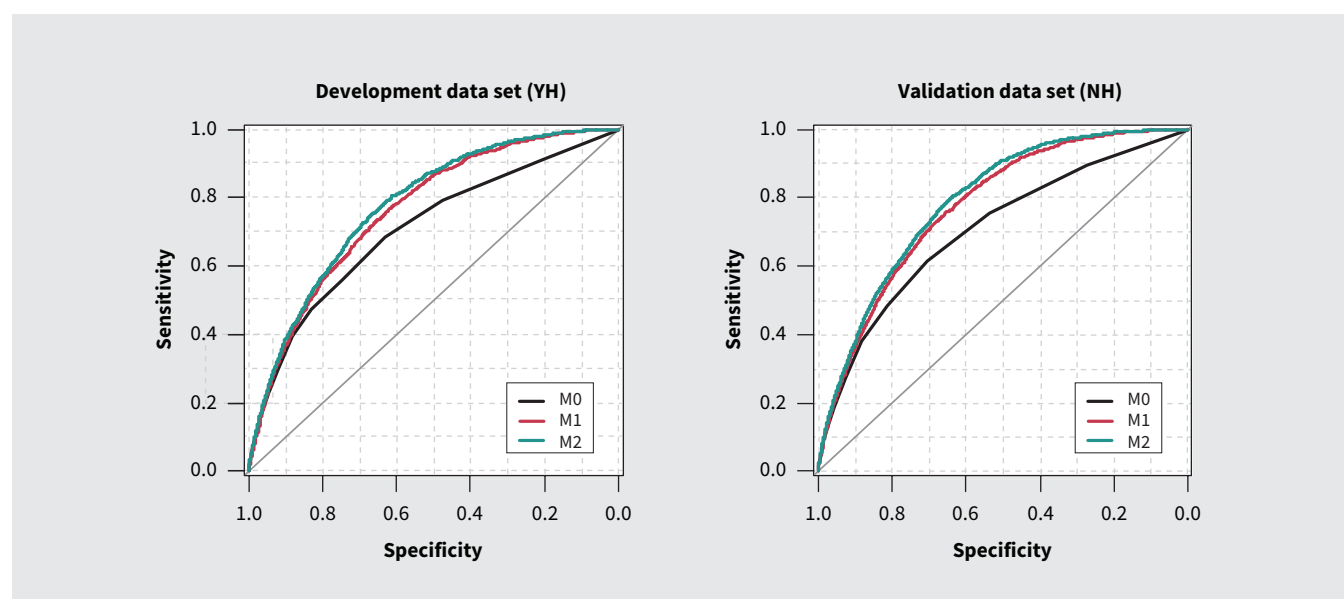

Figure 2: Receiver operating characteristic curve for 3 models (M0, M1 and M2) in predicting the risk of sepsis in the YH and NH hospitals. Note: NH = NLAG Hospitals, NLAG = Northern Lincolnshire and Goole NHS Foundation Trust, YH = York Hospital.

v. 86.96 v. 86.05). The positive likelihood ratio increased across models in the development data set (2.77 v. 2.99 v. 3.06) and the external validation data set (3.18 v. 3.32 v. 3.45). The negative likelihood ratio was reduced across models in the development data set (0.64 v. 0.59 v. 0.57) and the external validation data set (0.70 v. 0.65 v. 0.60). The positive predictive value increased in the development data set (11.44 v. 12.24 v. 12.49) and the external validation data set (22.75 v. 23.55 v. 24.21). The negative predictive value increased

Table 2: Performance of 3 computer-aided National Early Warning Score models in predicting the risk of sepsis in the development and validation data sets

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

Note: CI = confidence interval, eNEWS = electronic National Early Warning Score, NH = NLAG Hospitals, NLAG = Northern Lincolnshire and Goole NHS Foundation Trust, YH = York Hospital.

Table 3: Sensitivity, specificity and related analyses of M0, M1 and M2 to predict the risk of sepsis in the development and external validation data sets at selected eNEWS thresholds

NEWS threshold	Model	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Development data set, YH							
≥ 4 (equivalent predicted probability: YH = 0.051; NH = 0.11)	M0	55.95 (53.48–58.41)	74.91 (74.45–75.37)	9.42 (8.84–10.03)	97.33 (97.13–97.52)	2.23 (2.13–2.34)	0.59 (0.56–0.62)
	M1	60.96 (58.52–63.37)	75.53 (75.07–75.98)	10.41 (9.8–11.05)	97.65 (97.46–97.83)	2.49 (2.39–2.6)	0.52 (0.49–0.55)
	M2	63.22 (60.80–65.59)	75.55 (75.1–76.01)	10.77 (10.15–11.41)	97.78 (97.59–97.95)	2.59 (2.48–2.7)	0.49 (0.46–0.52)
≥ 5 (equivalent predicted probability: YH = 0.063; NH = 0.138)	M0	47.24 (44.77–49.73)	82.94 (82.54–83.34)	11.44 (10.68–12.23)	97.12 (96.92–97.31)	2.77 (2.62–2.93)	0.64 (0.61–0.67)
	M1	50.56 (48.08–53.05)	83.09 (82.69–83.49)	12.24 (11.46–13.06)	97.30 (97.11–97.48)	2.99 (2.83–3.16)	0.59 (0.57–0.63)
	M2	52.69 (50.21–55.17)	82.77 (82.37–83.17)	12.49 (11.71–13.3)	97.40 (97.21–97.58)	3.06 (2.9–3.22)	0.57 (0.54–0.60)
≥ 6 (equivalent predicted probability: YH = 0.079; NH = 0.169)	M0	39.22 (36.82–41.67)	88.1 (87.75–88.44)	13.33 (12.37–14.33)	96.88 (96.68–97.07)	3.30 (3.08–3.53)	0.69 (0.66–0.72)
	M1	40.23 (37.81–42.68)	88.22 (87.87–88.56)	13.74 (12.77–14.76)	96.94 (96.74–97.12)	3.41 (3.20–3.65)	0.68 (0.65–0.71)
	M2	42.04 (39.61–44.51)	88.13 (87.78–88.47)	14.18 (13.2–15.21)	97.02 (96.83–97.21)	3.54 (3.32–3.78)	0.66 (0.63–0.69)
External validation data set, NH							
≥ 4 (equivalent predicted probability: YH = 0.051; NH = 0.11	M0	48.58 (46.77–50.39)	81.29 (80.85–81.71)	19.39 (18.50–20.31)	94.46 (94.18–94.73)	2.60 (2.49–2.71)	0.63 (0.61–0.66)
	M1	57.09 (55.29–58.88)	79.77 (79.32–80.2)	20.73 (19.86–21.63)	95.25 (94.99–95.5)	2.82 (2.72–2.93)	0.54 (0.52–0.56)
	M2	59.50 (57.72–61.27)	79.38 (78.94–79.82)	21.11 (20.24–22.00)	95.48 (95.23–95.73)	2.89 (2.78–2.99)	0.51 (0.49–0.53)
≥ 5 (equivalent predicted probability: YH = 0.063; NH = 0.138)	M0	37.91 (36.17–39.68)	88.07 (87.71–88.42)	22.75 (21.59–23.94)	93.87 (93.59–94.13)	3.18 (3.01–3.36)	0.70 (0.69–0.73)
	M1	43.35 (41.56–45.15)	86.96 (86.58–87.32)	23.55 (22.43–24.70)	94.30 (94.03–94.57)	3.32 (3.16–3.49)	0.65 (0.63–0.67)
	M2	48.07 (46.27–49.88)	86.05 (85.67–86.43)	24.21 (23.12–25.32)	94.70 (94.44–94.96)	3.45 (3.29–3.61)	0.60 (0.58–0.62)
≥ 6 (equivalent predicted probability: YH = 0.079; NH = 0.169)	M0	27.49 (25.89–29.13)	92.65 (92.36–92.93)	25.74 (24.23–27.29)	93.24 (92.95–93.51)	3.74 (3.49–4.01)	0.78 (0.77–0.80)
	M1	31.38 (29.71–33.08)	91.91 (91.61–92.21)	26.46 (25.01–27.94)	93.53 (93.25–93.80)	3.88 (3.64–4.14)	0.75 (0.73–0.77)
	M2	35.67 (33.95–37.42)	90.80 (90.48–91.11)	26.44 (25.08–27.83)	93.84 (93.56–94.1)	3.88 (3.65–4.11)	0.71 (0.69–0.73)

Note: CI = confidence interval, eNEWS = electronic National Early Warning Score, N+ = the number of positive cases, NH = NLAG Hospitals, NLAG = Northern Lincolnshire and Goole NHS Foundation Trust, YH = York Hospital.

slightly in the development data set from (97.12 v. 97.30 v. 97.40) and the external validation data set (93.87 v. 94.30 v. 94.70).

We further explored the behaviour of the best performing cNEWS model (M2) across a range of cut-off probabilities (5%–15%) (Table 4). For YH, a cut-off point of 0.06 appears to offer reasonable performance; likewise, 0.12 appears to be a reasonable cut-off point for NH.

Interpretation

We developed 3 computer-aided versions of eNEWS models, which incorporated progressively more information, and found that M2 (M1 + subcomponents of NEWS + diastolic blood pressure) was the best model to predict sepsis. This is unsurprising because it incorporates additional information about the patient's age and other vital signs.

The main advantages of these cNEWS models is that they are designed to incorporate data that are already available in the patient's electronic health record and so place no additional data collection or computational burden on clinicians, and are readily automated. Nonetheless, we note that computer-aided risk scores are not designed to replace clinical judgment. They are intended and designed to support, not undermine, clinical decision-making and can be overridden by clinical concern.[9,26] However, cNEWS scores 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 Score,[27] which is based on physiologic variables and blood results, offers more accuracy in predicting sepsis than cNEWS, it has 2 main disadvantages. First, up to one-quarter of emergency medical admissions do not have a blood test undertaken within 24 hours of admission; second, there is usually a delay between drawing blood and reporting of results. These disadvantages would delay automated assessment of sepsis risk. The advantage of cNEWS is that it can trigger for sepsis

Table 4: Sensitivity, specificity and related analyses of M2 to predict the risk of sepsis in the development and external validation data sets at range of predicted probabilities cut-off points

Predicted probabilities cut-off point	N+	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Development data set, YH							
0.05	9558	63.85 (61.43–66.21)	75.04 (74.58–75.50)	10.66 (10.05–11.30)	97.80 (97.62–97.98)	2.56 (2.45–2.67)	0.48 (0.45–0.51)
0.06*	7337	55.20 (52.72–57.66)	81.13 (80.71–81.54)	12.01 (11.27–12.77)	97.49 (97.30–97.67)	2.93 (2.78–3.07)	0.55 (0.52–0.58)
0.07	5753	47.24 (44.77–49.73)	85.39 (85.01–85.76)	13.11(12.24–14.01)	97.20 (97.01–97.38)	3.23 (3.05–3.43)	0.62 (0.59–0.65)
0.08	4639	41.79 (39.36–44.26)	88.39 (88.05–88.73)	14.38 (13.38–15.42)	97.02 (96.82–97.21)	3.60 (3.37–3.84)	0.66 (0.63–0.69)
0.09	3834	36.97 (34.59–39.39)	90.52 (90.20–90.83)	15.39 (14.26–16.57)	96.85 (96.66–97.04)	3.90 (3.63–4.19)	0.70 (0.67–0.72)
0.10	3243	32.39 (30.10–34.75)	92.03 (91.74–92.32)	15.94 (14.70–17.25)	96.69 (96.49–96.88)	4.07 (3.75–4.40)	0.73 (0.71–0.76)
0.11	2677	28.13 (25.94–30.41)	93.49 (93.22–93.75)	16.77 (15.38–18.24)	96.54 (96.34–96.73)	4.32 (3.96–4.72)	0.77 (0.75–0.79)
0.12	2265	24.87 (22.77–27.07)	94.54 (94.29–94.78)	17.53 (15.98–19.16)	96.43 (96.22–96.62)	4.56 (4.14–5.01)	0.79 (0.77–0.82)
0.13	1913	21.68 (19.68–23.78)	95.42 (95.19–95.64)	18.09 (16.39–19.89)	96.31 (96.11–96.51)	4.73 (4.26–5.26)	0.82 (0.80–0.84)
0.14	1644	19.24 (17.33–21.26)	96.09 (95.88–96.29)	18.67 (16.82–20.64)	96.23 (96.02–96.43)	4.92 (4.39–5.51)	0.84 (0.82–0.86)
0.15	1407	16.92 (15.11–18.85)	96.68 (96.48–96.86)	19.19 (17.16–21.35)	96.15 (95.94–96.35)	5.09 (4.50–5.76)	0.86 (0.84–0.88)
External validation data set, NH							
0.05	20 864	93.66 (92.73–94.51)	43.84 (43.30–44.39)	13.39 (12.93–13.86)	98.68 (98.48–98.86)	1.67 (1.65–1.69)	0.14 (0.13–0.17)
0.06	18 160	89.78 (88.63–90.84)	51.89 (51.34–52.43)	14.75 (14.23–15.27)	98.21 (98.00–98.40)	1.87 (1.84–1.90)	0.20 (0.18–0.22)
0.07	15 471	83.04 (81.64–84.37)	59.62 (59.08–60.16)	16.01 (15.44–16.60)	97.43 (97.20–97.65)	2.06 (2.01–2.10)	0.28 (0.26–0.31)
0.08	13 331	77.98 (76.44–79.45)	65.80 (65.28–66.32)	17.45 (16.81–18.10)	96.99 (96.76–97.21)	2.28 (2.23–2.34)	0.33 (0.31–0.36)
0.09	11 477	71.61 (69.95–73.22)	70.97 (70.47–71.47)	18.61 (17.90–19.34)	96.42 (96.18–96.66)	2.47 (2.40–2.54)	0.40 (0.38–0.42)
0.10	9956	65.87 (64.14–67.58)	75.17 (74.69–75.64)	19.74 (18.96–20.53)	95.96 (95.71–96.20)	2.65 (2.57–2.74)	0.45 (0.43–0.48)
0.11	8644	60.41 (58.63–62.17)	78.74 (78.29–79.18)	20.85 (19.99–21.72)	95.55 (95.29–95.79)	2.84 (2.74–2.94)	0.50 (0.48–0.53)
0.12*	7540	55.88 (54.08–57.68)	81.75 (81.32–82.17)	22.11 (21.18–23.06)	95.24 (94.98–95.48)	3.06 (2.94–3.18)	0.54 (0.52–0.56)
0.13	6556	51.39 (49.58–53.20)	84.39 (83.99–84.78)	23.38 (22.36–24.43)	94.93 (94.67–95.18)	3.29 (3.15–3.44)	0.58 (0.55–0.60)
0.14	5776	47.07 (45.26–48.88)	86.41 (86.03–86.79)	24.31 (23.21–25.44)	94.63 (94.36–94.88)	3.46 (3.31–3.63)	0.61 (0.59–0.63)
0.15	5086	42.74 (40.96–44.54)	88.16 (87.80–88.51)	25.07 (23.88–26.28)	94.32 (94.05–94.58)	3.61 (3.43–3.80)	0.65 (0.63–0.67)

Note: N+ = the number of positive cases, NH = NLAG Hospitals, NLAG = Northern Lincolnshire and Goole NHS Foundation Trust, YH = York Hospital.
*Tentatively indicates a reasonable cut-off choice.

screening as soon as the first set of physiologic observations has been electronically recorded — usually within 30 minutes of admission for most patients.

Limitations

There are several important limitations of our study. We identified sepsis based on a validated optimized 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% confidence interval of the external calibration slope, despite adjustment for baseline differences in prevalence of sepsis between our 2 hospitals, did not include the ideal value of 1; this indicates some differences between observed and predicted risk of sepsis in the external data set. Further work is required to understand why this is the case (e.g., it may be attributable to different ways of recording sepsis between the 2 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 point that is similar to the current threshold of a NEWS of 5 or higher,[15,16] but 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, voice, pain, unresponsive" scale becomes ACVPU). The extent to which these changes to NEWS enhance our cNEWS models requires further investigation. A crucial next phase of this work is to field test cNEWS by carefully engineering it to build on 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.

Conclusion

From the 3 cNEWS models evaluated, we found model M2 to be the most accurate. Given that using the model 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, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour. *Crit Care Med* 2014;42:1749–55. Available: www.ncbi.nlm.nih.gov/pubmed /24717459 (accessed 2018 Oct. 24).

2. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. *Crit Care Med* 2006;34:1589–96. Available: www.ncbi.nlm.nih.gov/pubmed/16625125 (accessed 2018 Oct. 24).

3. Marik PE. Don't miss the diagnosis of sepsis! *Crit Care* 2014;18:529. Available: www.ncbi.nlm.nih.gov/pubmed/25675360 (accessed 2019 Feb. 24).

4. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; 315:801-10.

5. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Med* 2003;29:530–8.

6. Serafim R, Gomes JA, Salluh J, et al. A comparison of the Quick-SOFA and Systemic Inflammatory Response Syndrome criteria for the diagnosis of sepsis and prediction of mortality. *Chest* 2018;153:646–55. Available: www.ncbi.nlm. nih.gov/pubmed/29289687 (accessed 2019 Feb. 24).

7. Churpek MM, Snyder A, Han X, 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* 2017;195:906-11.

8. Goulden R, Hoyle M-C, Monis J, et al. qSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. *Emerg Med J* 2018;35:345-9.

9. *National Early Warning Score (NEWS): Standardising the assessment of acute illness severity in the NHS — report of a working party*. London: Royal College of Physicians; 2012.

10. Smith GB, Prytherch DR, Schmidt P, et al. Hospital-wide physiological surveillance — a new approach to the early identification and management of the sick patient. *Resuscitation* 2006;71:19-28.

11. Edwards M, McKay H, Van Leuvan C, et al. Modified early warning scores: Inaccurate summation or inaccurate assignment of score? *Crit Care* 2010; 14(Suppl 1):257.

12. Prytherch DR, Smith GB, Schmidt P, et al. Calculating early warning scores — a classroom comparison of pen and paper and hand-held computer methods. *Resuscitation* 2006;70:173-8.

13. Mohammed M, Hayton R, Clements G, et al. Improving accuracy and efficiency of early warning scores in acute care. *Br J Nurs* 2009;18:18-24.

14. NHS Improvement. Patient Safety Measurement Unit. United Kingdom: National Health Service; 2019. Available: http://psmu.improvement.nhs.uk/ workstreams/deterioration/news-baseline-survey (accessed 2017 January).

15. *Sepsis: quality standard*. London (UK) National Institute for Health and Care Excellence; 2017. Available: www.nice.org.uk/guidance/qs161/resources/ sepsis-pdf-75545595402181 (accessed 2017 January).

16. *Sepsis guidance implementation advice for adults*. England: National Health Service; 2017. Available: www.england.nhs.uk/wp-content/uploads/2017/09/sepsis -guidance-implementation-advice-for-adults.pdf (accessed 2017 January).

17. Jolley RJ, Quan H, Jetté N, et al. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. *BMJ Open* 2015;5:e009487.

18. Jolley RJ, Sawka KJ, Yergens DW, et al. Validity of administrative data in recording sepsis: a systematic review. *Crit Care* 2015;19:139.

19. 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*. New York: Springer-Verlag New York; 2008.

21. Faisal M, Howes R, Steyerberg EW, et al. Using routine blood test results to predict the risk of death for emergency medical admissions to hospital: an external model validation study. *QJM* 2017;110:27–31.

22. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.

23. R: a language and environment for statistical computing. Vienna (Austria): R Project for Statistical Computing; 2015. Available: www.r-project.org (accessed 2017 January).

24. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128-38.

25. Sing T, Sander O, Beerenwinkel N, et al. ROCR: visualizing classifier performance in R. *Bioinformatics* 2005;21:3940-1.

26. Balamuth F, Alpern ER, Abbadessa MK, 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* 2017;70:759–68.e2.

27. Faisal M, Scally A, Richardson D, 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;46:612-8.

28. Corfield AR, Lees F, Zealley I, et al. Utility of a single early warning score in patients with sepsis in the emergency department. *Emerg Med J* 2014;31:482-7.

29. *National Early Warning Score (NEWS) 2: Standardising the assessment of acute illness severity in the NHS.* Updated report of a working party. London: Royal College of Physicians; 2017

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

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