CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Development and validation of a novel computer-aided score to predict the risk of in-hospital mortality for acutely ill medical admissions in two acute hospitals using their first electronically recorded blood test results and vital signs: a cross-sectional study
Authors	Faisal, Muhammad;Scally, Andrew;Jackson, Natalie;Richardson, Donald;Beatson, Kevin;Howes, Robin;Speed, Kevin;Menon, Madhav;Daws, Jeremey;Dyson, Judith;Marsh, Claire;Mohammed, Mohammed A.
Publication date	2018-12-06
Original Citation	Faisal, M., Scally, A.J., Jackson, N., Richardson, D., Beatson, K., Howes, R., Speed, K., Menon, M., Daws, J., Dyson, J. and Marsh, C., 2018. Development and validation of a novel computer-aided score to predict the risk of in-hospital mortality for acutely ill medical admissions in two acute hospitals using their first electronically recorded blood test results and vital signs: a cross-sectional study. BMJ open, 8(12): e022939, (8pp.) DOI: 10.1136/bmjopen-2018-022939
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://bmjopen.bmj.com/content/8/12/e022939 - 10.1136/bmjopen-2018-022939
Rights	© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ. This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/. - http://creativecommons.org/licenses/by-nc/4.0/
Download date	2026-11-16 19:11:17

Item downloaded from	https://hdl.handle.net/10468/8087



BMJ Open

Development and validation of a novel computer-aided score to predict the risk of in-hospital mortality for acutely ill medical admissions in two acute hospitals using their first electronically recorded blood test results and vital signs: a cross-sectional study

Muhammad Faisal,[1,2] Andrew J Scally,[3] Natalie Jackson,[2] Donald Richardson,[4] Kevin Beatson,[4,5] Robin Howes,[6] Kevin Speed,[7] Madhav Menon,[7] Jeremey Daws,[7] Judith Dyson,[8] Claire Marsh,[2] Mohammed A Mohammed[1,2]

To cite: Faisal M, Scally AJ, Jackson N, *et al*. Development and validation of a novel computer-aided score to predict the risk of in-hospital mortality for acutely ill medical admissions in two acute hospitals using their first electronically recorded blood test results and vital signs: a cross-sectional study. *BMJ Open* 2018;**8**:e022939. doi:10.1136/bmjopen-2018-022939

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-022939).

Received 15 March 2018
Revised 24 September 2018
Accepted 16 October 2018

Check for updates

© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Mohammed A Mohammed; m.a.mohammed5@bradford.ac.uk

ABSTRACT

Objectives There are no established mortality risk equations specifically for emergency medical patients who are admitted to a general hospital ward. Such risk equations may be useful in supporting the clinical decision-making process. We aim to develop and externally validate a computer-aided risk of mortality (CARM) score by combining the first electronically recorded vital signs and blood test results for emergency medical admissions.

Design Logistic regression model development and external validation study.

Setting Two acute hospitals (Northern Lincolnshire and Goole NHS Foundation Trust Hospital (NH)—model development data; York Hospital (YH)—external validation data).

Participants Adult (aged ≥16 years) medical admissions discharged over a 24-month period with electronic National Early Warning Score(s) and blood test results recorded on admission.

Results The risk of in-hospital mortality following emergency medical admission was 5.7% (NH: 1766/30 996) and 6.5% (YH: 1703/26 247). The C-statistic for the CARM score in NH was 0.87 (95% CI 0.86 to 0.88) and was similar in an external hospital setting YH (0.86, 95% CI 0.85 to 0.87) and the calibration slope included 1 (0.97, 95% CI 0.94 to 1.00).

Conclusions We have developed a novel, externally validated CARM score with good performance characteristics for estimating the risk of in-hospital mortality following an emergency medical admission using the patient's first, electronically recorded, vital signs and blood test results. Since the CARM score places no additional data collection burden on clinicians and is readily automated, it may now be carefully introduced and evaluated in hospitals with sufficient informatics infrastructure.

Strengths and limitations of this study

- ► This study provides a novel computer-aided risk of mortality (CARM) score by combining the first electronically recorded vital signs and blood test results for emergency medical admissions.
- ► CARM is externally validated and places no additional data collection burden on clinicians and is readily automated.
- ► About 20%–30% of admissions do not have both National Early Warning Score(s) and blood test results and so CARM is not applicable to these admissions.

INTRODUCTION

Unplanned or emergency medical admissions to hospital involve patients with a broad spectrum disease and illness severity.[1] The appropriate early assessment and management of such admissions can be a critical factor in ensuring high-quality care.[2] A number of scoring systems have been developed which may support this clinical decision-making process, but few have been externally validated.[1] We propose to develop a computer-aided risk of in-hospital mortality score, following emergency medical admission that automatically combines two routinely collected, electronically recorded, clinical datasets—vital signs and blood test results. There is some evidence to suggest that the results of routinely undertaken blood tests and/or vital signs data may be useful in predicting the risk of death.[1]

In the UK National Health Service (NHS), the patient's vital signs are monitored and summarised into a National Early Warning Score(s) (NEWS) that is mandated by the

BMJ Open: first published as 10.1136/bmjopen-2018-022939 on 6 December 2018. Downloaded from http://bmjopen.bmj.com/ on 15 May 2019 by guest. Protected by copyright.

Royal College of Physicians (London).[3] NEWS is derived from six 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 for all patients, and then repeated thereafter depending on local hospital protocols.[3] The use of NEWS is relevant because 'patients die not from their disease but from the disordered physiology caused by the disease'.[4] 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 online supplementary material 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 are highly reliable and accurate when compared with paper-based methods.[5–8]

Blood tests are an integral part of clinical medicine, and are routinely undertaken during a patient's stay in hospital. Typically, routine blood tests consist of a core list of seven biochemical and haematological tests (albumin, creatinine, potassium, sodium, urea, haemoglobin, white blood cell count) and, in the absence of contraindications and subject to patient consent, almost all patients admitted to hospital undergo these tests on admission. Furthermore, in the UK NHS creatinine blood test results are now used to identify patients at risk of acute kidney injury (AKI),[9] which is an important cause of avoidable patient harm.[10]

In this paper, we investigate the extent to which the vital signs and blood test results of acutely ill patients can be used to predict the risk of in-hospital mortality following emergency admission to hospital. Our aim is to develop and validate an automated, computer-aided risk of mortality (CARM) model, using the patient's first, electronically recorded, vital signs and blood test results, which are usually available within a few hours of emergency admission without requiring any additional data items or prompts from clinicians. CARM, therefore, is designed for use in hospitals with sufficient informatics infrastructure.

METHODS

Setting and data

Our cohorts of emergency medical admissions are from three acute hospitals which are approximately 100 km apart in the Yorkshire and 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). The data from the two acute hospitals from NLAG are combined because this reflects how the hospitals are managed and are referred to as NLAG Hospitals (NH), which essentially places our study in two acute hospitals. Our study hospitals (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 chose 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 did not approach any other hospital.

We considered all adult (aged ≥16 years) emergency medical admissions, discharged during a 24-month period (1 January 2014 to 31 December 2015), with blood test results and NEWS. For each admission, we obtained a pseudonymised patient identifier, the patient's age (years), sex (male/female), discharge status (alive/dead), admission and discharge date and time and electronic NEWS. The NEWS ranged from 0 (indicating the lowest severity of illness) to 19 (the maximum NEWS value possible is 20). The admission/discharge date and electronically recorded NEWS are date and time stamped and the index NEWS was defined as the first electronically recorded score within ±24 hours of the admission time. The first blood test results were defined as the first full set of blood test results recorded within 4 days (96 hours) of admission (>90% of blood test results were within ±24 hours of admission—see online supplementary table S1).

For model development purposes, we were unable to consider emergency admissions without complete blood test results and NEWS recorded—this constituted 16.5% (6104/37 100) of records in NH and 28.6% (10 504/36 751) of records in YH. We excluded records for the following reasons: (1) records where the first NEWS was after 24 hours of admission and/or (2) where the first blood test was after 4 days of admission because these 'delayed' data were considered less likely to reflect the sickness profile of patients on admission. Moreover, the time from admission to first blood test results was usually several hours earlier than the actual time of admission because blood tests can be ordered in the emergency department before formal admission (see online supplementary figure S1).

Development of a CARM score

We began with exploratory analyses including line plots and box plots that showed the relationship between covariates and risk of in-hospital death in our hospitals. We developed a logistic regression model, known as CARM, to predict the risk of in-hospital death with the following covariates: age (years), sex (male/female), NEWS (including its components, plus diastolic blood pressure, as separate covariates), blood test results (albumin, creatinine, haemoglobin, potassium, sodium, urea and white cell count) and AKI score. The primary rationale for using these variables is that they are routinely collected as part of process of care and their inclusion in our statistical models is on clinical grounds as opposed

BMJ Open: first published as 10.1136/bmjopen-2018-022939 on 6 December 2018. Downloaded from http://bmjopen.bmj.com/ on 15 May 2019 by guest. Protected by copyright.

to the statistical significance of any given covariate. The widespread use of these variables in routine clinical care means that our model is more likely to be generalisable to other settings.

We used the *qladder* function (Stata, StatCorp, 2014), which displays the quantiles of transformed variable against the quantiles of a normal distribution according to the ladder powers $\left(x^3, x^2, x^1, x, \sqrt{x}, \log\left(x\right), x^{-1}, x^{-2}, x^{-3}\right)$ for each variable continious covariate and chose the following transformations: $(\text{creatinine})^{-1/2}$, $\log_e$(potassium), $\log_e$(white cell count), $\log_e$(urea), $\log_e$(respiratory rate), $\log_e$(pulse rate), $\log_e$(systolic blood pressure) and $\log_e$(diastolic blood pressure). We used an automated approach to search for all two-way interactions and incorporated those interactions which were statistically significant (p<0.001) implemented in the MASS library[11] in R.[12]

We developed the CARM model to predict the risk of in-hospital mortality following emergency medical admission using data from NH (the development dataset) and we externally validated this model, reporting discrimination and calibration characteristics,[13] using data from another hospital (YH) (the external validation dataset). The data from YH are not used for model development but as an external validation dataset only. We internally validated the CARM using a bootstrapping method that is implemented in the *rms* library[14] in R to estimate statistical optimism.[13 14]

Discrimination relates to how well a model can separate (or discriminate between), those who died and those who did not. Calibration measures a model's ability to generate predictions that are on average close to the average observed outcome. Overall statistical performance was assessed using the scaled Brier score, which incorporates both discrimination and calibration.[13] The Brier score is the squared difference between actual outcomes and predicted risk of death, scaled by the maximum Brier score such that the scaled Brier score ranges from 0% to 100%. Interpretation of the scaled Brier score is similar to R^2. Higher values indicate superior models. Calibration is the relationship between the observed and predicted risk of death 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'.[15] At model development, a=0 and b=1, but at 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.[16]

The concordance statistic (C-statistic) is a commonly used measure of discrimination. For a binary outcome, the C-statistic is the area under the receiver operating characteristics (ROC) curve. The ROC curve is a plot of the sensitivity (true positive rate) versus 1–specificty (false positive rate), for consecutive predicted risks.[13] The area under the ROC curve is interpreted as the probability that a deceased patient has a higher predicted risk of death than a randomly chosen non-deceased patient. A C-statistic of 0.5 is no better than tossing a coin, while a perfect model has a C-statistic of 1. The higher the C-statistic, the better the model. In general, values <0.7 are considered to show poor discrimination, values of 0.7–0.8 can be described as reasonable and values >0.8 suggest good discrimination.[17] The 95% CI for the C-statistic was derived using DeLong's method as implemented in the *pROC* library[18] in R.[12] Box plots showing the risk of death for those discharged alive and dead are a simple way to visualise the discrimination of each model. The difference in the mean predicted risk of death for those who were discharged alive and dead is a measure of the discrimination slope. The higher the slope, the better the discrimination.[13] We followed the TRIPOD guidelines for model development and validation.[19] All analyses were carried using R[12] and Stata.

Patient and public involvement

A workshop with a patient and service user group, linked to the University of Bradford, was involved at the start of this project to co-design the agenda for the patient and staff focus groups, which were subsequently held at each hospital site. Patients were invited to attend the patient focus group through existing patient and public involvement groups. The criteria used for recruitment to these focus group was any member of the public who had been a patient or carer in the last 5 years. The patient and public voice continued to be included throughout the project with three patient representatives invited to sit on

Table 1 Number and mortality of emergency medical admissions included/excluded

	Development dataset	Validation dataset
Characteristic	**N (%)**	**N (%)**
Total emergency medical admissions	37 100	36 751
Excluded: no NEWS recorded (%)	1305 (3.5)	772 (2.1)
Excluded: first NEWS after 24 hours of admission (%)	634 (1.7)	172 (0.5)
Excluded: first blood test results after 4 days of admission (%)	464 (1.3)	673 (1.8)
Excluded: no or incomplete blood test results recorded (%)	3701 (10.0)	8887 (24.2)
Total excluded (%)	6104 (16.5)	10 504 (28.6)
Total included (%)	30 996 (83.5)	26 247 (71.4)

NEWS, National Early Warning Score(s).

BMJ Open: first published as 10.1136/bmjopen-2018-022939 on 6 December 2018. Downloaded from http://bmjopen.bmj.com/ on 15 May 2019 by guest. Protected by copyright.

Table 2 Characteristics of emergency admissions for development and validation datasets

	Development dataset (NH)		Validation dataset (YH)	
Characteristic	Discharged alive	Discharged died	Discharged alive	Discharged died
N	29230	1766	24544	1703
Median length of stay (days) (IQR)	4.3 (8.3)	8.3 (13.3)	3.9 (7.7)	8.1 (14.1)
Male (%)	14557 (49.8)	887 (50.2)	11646 (47.5)	845 (49.6)
Mean NEWS (SD)	2.1 (2.2)	4.5 (3.2)	2.5 (2.5)	5.0 (3.6)
Alertness				
Alert (%)	28788 (98.5)	1613 (91.3)	23953 (97.6)	1503 (88.3)
Pain (%)	80 (0.3)	31 (1.8)	131 (0.5)	49 (2.9)
Voice (%)	315 (1.1)	83 (4.7)	357 (1.5)	106 (6.2)
Unconscious (%)	47 (0.2)	39 (2.2)	103 (0.4)	45 (2.6)
AKI score				
0 (%)	27063 (92.6)	1326 (75.1)	22133 (90.2)	936 (55.0)
1 (%)	1358 (4.7)	204 (11.6)	1482 (6.0)	451 (26.5)
2 (%)	429 (1.5)	129 (7.3)	369 (1.5)	191 (11.2)
3 (%)	380 (1.3)	107 (6.1)	560 (2.3)	125 (7.3)
Oxygen supplementation (%)	5364 (18.4)	900 (51.0)	2549 (10.4)	582 (34.2)
Mean age (years) (SD)	66.2 (19.5)	79.8 (11.1)	67.5 (19.4)	80 (11.7)
Mean albumin (g/L) (SD)	33.7 (5.9)	27.3 (6.4)	38.2 (5.7)	32.9 (6)
Mean creatinine (µmol/L) (SD)	103.3 (78.2)	148.9 (124.4)	100.8 (90.6)	138.7 (119)
Mean haemoglobin (g/L) (SD)	127.8 (22.2)	117.1 (22.8)	125.2 (22)	117.1 (23.2)
Mean potassium (mmol/L) (SD)	4.1 (0.6)	4.3 (0.8)	4.3 (0.6)	4.4 (0.8)
Mean sodium (mmol/L) (SD)	137 (5.1)	136 (7)	136.6 (4.6)	136.1 (6.2)
Mean white cell count (10^9 cells/L) (SD)	9.8 (6.5)	13.2 (13.3)	10.2 (10.7)	13.9 (21.1)
Mean urea (mmol/L) (SD)	7.5 (5.6)	14.1 (10.5)	7.8 (5.6)	13.3 (8.9)
Mean respiratory rate (breaths per min) (SD)	18 (3.5)	20.1 (4.8)	18.6 (4.6)	21.7 (6.8)
Mean temperature (°C) (SD)	36.5 (0.7)	36.3 (0.8)	36.3 (0.8)	36.1 (1.1)
Mean systolic pressure (mm Hg) (SD)	129.6 (22.7)	119.8 (24.8)	136.1 (27.2)	128.5 (30.3)
Mean diastolic pressure (mm Hg) (SD)	75 (14.8)	69.5 (15.8)	75.4 (15.5)	71.3 (17.7)
Mean pulse rate (beats per min) (SD)	81.3 (17.7)	86.5 (19.7)	86.2 (20.9)	92.1 (23.3)
Mean % oxygen saturation (SD)	96.0 (2.9)	94.6 (4.7)	96.3 (2.9)	95.0 (4.4)

AKI, acute kidney injury; NEWS, National Early Warning Score(s); NH, Northern Lincolnshire and Goole NHS Foundation Trust Hospital; YH, York Hospital.

the project steering group. Participants will be informed of the results of this study through the patient and public involvement leads at each hospital site and the project team have met with the Bradford patient and service user group to discuss the results.

RESULTS

Cohort description

We considered emergency medical admissions in each hospital (NH: n=37100, YH: n=36751) over the 24-month period. Of these, 16.5% (6104/37 100) in NH and 28.6% (10 504/36 751) in YH were not eligible for our study because they did not have NEWS recorded within ±24 hours of admission and/or full complement of blood test results within ±96 hours of admission (see table 1, online supplementary table S1 and figure S1). At YH, 24.2% of records were excluded because no or incomplete blood test results were recorded compared with only 10% in NH. Exclusions due to lack of NEWS data were less marked between YH and NH (see online supplementary table S2 for characteristic of emergency admissions with incomplete data).

The in-hospital mortality was 5.7% (1766/30 996) in NH and 6.5% (1703/26 247) in YH. The age, sex, NEWS and blood test results profile is shown table 2. Admissions in YH were older, with higher NEWS, higher AKI scores (AKI stage 3 is more common than stage 2 in YH) but higher albumin blood test results than NH. YH has a renal unit whereas NH does not.

BMJ Open: first published as 10.1136/bmjopen-2018-022939 on 6 December 2018. Downloaded from http://bmjopen.bmj.com/ on 15 May 2019 by guest. Protected by copyright.

Table 3 Comparing calibration and discrimination of CARM model to predict in-hospital mortality in development and validation datasets

Dataset	Mean predicted risk: alive	Mean predicted risk: died*	Discrimination slope†	Scaled Brier score	AUC (95% CI)	Median imputed AUC (95% CI)
Development dataset	0.047	0.229	0.183	0.175	0.874‡ (0.866 to 0.881)	0.915 (0.888 to 0.941)
Validation dataset	0.053	0.231	0.178	0.165	0.861 (0.852 to 0.869)	0.900 (0.880 to 0.919)

*Died in-hospital following emergency admission.
†Mean predicted risk difference between who discharged died and discharged alive.
‡Corrected optimism (original=0.874 and corrected=0.873).
AUC, area under the curve; CARM, computer-aided risk of mortality.

Online supplementary figures S2 to S5 show box plots and line plots for each continuous (untransformed) covariate that was included in the CARM model for NH and YH, respectively. The box plots (see online supplementary figures S2 and S3) show a similar pattern in each hospital. Compared with patients discharged alive, the deceased patients were aged older, with lower albumin, haemoglobin and sodium values and higher creatinine, potassium, white cell count and urea values. NEWS was higher in deceased patients compared with patients discharged alive, as respiratory rate and pulse rate were higher in deceased patients. However, the temperature, blood pressure and oxygen saturation were lower in deceased patients. The line plots in online supplementary figures S4 and S5 show that the relationship between a given continuous covariate and the risk of death is similar in each hospital.

Statistical modelling of CARM

We assessed the performance of the CARM model to predict the risk of in-hospital mortality. The model coefficients in logit scale with examples are shown in online supplementary table S3. Table 3 shows the performance of the model in the development and validation dataset. Figure 1 shows the ROC plots of CARM in the development and validation datasets (see online supplementary figure S6 for ROC plots comparing CARM vs NEWS). The C-statistic was high in the development dataset 0.87 (95% CI 0.86 to 0.88) and the external validation dataset 0.86 (95% CI 0.85 to 0.87). Likewise, the scaled Brier score and discrimination were similar in the development and external validation datasets. The calibration slope is 0.97 (95% CI 0.94 to 1.00), which is good (see online supplementary figure S7). The final CARM model, which is not intended for paper-based use, is shown in the online supplementary figure S7).

We excluded 10.0% (NH) and 24.2% (YH) of emergency admissions from the development and validation dataset, respectively, because they had no or incomplete set of blood test results reported. We examined the performance of the CARM model in these excluded records by first imputing age and sex-specific median blood test results, and then applying the CARM model to these admissions only. The last column in table 3 shows the subsequent C-statistics in these imputed records only. The C-statistics for these imputed records were not markedly different in the development and validation dataset

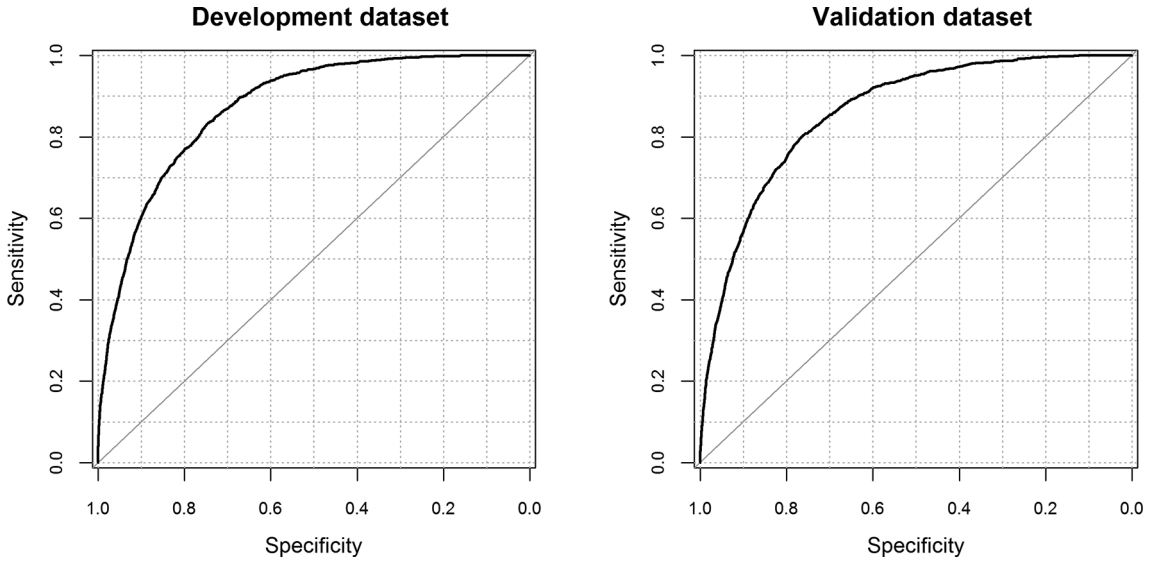


Figure 1 Area under the receiver operating characteristic curve for development dataset (0.87) and validation dataset (0.86).

BMJ Open: first published as 10.1136/bmjopen-2018-022939 on 6 December 2018. Downloaded from http://bmjopen.bmj.com/ on 15 May 2019 by guest. Protected by copyright.

(see online supplementary figure S8 for corresponding ROC plots).

Table 4 shows the sensitivity, specificity and positive and negative predictive values along with likelihood ratio (LR+/LR−) for a selected range of cut-off values for the risk of dying, which tentatively suggests that a threshold risk of 8% provides a reasonable balance between sensitivity (around 70%) and specificity (>80% in development and validation datasets—see table 4 and online supplementary figure S9). Furthermore, the CARM model performance is good in each hospital in various subgroups such as by sex, age, seasons, longer versus shorter length of stay admissions, day of the week and 16 Charlson Comorbidity Index (CCI) disease groups (see online supplementary table S4).

DISCUSSION

We have shown that it is feasible to use the first electronically recorded vital signs and blood test results of an emergency medical patient to predict the risk of in-hospital mortality following emergency medical admission. We developed our CARM model in one hospital and externally validated in data from another hospital. We found that CARM has good performance and our findings tentatively suggest that a cut-off of 8% predicted risk of in-hospital mortality death appears to strike a reasonable balance between sensitivity and specificity.

While several previous studies[1] have used blood test results[20–27] or patient physiology[28 29] to predict the risk of in-hospital mortality, few studies have combined these two data sources[2 30–32] and even fewer reported external validation.[1] Our study is based on data from two different hospitals with material differences in recording of blood test results but still yielding similar performance of CARM. This suggests that our approach, which merits further study, may be generalisable to other UK NHS hospitals with electronically recorded blood test results and NEWS, especially as the use of NEWS in the UK NHS is mandated and that our approach does not rely on reference ranges from blood tests which can vary between hospitals. Indeed, a recent paper with sepsis as the outcome variable also showed promising results by combining the first blood test results and NEWS.[33]

There are a number of limitations in our study. There appears to be a systematic difference in the prevalence of oxygen supplementation in the development and validation datasets, which may warrant further investigation. However, the prevalence ratios (dead/alive) are similar in both groups (2.77 and 3.29 for NH and YH, respectively) and therefore this should have no significant detrimental effect on the validity of our model. Although we focused on in-hospital mortality (because we aimed to aid clinical decision making in the hospital), the impact of this selection bias needs to be assessed by capturing out-of-hospital mortality by linking death certification data and hospital data. CARM, like other risk scores, can only be an aid to the decision-making process of clinical teams[1 17] and

Table 4 Sensitivity, specificity and predictive values for the CARM model at various cut-offs in the development dataset and validation dataset

Dataset	Risk value cut-off	No. of patients >cut-off	% Sensitivity (95% CI)	% Specificity (95% CI)	% PPV (95% CI)	% NPV (95% CI)	LR+ (95% CI)	LR− (95% CI)
Development dataset	0.01	19876	98.5 (97.9 to 99)	38 (37.4 to 38.5)	8.8 (8.4 to 9.2)	99.8 (99.7 to 99.8)	1.6 (1.6 to 1.6)	0 (0 to 0.1)
	0.02	15297	95.7 (94.6 to 96.6)	53.4 (52.9 to 54)	11 (10.6 to 11.6)	99.5 (99.4 to 99.6)	2.1 (2 to 2.1)	0.1 (0.1 to 0.1)
	0.04	10382	87.3 (85.6 to 88.8)	69.8 (69.2 to 70.3)	14.8 (14.2 to 15.5)	98.9 (98.8 to 99)	2.9 (2.8 to 3)	0.2 (0.2 to 0.2)
	0.08	6070	72.2 (70 to 74.3)	83.6 (83.2 to 84)	21 (20 to 22.1)	98 (97.8 to 98.2)	4.4 (4.2 to 4.6)	0.3 (0.3 to 0.4)
	0.20	2190	42 (39.6 to 44.3)	95 (94.8 to 95.3)	33.8 (31.9 to 35.9)	96.4 (96.2 to 96.7)	8.5 (7.9 to 9.1)	0.6 (0.6 to 0.6)
Validation dataset	0.01	18338	98.4 (97.7 to 99)	32.1 (31.5 to 32.7)	9.1 (8.7 to 9.6)	99.7 (99.5 to 99.8)	1.4 (1.4 to 1.5)	0 (0 to 0.1)
	0.02	14537	95.9 (94.9 to 96.8)	47.4 (46.8 to 48.1)	11.2 (10.7 to 11.8)	99.4 (99.3 to 99.5)	1.8 (1.8 to 1.9)	0.1 (0.1 to 0.1)
	0.04	10047	89 (87.4 to 90.4)	65.2 (64.6 to 65.8)	15.1 (14.4 to 15.8)	98.8 (98.7 to 99)	2.6 (2.5 to 2.6)	0.2 (0.1 to 0.2)
	0.08	5871	73.2 (71 to 75.3)	81.2 (80.7 to 81.6)	21.2 (20.2 to 22.3)	97.8 (97.5 to 98)	3.9 (3.7 to 4)	0.3 (0.3 to 0.4)
	0.20	2158	43.1 (40.7 to 45.5)	94.2 (93.9 to 94.5)	34 (32 to 36.1)	96 (95.7 to 96.2)	7.4 (6.9 to 8)	0.6 (0.6 to 0.6)

CARM, computer-aided risk of mortality; LR+, positive likelihood ratio; LR−, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

BMJ Open: first published as 10.1136/bmjopen-2018-022939 on 6 December 2018. Downloaded from http://bmjopen.bmj.com/ on 15 May 2019 by guest. Protected by copyright.

its usefulness in clinical practice remains to be seen. We found that up to about quarter of emergency medical admissions had no (or an incomplete set of) recorded blood test results for whom we tested a simple median imputation strategy without knowing why such data were missing. We found that the performance of CARM did not materially deteriorate in these admissions. We do not suggest that our imputation method is an optimal imputation strategy. Rather, we offer it as a simple, pragmatic, preliminary imputation strategy, which is akin to the AKI detection algorithm which also imputes the median creatinine value where required.[34] Further work on how to optimally address the issue of missing data is required. We did not undertake an imputation exercise for patients with no recorded NEWS because they constituted a much smaller proportion of missing data (<5%), and NEWS is not recommended in patients requiring immediate resuscitation, direct admission to intensive care, and patients with end-stage renal failure or with acute intracranial conditions.[35] We have used the first set of electronically recorded vital signs and blood test results to develop CARM, but updating CARM scores in real-time when new data become available is likely to be important to clinical teams and so warrants further study. Finally, our external validation was undertaken by the same research team in a similar context of the NHS. Further external validation by different research teams in different settings would be useful.

We have designed CARM to be used in hospitals with sufficient informatics infrastructure (eg, electronic health records).[36 37] CARM is not targeting specific emergency medical patients only. Rather, we are seeking to raise situational awareness of the risk of death in-hospital as early as possible, without requiring any additional data items or prompts from clinicians. While we have demonstrated that CARM has potential, we have yet to test its use in routine clinical practice. This is important because we need to demonstrate that CARM does more 'good' than 'harm' in practice.[36 37] For example, while routine blood tests are not indicated in a considerable number of emergency medical admissions, it is nevertheless possible that for a given patient, some clinicians (eg, less experienced) may be tempted to order routine blood tests so that they can obtain a CARM score to support their clinical decision-making process. So, the next phase of this work is to field test CARM by carefully engineering it into routine clinical practice to see if it does enhance the quality of care for acutely ill patients, while noting any unintended consequences.

CONCLUSION

We have developed a novel, externally validated CARM model, with good performance for estimating the risk of in-hospital mortality following emergency medical admission using the patient's first, electronically recorded, vital signs and blood test results. Since CARM places no additional data collection burden on clinicians and is readily automated, it may now be carefully introduced and evaluated in hospitals with electronic health records.

Author affiliations
[1]Faculty of Health Studies, University of Bradford, Bradford, UK
[2]Bradford Institute for Health Research, Bradford, UK
[3]School of Clinical Therapies, University College Cork, Cork, Ireland
[4]Department of Renal Medicine, York Teaching Hospital NHS Foundation Trust Hospital, York, UK
[5]York Teaching Hospital NHS Foundation Trust Hospital, York, UK
[6]Department of Strategy and Planning, Northern Lincolnshire and Goole NHS Foundation Trust, Scunthorpe, UK
[7]Northern Lincolnshire and Goole NHS Foundation Trust, Scunthorpe, UK
[8]School of Health and Social Work, University Of Hull, Hull, UK

Contributors MAM and DR had the original idea for this work. NJ was overall study coordinator with JeD as local NLAG coordinator. MF, AJS and MAM undertook the statistical analyses. JuD, CM and NJ are leads for qualitative studies. RH and KB extracted the necessary data frames. DR, MM and KS gave a clinical perspective. MAM and MF 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 healthcare 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).

Disclaimer 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 None declared.

Patient consent Not required.

Ethics approval This study received ethical approval from The Yorkshire & Humberside Leeds West Research Ethics Committee on 17 September 2015 (ref. 173753), with NHS management permissions received January 2016.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Our data sharing agreement with the two hospitals (York hospital & NLAG hospital) does not permit us to share this data with other parties. Nonetheless if anyone is interested in the data, then they should contact the R&D offices at each hospital in the first instance.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

1. Brabrand M, Folkestad L, Clausen NG, *et al*. Risk scoring systems for adults admitted to the emergency department: a systematic review. *Scand J Trauma Resusc Emerg Med* 2010;18:8.
2. Silke B, Kellett J, Rooney T, *et al*. An improved medical admissions risk system using multivariable fractional polynomial logistic regression modelling. *QJM* 2010;103:23–32.
3. Royal College of Physicians. National Early Warning Score (NEWS): Standardising the assessment of acuteillness severity in the NHS. https://www.rcplondon.ac.uk/news/nhs-england-approves-use-national-early-warning-score-news-2-improve-detection-acutely-il
4. McGinley A, Pearse RM. A national early warning score for acutely ill patients. *BMJ* 2012;345:e5310.
5. 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.
6. Edwards M, McKay H, Van Leuvan C, *et al*. Modified early warning scores: inaccurate summation or inaccurate assignment of score? *Critical Care* 2010;14:P257.
7. 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.

BMJ Open: first published as 10.1136/bmjopen-2018-022939 on 6 December 2018. Downloaded from http://bmjopen.bmj.com/ on 15 May 2019 by guest. Protected by copyright.

8. 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.
9. NHS England. Acute Kidney Injury (AKI) Programme. 2014. http://www.england.nhs.uk/ourwork/patientsafety/akiprogramme/
10. NCEPOD. Acute kidney injury: adding insult to injury. National confidential enquiry into patient outcome and death. 2009.
11. Venables W, Ripley B. *Modern Applied Statistics with S*. 4 edn. New York: Springer, 2002.
12. R Development Core Team. R: A language and environment for statistical computing. R foundation for statistical computing. 2015. http://www.r-project.org/
13. Steyerberg EW. Clinical Prediction Models. In: *A practical approach to development, validation and updating*: Springer, 2008.
14. Harrell FE. rms: Regression Modeling Strategies. 2015. http://cran.r-project.org/package=rms
15. Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;45:562–5.
16. 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.
17. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
18. 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.
19. Moons KG, Altman DG, Reitsma JB, *et al*. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1.
20. Prytherch DR, Sirl JS, Weaver PC, *et al*. Towards a national clinical minimum data set for general surgery. *Br J Surg* 2003;90:1300–5.
21. Prytherch DR, Sirl JS, Schmidt P, *et al*. The use of routine laboratory data to predict in-hospital death in medical admissions. *Resuscitation* 2005;66:203–7.
22. Pine M, Jones B, Lou YB. Laboratory values improve predictions of hospital mortality. *Int J Qual Health Care* 1998;10:491–501.
23. Vroonhof K, van Solinge WW, Rovers MM, *et al*. Differences in mortality on the basis of laboratory parameters in an unselected population at the Emergency Department. *Clin Chem Lab Med* 2005;43:536–41.
24. Hucker TR, Mitchell GP, Blake LD, *et al*. Identifying the sick: can biochemical measurements be used to aid decision making on presentation to the accident and emergency department. *Br J Anaesth* 2005;94:735–41.
25. Froom P, Shimoni Z. Prediction of hospital mortality rates by admission laboratory tests. *Clin Chem* 2006;52:325–8.
26. Asadollahi K, Hastings IM, Gill GV, *et al*. Prediction of hospital mortality from admission laboratory data and patient age: a simple model. *Emerg Med Australas* 2011;23:354–63.
27. ten Boekel E, Vroonhof K, Huisman A, *et al*. Clinical laboratory findings associated with in-hospital mortality. *Clin Chim Acta* 2006;372:1–13.
28. Duckitt RW, Buxton-Thomas R, Walker J, *et al*. Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study. *Br J Anaesth* 2007;98:769–74.
29. Hodgetts TJ, Kenward G, Vlachonikolis IG, *et al*. The identification of risk factors for cardiac arrest and formulation of activation criteria to alert a medical emergency team. *Resuscitation* 2002;54:125–31.
30. Prytherch DR, Briggs JS, Weaver PC, *et al*. Measuring clinical performance using routinely collected clinical data. *Med Inform Internet Med* 2005;30:151–6.
31. O'Sullivan E, Callely E, O'Riordan D, *et al*. Predicting outcomes in emergency medical admissions - role of laboratory data and co-morbidity. *Acute Med* 2012;11:59–65.
32. Mohammed MA, Rudge G, Watson D, *et al*. Index blood tests and national early warning scores within 24 hours of emergency admission can predict the risk of in-hospital mortality: a model development and validation study. *PLoS One* 2013;8.
33. 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.
34. NHS England. Algorithm for detecting Acute Kidney Injury (AKI) based on serum creatinine changes with time. 2014. https://www.england.nhs.uk/wp-content/uploads/2014/06/psa-aki-alg.pdf
35. Silcock DJ, Corfield AR, Gowens PA, *et al*. Validation of the National early warning score in the prehospital setting. *Resuscitation* 2015;89:31–5.
36. Escobar GJ, Dellinger RP. Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med* 2016;11:S5–10.
37. 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;1124:S18–24.

BMJ Open: first published as 10.1136/bmjopen-2018-022939 on 6 December 2018. Downloaded from http://bmjopen.bmj.com/ on 15 May 2019 by guest. Protected by copyright.