BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-022939
Article Type:	Research
Date Submitted by the Author:	15-Mar-2018
Complete List of Authors:	Faisal, Muhammad; University of Bradford; Bradford Institute for Health Research Scally, Andrew ; University of Bradford, Faculty of Health Studies; Bradford Institute for Health Research Jackson, Natalie; Bradford Institute for Health Research Richardson, Donald; York Teaching Hospital NHS Foundation Trust Beatson, Kevin; York Teaching Hospital NHS Foundation Trust Howes, Robin; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Speed, Kevin; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Menon, Madhav; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Daws, Jeremey; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Dyson, Judith; University of Hull, Department of Community Health and Social Care Marsh, Claire; Bradford Institute for Health Research Mohammed, Mohammed ; University of Bradford Faculty of Health Studies; Bradford Institute for Health Research
Keywords:	computer aided risk score, hospital mortality, vital signs and blood test, national early warning score, emergency admission

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

Authors

Muhammad Faisal
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

Andy Scally
Medical Statistician
Faculty of Health Studies, University of Bradford, Bradford, UK
Bradford Institute for Health Research
E-mail: A.J.Scally@bradford.ac.uk

Natalie Jackson
Health Informatics Senior Project Manager
Improvement Academy, Bradford Institute for Health Research
E-mail: Natalie.jackson@yhahsn.nhs.uk

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

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

Robin Howes
Operational Manager for Electronic Patient Records
Department of Strategy & Planning
Northern Lincolnshire and Goole Hospitals
E-mail: robin.howes@nhs.net

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

Madhav Menon
Consultant in respiratory medicine
Northern Lincolnshire and Goole Hospitals

E-mail: madhav.menon1@nhs.net

Jeremy Daws
Head of Quality Assurance
Northern Lincolnshire and Goole Hospitals
E-mail: Jeremy.daws@nhs.net

Judith Dyson
Senior Lecturer Mental Health
University Of Hull
E-mail: J.Dyson@hull.ac.uk

Claire Marsh
Senior Research Fellow/ Patient & Public Engagement Lead
Improvement Academy, Bradford Institute for Health Research
E-mail: claire.marsh@bthft.nhs.uk

Mohammed A Mohammed
Professor of Healthcare Quality & Effectiveness
Faculty of Health Studies, University of Bradford, Bradford, UK
Deputy Director of the Bradford Institute for Health Research
Academic Director to the Yorkshire & Humberside Academic Health Sciences Network
E-mail: M.A.Mohammed5@Bradford.ac.uk

Corresponding author: Muhammad Faisal M.Faisal1@bradford.ac.uk

Abstract

Objective: 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.

Materials: We extracted details of all adult emergency medical admissions from two acute hospitals (NH – model development data; YH – external validation data) discharged over a 24-month period with vital signs and blood test results. We report the performance of the CARM score in terms of the c-statistic.

Results: The risk of in-hospital mortality following emergency medical admission was 5.7% (NH: 1766/30996) and 6.5% (YH: 1703/26247). 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).

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 tests 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.

Key words: computer aided risk score, hospital mortality, vital signs and blood test, national early warning score, emergency admission

Article Summary

- There are no established mortality risk equations specifically for emergency medical patients who are admitted to a general ward, usually via a medical admissions unit.
- 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.
- CARM is not intended for hospitals without sufficient IT infrastructure
- About 20-30% of admissions do not have both NEWS 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 data sets – 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 United Kingdom (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 Royal College of Physicians (London) [3]. 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 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 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 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 National Health Service (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 & 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). 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, YH respectively) 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 (age≥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 tests results recorded within 4 days (96 hours) of admission (>90% of blood test results were within ±24 hours of admission - see table S1 in appendix).

For model development purposes, we were unable to consider emergency admissions without complete blood test results and NEWS recorded – this constituted 16.5% (6104/37100) of records in NH and 28.6% (10504/36751) 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 figure S1 in appendix).

Development of a Computer Aided Risk of Mortality (CARM) Score

We began with exploratory analyses including scatter 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 Acute Kidney Injury (AKI) score. We used the *qladder* function (*Stata* [11]), 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 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 [12] in *R* [13].

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 [14], using data from another hospital (YH) (the external validation dataset). The data from YH is 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 [15] in R to estimate statistical optimism [14,15].

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 [14]. 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 0–100%. Higher values indicate superior models. For calibration, we used the popular Hosmer-Lemeshow (HL) deciles of risk goodness of fit test [16] that compares observed versus predicted number of deaths, although studies have noted that this test is less useful for assessing calibration when models are developed

on large sample sizes because small differences between observed and predicted values reach statistical significance [17,18].

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 [14]. The area under the ROC curve is interpreted as the probability that a deceased patient has a higher risk of death than a randomly chosen non-deceased patient. A c-statistic of 0.5 is no better than tossing a coin, whilst a perfect model has a c-statistic of 1. The higher the c-statistic, the better the model. In general, values less than 0.7 are considered to show poor discrimination, values of 0.7–0.8 can be described as reasonable, and values above 0.8 suggest good discrimination [19]. The 95% confidence interval for the c-statistic was derived using DeLong's method as implemented in the *pROC* library [20] in *R* [13]. 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 [14]. All analyses were carried using *R* [13] and *Stata* [11].

Ethical 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.

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 five years. The patient and public voice continued to be included throughout the project with three patient representatives invited to sit on 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.

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.

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% (6104/37100) in NH and 28.6% (10504/36751) 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 (Table 1). 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. Missing blood tests were seen to occur more frequently with patients discharged alive versus those that died.

The in-hospital mortality was 5.7% (1766/30996) in NH and 6.5% (1703/26247) 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.

Figures S2 to S5 (see appendix) show box plots and scatter plots for each continuous (untransformed) covariate that was included in the CARM model for NH and YH respectively. The box plots (figures S2 & S3 in appendix) 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 were temperature, blood pressure and oxygen saturation values. The respiratory rate and pulse rate were lower in deceased patients. The scatter plots in the appendix (figures S4 & S5 in appendix) 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 (logit scale) and external validation plots are shown in the appendix. Table 3 shows the performance of the model in the development and validation dataset. Figure 1 shows the ROC plots in the development and validation datasets. The c-statistic was high in the development dataset 0.87 [95% CI 0.86 - 0.88] and the external validation dataset 0.86 [95% CI 0.85 - 0.87]. Likewise, the scaled Brier score and discrimination were similar in the development and external validation datasets. The Hosmer-Lemeshow (HL) showed poor calibration (p-value<0.001). However, after adjusting for 'calibration-in-the-large' (see appendix), our calibration slope is 1.0, which is ideal (see appendix Figure S6).

The final CARM model, which is not intended for paper-based use, is shown in the appendix with accompanying internal and external validation plots (see appendix figure S6).

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 (see Figure S7 appendix for corresponding ROC plots).

Table 4 shows the sensitivity, specificity and positive predictive value 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 (more than 80% in development and validation datasets – see table 4 and figure S8 in appendix).

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.

Whilst several previous studies [1] have used blood test results [21–28] or patient physiology [29,30] to predict the risk of in-hospital mortality, few studies have combined these two data sources [31–34] 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 [35].

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 [1,19] and its usefulness in clinical practice remains to be seen. We found that up to about ¼ 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 was 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 [36]. 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 [37]. 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 becomes available is likely to be important to clinical teams and so warrants further study.

We have designed CARM to be used in hospitals with sufficient informatics infrastructure (eg electronic health records) [38,39]. 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. Whilst 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 [38,39]. For example, whilst 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, whilst 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.

Acknowledgements

The authors would like to express their gratitude to Sue Rushbrook and Gary Hardcastle at York Hospital for their support in obtaining the data frame, and to Laura Lamming at the University of Bradford for her support with the Ethics application.

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.

Contributorship

MAM & DR had the original idea for this work. NJ was overall study coordinator with JeD as local NLAG coordinator. MF, AS and MAM undertook the statistical analyses. JuD, CM & 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.

Competing Interests: The authors declare no conflicts of interest.

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

Characteristic	Development dataset (NH)		Validation dataset (YH)	
	N (%)	Died (%)	N (%)	Died (%)
Total emergency medical admissions	37100	2171 (5.9)	36751	2137 (5.8)
Excluded: No NEWS recorded (%)	1305 (3.5)	212 (16.3)	772 (2.1)	47 (6.1)
Excluded: First NEWS after 24 hours of admission (%)	634 (1.7)	59 (9.3)	172 (0.5)	10 (5.8)
Excluded: First blood test results after 4 days of admission (%)	464 (1.3)	31 (6.7)	673 (1.8)	83 (12.3)
Excluded: No or incomplete blood test results recorded (%)	3701 (10.0)	103 (2.8)	8887 (24.2)	294 (3.3)
Total excluded (%)	6104 (16.5)	405 (6.6)	10504 (28.6)	434 (4.1)
Total included (%)	30996 (83.5)	1766 (5.7)	26247 (71.4)	1703 (6.5)

Table 2 Characteristics of emergency admissions for development and validation datasets.

Characteristic	Development dataset (NH)		Validation dataset (YH)	
	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 [umol/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 minute] (SD)	18 (3.5)	20.1 (4.8)	18.6 (4.6)	21.7 (6.8)
Mean Temperature [^{o}C] (SD)	36.5 (0.7)	36.3 (0.8)	36.3 (0.8)	36.1 (1.1)
Mean Systolic pressure [mmHg] (SD)	129.6 (22.7)	119.8 (24.8)	136.1 (27.2)	128.5 (30.3)
Mean Diastolic pressure [mmHg] (SD)	75 (14.8)	69.5 (15.8)	75.4 (15.5)	71.3 (17.7)
Mean Pulse rate [beats per minute] (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)

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

Dataset	Chi-square†	p-value†	Mean predicted risk: Alive	Mean predicted risk: Died‡	Discrimination	Scaled Brier Score	AUC [95% CI]	Median Imputed AUC [95% CI]
Development dataset	24.64	0.002	0.047	0.229	0.183	0.175	0.874# [0.866 to 0.881]	0.915 [0.888 to 0.941]
Validation dataset	15.82	0.045	0.053	0.231	0.178	0.165	0.861 [0.852 to 0.869]	0.900 [0.880 to 0.919]

NB:

† is based on the Hosmer-Lemeshow deciles of risk goodness of fit test with 8 degrees of freedom.
‡ Died in-hospital following emergency admission
corrected optimism (original = 0.874, and corrected=0.873).

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	Sensitivity %	Specificity %	PPV	NPV
Development dataset	0.01	98.53	37.95	8.75	99.77
	0.02	95.70	53.45	11.05	99.52
	0.04	87.26	69.75	14.84	98.91
	0.08	72.20	83.59	21.00	98.03
	0.20	41.96	95.04	33.82	96.44
Validation dataset	0.01	98.41	32.11	9.14	99.66
	0.02	95.95	47.43	11.24	99.41
	0.04	88.96	65.24	15.08	98.84
	0.08	73.17	81.16	21.22	97.76
	0.20	43.10	94.19	34.00	95.98

PPV = Positive Predictive Value

NPV = Negative Predictive Value

Figure 1 Area under the Receiver Operating Characteristic curve for development dataset (0.87) and validation dataset (0.86).

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. doi:10.1186/1757-7241-18-8

2 Silke B, Kellett J, Rooney T, *et al.* An improved medical admissions risk system using multivariable fractional polynomial logistic regression modelling. *QJM An Int J Med* 2010;**103**:23–32. doi:10.1093/qjmed/hcp149

3 Physicians RC of. National. Early Warning Score (NEWS): Standardising the assessment of acuteillness severity in the NHS https://www.rcplondon.ac.uk/sites/default/files/documents/national-early-warning-score-standardising-assessment-acute-illness-severity-nhs.pdf. 2012.

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. doi:10.1016/j.resuscitation.2006.03.008

6 Edwards M, McKay H, Van Leuvan C, *et al.* Modified Early Warning Scores: inaccurate summation or inaccurate assignment of score? *Crit Care* 2010;**14**:P257. doi:10.1186/cc8489

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. doi:10.1016/j.resuscitation.2005.12.002

8 Mohammed M, Hayton R, Clements G, *et al.* Improving accuracy and efficiency of early warning scores in acute care. *Br J Nurs*;**18**:18–24. doi:10.12968/bjon.2009.18.1.32072

9 NHS England: Acute Kidney Injury (AKI) Programme. http://www.thinkkidneys.nhs.uk, 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 StatCorp. Stata: Release 13. Statistical Software. College Station, TX: StataCorp LP; 2014.

12 Venables W, Ripley B. *Modern Applied Statistics with S*. Fourth Edi. New York: : Springer 2002.

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

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

15 Harrell FE. rms: Regression Modeling Strategies http://cran.r-project.org/package=rms. 2015.

16 Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;**115**:92–106.

17 Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited*. *Crit Care Med* 2007;**35**:2052–6.

doi:10.1097/01.CCM.0000275267.64078.B0

18 Paul P, Pennell ML, Lemeshow S. Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. *Stat Med* 2013;**32**:67–80. doi:10.1002/sim.5525

19 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36. doi:10.1148/radiology.143.1.7063747

20 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. doi:10.1186/1471-2105-12-77

21 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. doi:10.1002/bjs.4274

22 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. doi:10.1016/j.resuscitation.2005.02.011

23 Pine M, Jones B, Lou YB. Laboratory values improve predictions of hospital mortality. *Int J Qual Heal care J Int Soc Qual Heal Care* 1998;**10**:491–501.

24 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. doi:10.1515/CCLM.2005.093

25 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. doi:10.1093/bja/aei122

26 Froom P, Shimoni Z. Prediction of Hospital Mortality Rates by Admission Laboratory Tests. *Clin Chem* 2005;**52**:325–8. doi:10.1373/clinchem.2005.059030

27 Asadollahi K, Hastings IM, Gill G V, *et al.* Prediction of hospital mortality from admission laboratory data and patient age: A simple model. *Emerg Med Australas* 2011;**23**:354–63. doi:10.1111/j.1742-6723.2011.01410.x

28 ten Boekel E, Vroonhof K, Huisman A, *et al.* Clinical laboratory findings associated with in-hospital mortality. *Clin Chim Acta* 2006;**372**:1–13. doi:10.1016/J.CCA.2006.03.024

29 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†‡. *BJA Br J Anaesth* 2007;**98**:769–74. doi:10.1093/bja/aem097

30 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.

31 Prytherch DR, Briggs JS, Weaver PC, *et al.* Measuring clinical performance using routinely collected clinical data. *Med Inform Internet Med* 2005;**30**:151–6. doi:10.1080/14639230500298966

32 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. doi:10.1093/qjmed/hcp149

33 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.

34 Mohammed M a., 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**. doi:10.1371/journal.pone.0064340

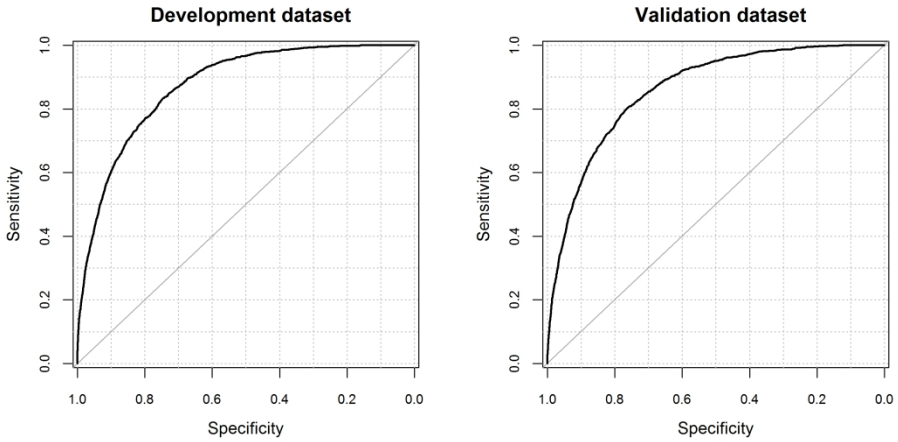
35 Faisal M, Scally A, Richardson D, *et al.* Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test Results. *Crit Care Med* 2018;:1. doi:10.1097/CCM.0000000000002967

36 Algorithm for detecting Acute Kidney Injury (AKI) based on serum creatinine changes with time https://www.england.nhs.uk/wp-content/uploads/2014/06/psa-aki-alg.pdf. 2014.

37 Silcock DJ, Corfield AR, Gowens PA, *et al.* Validation of the National Early Warning Score in the prehospital setting. *Resuscitation* 2015;**89**:31–5. doi:10.1016/j.resuscitation.2014.12.029

38 Escobar GJ, Dellinger RP. Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med* 2016;**11**:S5–10. doi:10.1002/jhm.2653

39 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. doi:10.1002/jhm.2652



Area under the Receiver Operating Characteristic curve for development dataset (0.87) and validation dataset (0.86).

254x127mm (300 x 300 DPI)

Supplementary Material

$$Logit(Died) = -3.220 + 0.140 * \boldsymbol{Male} + 0.077 * \boldsymbol{Age} - 0.104 * \boldsymbol{Albumin} + 9.883 * (\boldsymbol{Creatinine})^{-0.5} + 0.002 * \boldsymbol{Haemoglobin} - 0.024 * \boldsymbol{log}(\boldsymbol{Potassium}) - 0.023 * \boldsymbol{Sodium} + 1.167 * \boldsymbol{log}(\boldsymbol{White\ Cell\ Count}) + 1.211 * \boldsymbol{log}(\boldsymbol{Urea}) + 0.131 * \boldsymbol{AKI\ stage\ 1} + 0.443 * \boldsymbol{AKI\ stage\ 2} - 0.388 * \boldsymbol{AKI\ stage\ 3} + 0.093 * \boldsymbol{NEWS} + 0.569 * \boldsymbol{log}(\boldsymbol{Respiration\ Rate}) - 0.145 * \boldsymbol{Temperature} - 0.919 * \boldsymbol{log}(\boldsymbol{Systoloic\ pressure}) + 0.777 * \boldsymbol{log}(\boldsymbol{Diastolic\ pressure}) + 0.511 * \boldsymbol{log}(\boldsymbol{Pulse\ rate}) - 0.016 * \boldsymbol{Oxygen\ Saturations} + 0.606 * \boldsymbol{Supplemntal\ oxygen} + 0.716 * \boldsymbol{Pain} + 0.395 * \boldsymbol{Voice} + 1.925 * \boldsymbol{Unconscious} - 0.015 * \boldsymbol{Age} * \boldsymbol{log}(\boldsymbol{White\ Cell\ Count}) + 1.481 * (\boldsymbol{Creatinine})^{-0.5} * \boldsymbol{log}(\boldsymbol{White\ Cell\ Count}) + 15.551 * \boldsymbol{AKI\ stage\ 3} * (\boldsymbol{Creatinine})^{-0.5}$$

We accounted for baseline difference in risk of death in the external validation data by adding 0.52 to the CARM logit model using an iterative procedure described elsewhere[1]

1. 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 [Internet]. 2017 Jan 1 [cited 2017 Oct 2];110(1):27–31. Available from: https://academic.oup.com/qjmed/article-lookup/doi/10.1093/qjmed/hcw110

The NEWS [https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news] is based on a scoring system in which a score is allocated to vital signs physiological measurements already undertaken when patients present to, or are being monitored in hospital. Six physiological parameters form the basis of the scoring system:

Physiological Parameters	3	2	1	0	1	2	3
Respiration Rate	≤8		9 - 11	12 - 20		21 - 24	≥25
Oxygen Saturations	≤91	92 - 93	94 - 95	≥96			
Any Supplemental Oxygen		Yes		No			
Temperature	≤35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥39.1	
Systolic BP	≤90	91 - 100	101 - 110	111 - 219			≥220
Heart Rate	≤40		41 - 50	51-90	91 - 110	111 - 130	≥131
Level of Consciousness				Alert			Voice, Pain, or Unconscious

A score is allocated to each as they are measured, the magnitude of the score reflecting how extreme the parameter varies from the norm. This score is then aggregated, and uplifted for people requiring oxygen.

Blood tests results recorded	Development dataset N(%)	Validation dataset N(%)
within 24 hours	29255 (94.4)	24341 (92.7)
within 48 hours	894 (2.9)	1098 (4.2)
within 72 hours	512 (1.7)	495 (1.9)
within 96 hours	335 (1.1)	313 (1.2)

Table S1 Distribution of time to the first set of Blood test results recorded within 4 days for development and validation datasets.

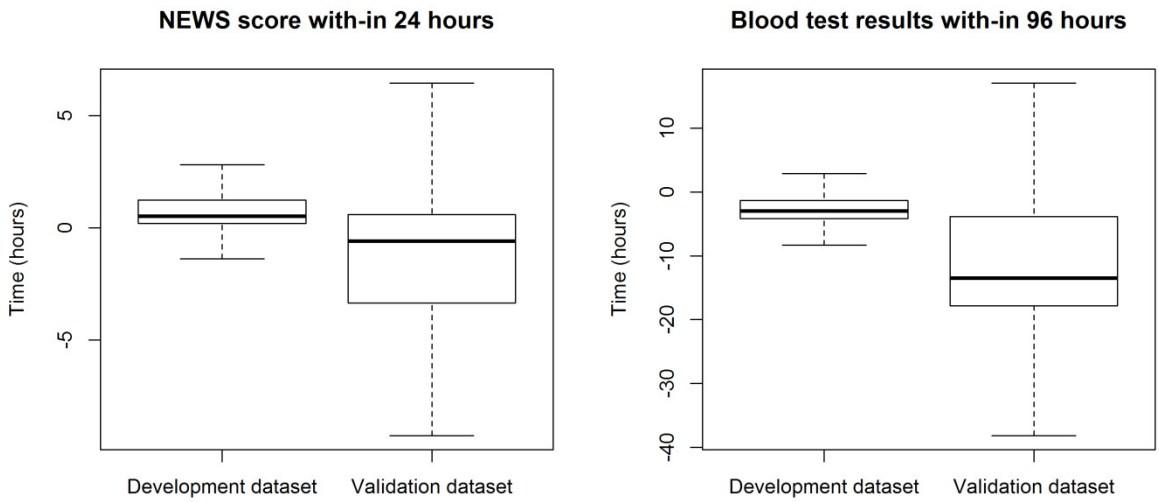


Figure S1 Distribution of time to first NEWS score and Blood test results for development and validation datasets.

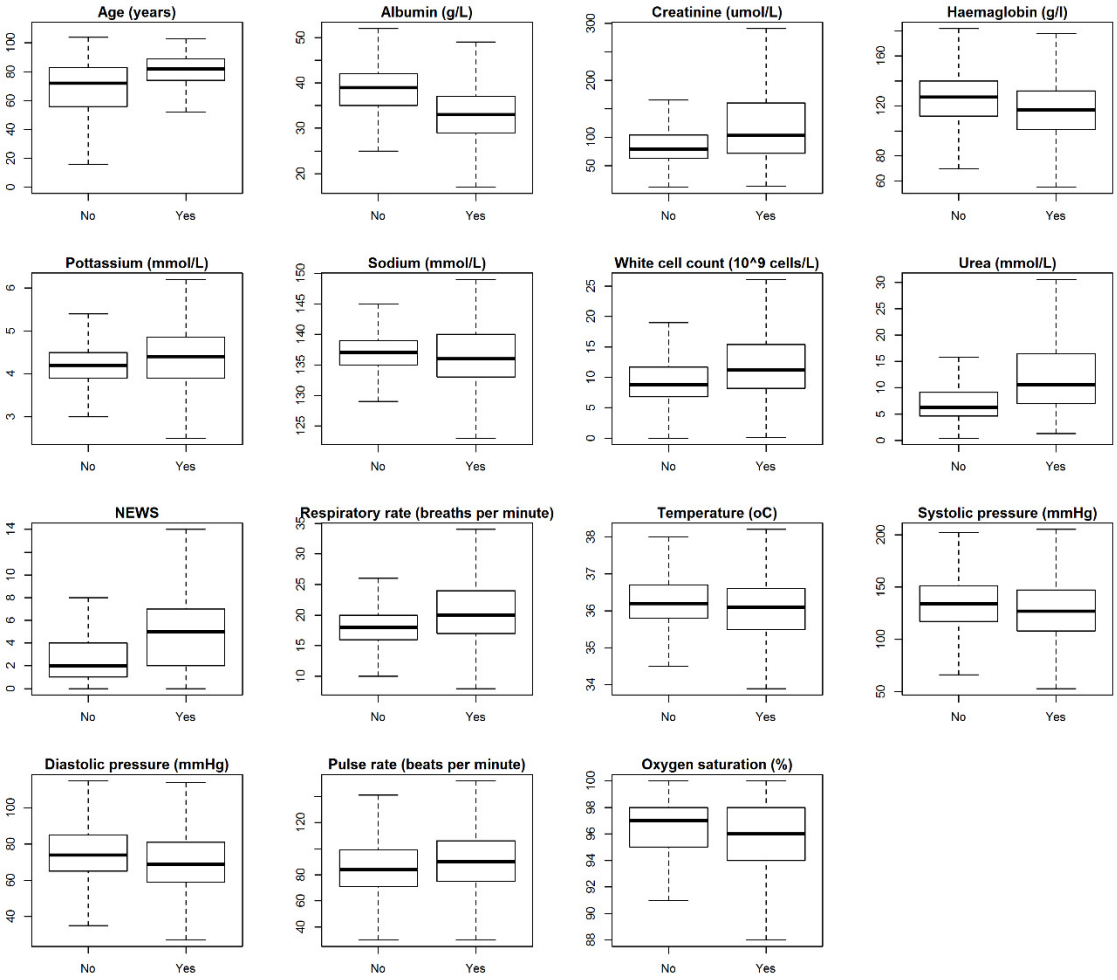


Figure S2 Boxplot without outliers for continuous covariates with respect to patient's discharge status (Alive/Died) for NLAG hospitals

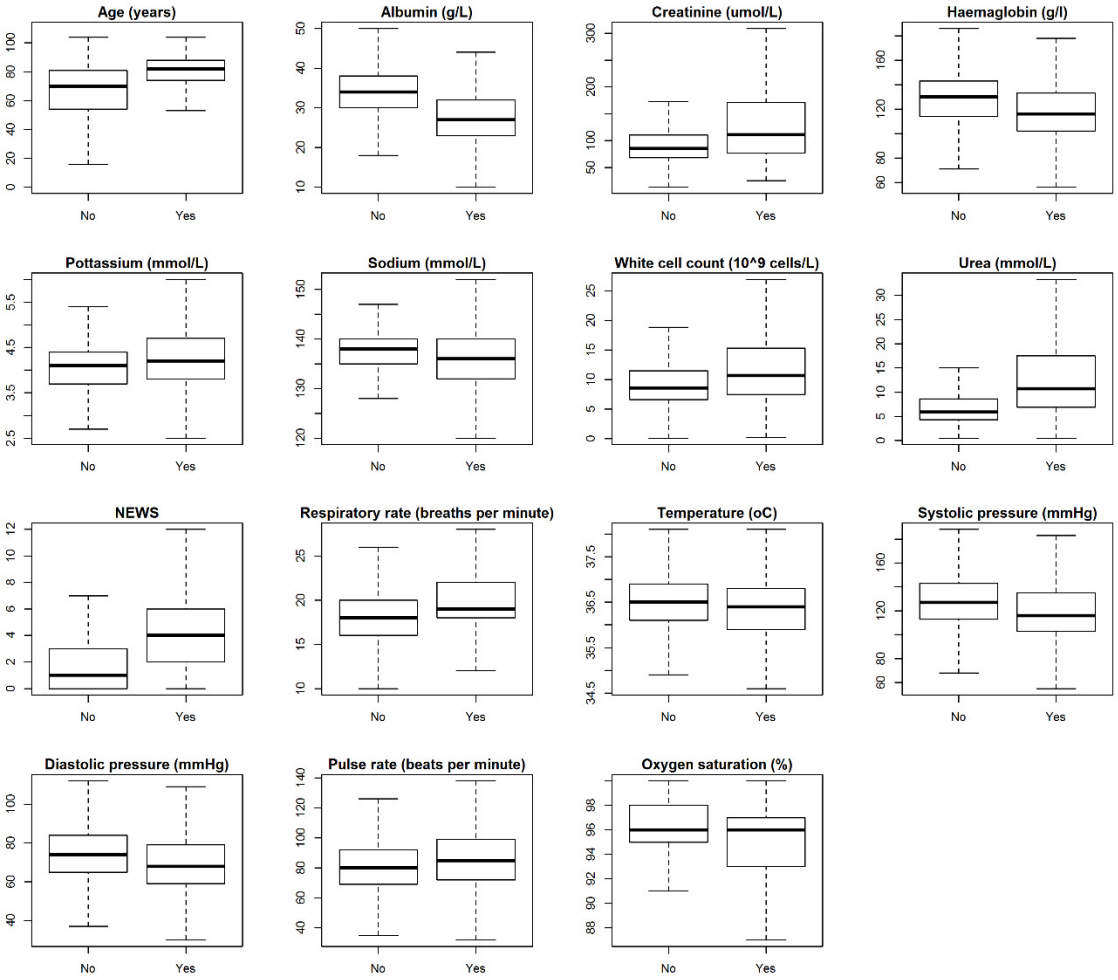


Figure S3 Boxplot without outliers for continuous covariates with respect to patient's discharge status (Alive/Died) for York hospital

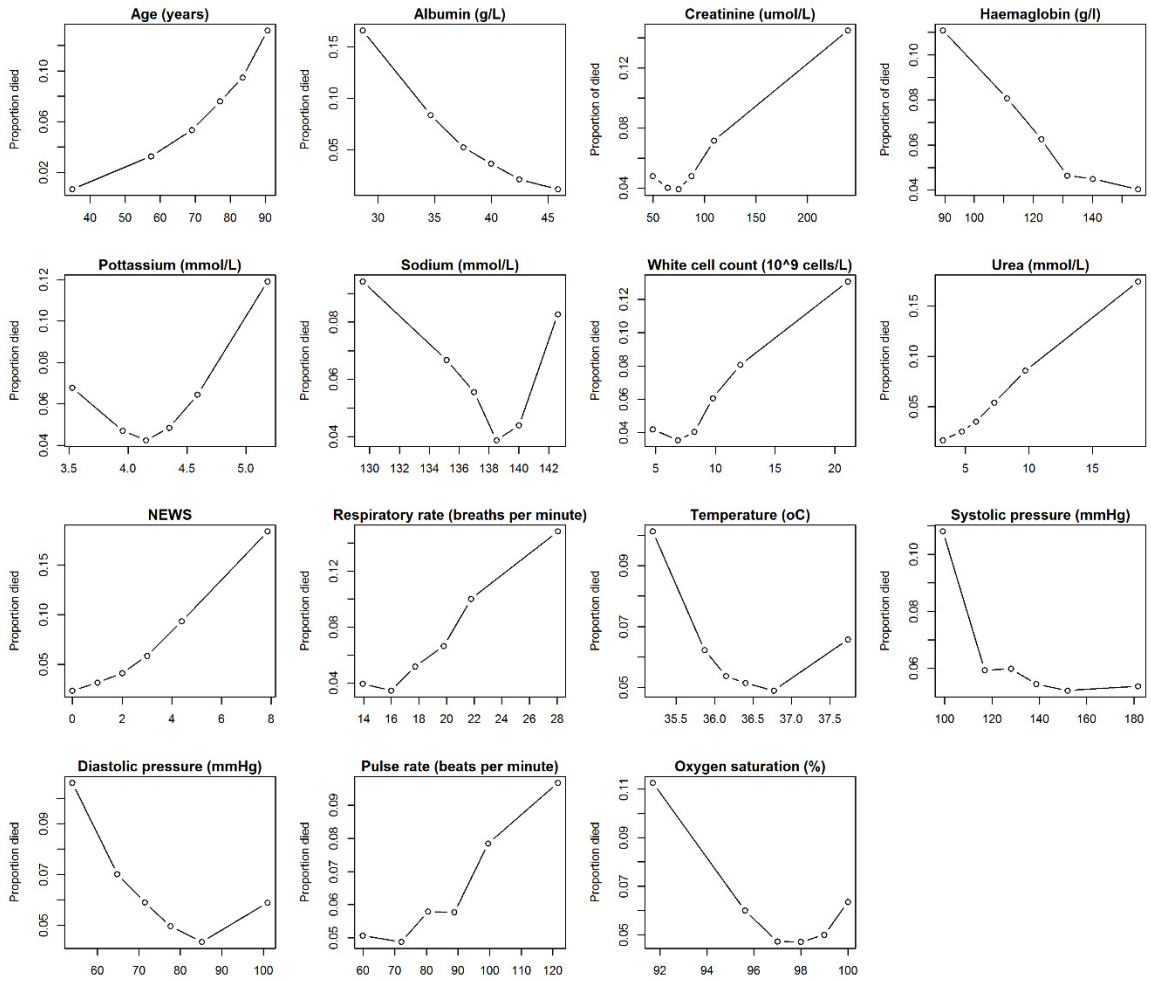


Figure S4 Scatter plots showing the observed risk of death with continuous covariates for NLAG hospitals

NB: y-axis range changes in each plot.

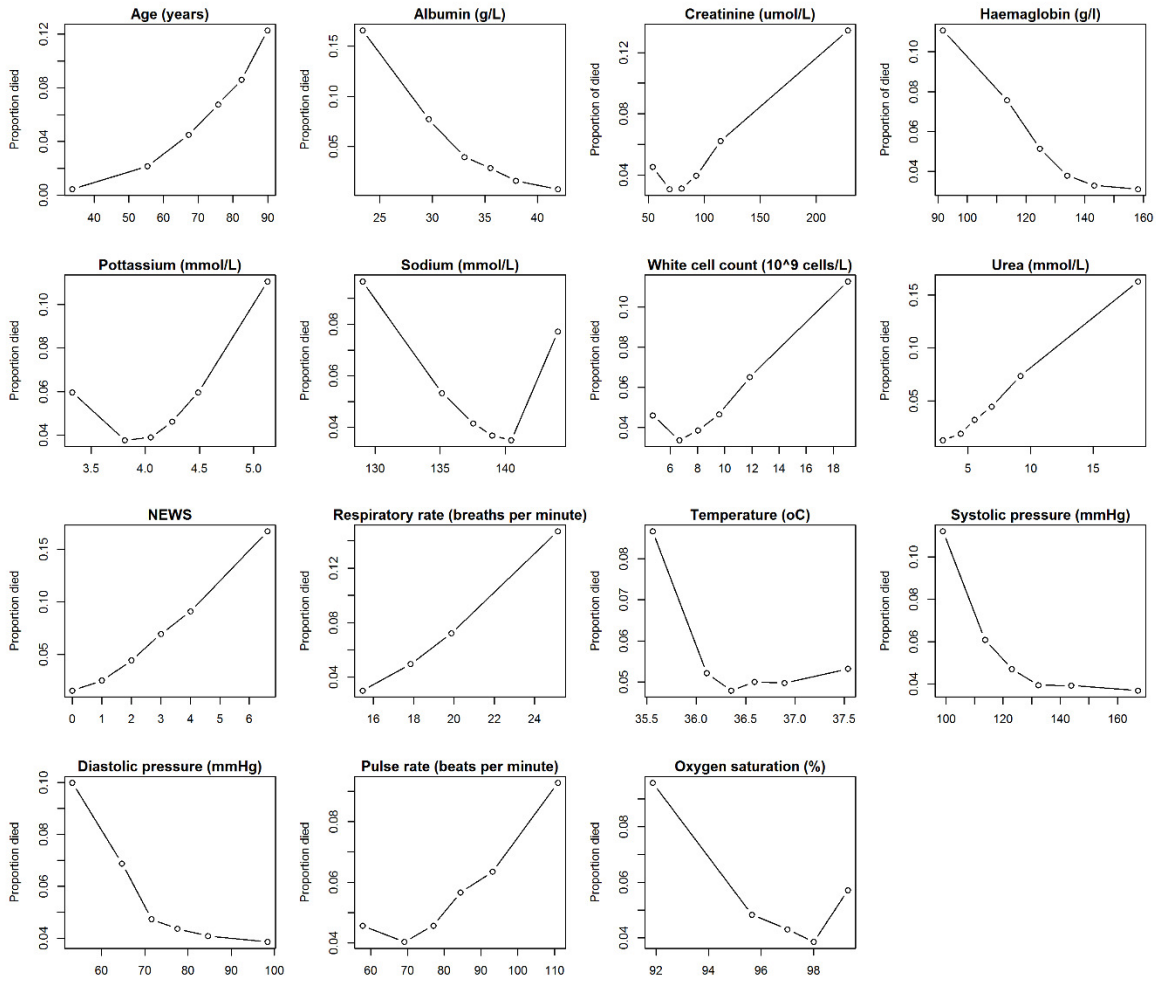


Figure S5 Scatter plots showing the observed risk of death with continuous covariates for York hospital.

NB: y-axis range changes in each plot.

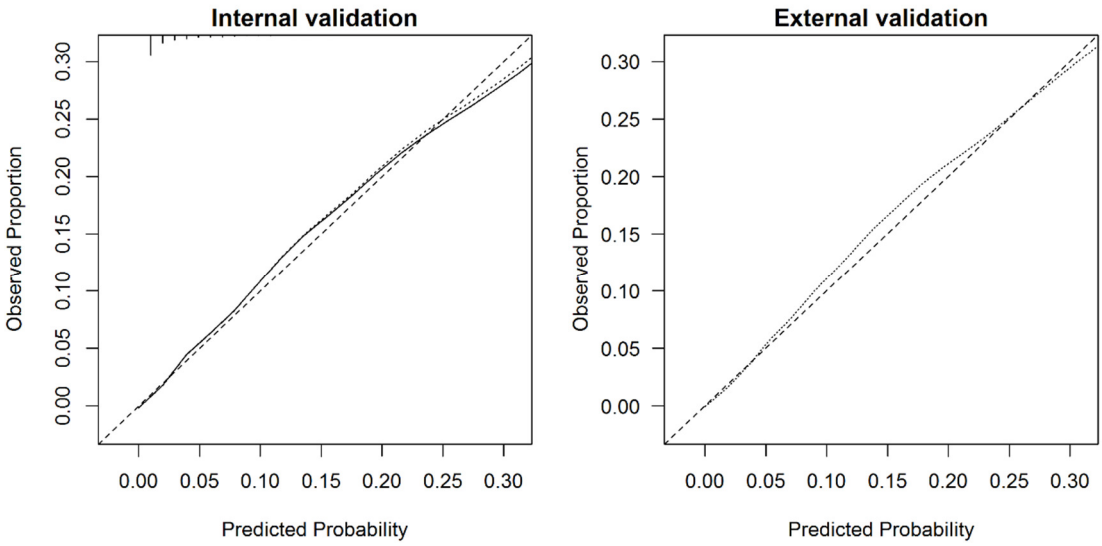


Figure S6 Internal and external validation sepsis model on development dataset and validation dataset.

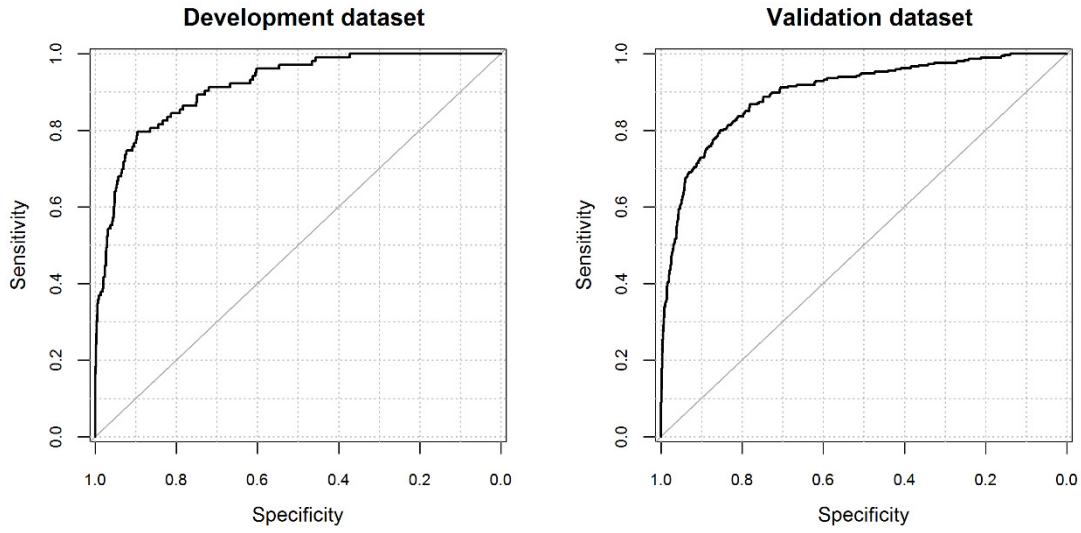


Figure S7 Receiver Operating Characteristic curve of (median) imputed blood tests results on development dataset and validation dataset.

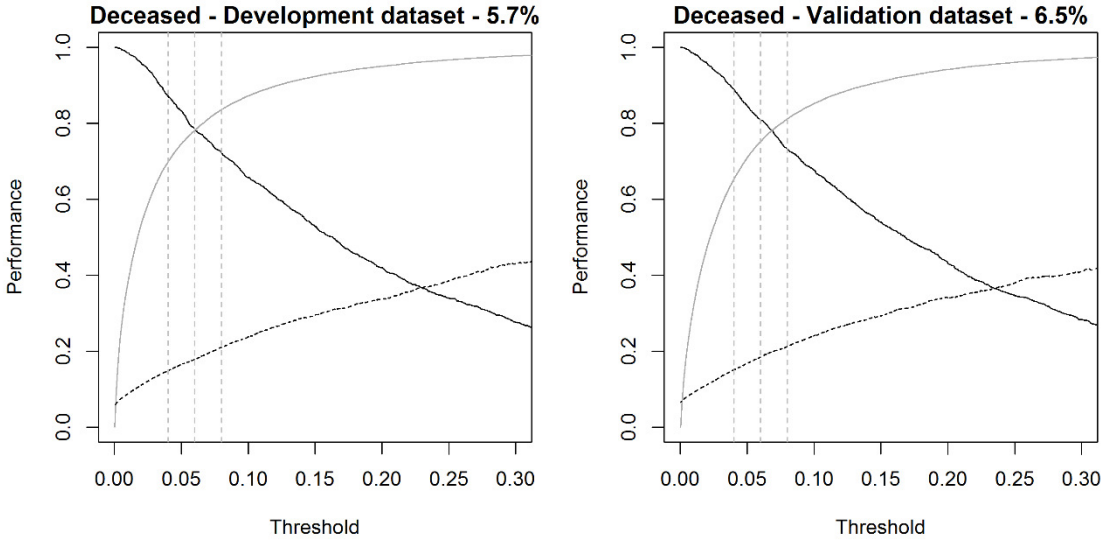


Figure S8: Sensitivity analysis of CARM model at various thresholds (0.0, 0.01, …, 0.30) on development dataset and validation dataset.

Black solid line is for sensitivity and black dashed line is for positive predictive value (PPV). Grey solid line is specificity and grey dashed vertical lines are at thresholds (0.04, 0.06, and 0.08).

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-022939.R1
Article Type:	Research
Date Submitted by the Author:	03-Jul-2018
Complete List of Authors:	Faisal, Muhammad; University of Bradford; Bradford Institute for Health Research Scally, Andrew ; University of Bradford, Faculty of Health Studies Jackson, Natalie; Bradford Institute for Health Research Richardson, Donald; York Teaching Hospital NHS Foundation Trust Beatson, Kevin; York Teaching Hospital NHS Foundation Trust Howes, Robin; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Speed, Kevin; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Menon, Madhav; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Daws, Jeremey; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Dyson, Judith; University of Hull, Department of Community Health and Social Care Marsh, Claire; Bradford Institute for Health Research Mohammed, Mohammed ; University of Bradford Faculty of Health Studies; Bradford Institute for Health Research
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health informatics
Keywords:	computer aided risk score, hospital mortality, vital signs and blood test, national early warning score, emergency admission

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

Muhammad Faisal
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

Andrew J Scally
Medical Statistician
Faculty of Health Studies, University of Bradford, Bradford, UK
Bradford Institute for Health Research
E-mail: A.J.Scally@bradford.ac.uk

Natalie Jackson
Health Informatics Senior Project Manager
Improvement Academy, Bradford Institute for Health Research
E-mail: Natalie.jackson@yhahsn.nhs.uk

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

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

Robin Howes
Operational Manager for Electronic Patient Records
Department of Strategy & Planning
Northern Lincolnshire and Goole Hospitals
E-mail: robin.howes@nhs.net

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

Madhav Menon
Consultant in respiratory medicine
Northern Lincolnshire and Goole Hospitals

E-mail: madhav.menon1@nhs.net

Jeremy Daws
Head of Quality Assurance
Northern Lincolnshire and Goole Hospitals
E-mail: Jeremy.daws@nhs.net

Judith Dyson
Senior Lecturer Mental Health
University Of Hull
E-mail: J.Dyson@hull.ac.uk

Claire Marsh
Senior Research Fellow/ Patient & Public Engagement Lead
Improvement Academy, Bradford Institute for Health Research
E-mail: claire.marsh@bthft.nhs.uk

Mohammed A Mohammed
Professor of Healthcare Quality & Effectiveness
Faculty of Health Studies, University of Bradford, Bradford, UK
Deputy Director of the Bradford Institute for Health Research
Academic Director to the Yorkshire & Humberside Academic Health Sciences Network
E-mail: M.A.Mohammed5@Bradford.ac.uk

Corresponding author: 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 (NH – model development data; YH – external validation data).

Participants: Adult (>=16 years) medical admissions discharged over a 24 month period with electronic NEWS and blood test results recorded on admission.

Results: The risk of in-hospital mortality following emergency medical admission was 5.7% (NH: 1766/30996) and 6.5% (YH: 1703/26247). 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 tests 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.

Key words: computer aided risk score, hospital mortality, vital signs and blood test, national early warning score, emergency admission

Article Summary

- 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 NEWS 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 data sets – 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 United Kingdom (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 Royal College of Physicians (London) [3]. 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 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 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 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 National Health Service (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 & 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). 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, YH respectively) 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 (age≥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 tests results recorded within 4 days (96 hours) of admission (>90% of blood test results were within ±24 hours of admission - see table S1 in appendix).

For model development purposes, we were unable to consider emergency admissions without complete blood test results and NEWS recorded – this constituted 16.5% (6104/37100) of records in NH and 28.6% (10504/36751) 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 figure S1 in appendix).

Development of a Computer Aided Risk of Mortality (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 Acute Kidney Injury (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 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* [11]), 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 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 [12] in *R* [13].

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 [14], using data from another hospital (YH) (the external validation dataset). The data from YH is 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 [15] in R to estimate statistical optimism [14,15].

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 [14]. 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 0–100%. 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' [16]. 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 [17].

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 [14]. 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, whilst a perfect model has a c-statistic of 1. The higher the c-statistic, the better the model. In general, values less than 0.7 are considered to show poor discrimination, values of 0.7–0.8 can be described as reasonable, and values above 0.8 suggest good discrimination [18]. The 95% confidence interval for the c-statistic was derived using DeLong's method as implemented in the *pROC* library [19] in *R* [13]. 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 [14]. We followed the TRIPOD guidelines for model development and validation [20]. All analyses were carried using *R* [13] and *Stata* [11].

Ethical 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.

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 five years. The patient and public voice continued to be included throughout the project with three patient representatives invited to sit on 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.

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.

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% (6104/37100) in NH and 28.6% (10504/36751) 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, 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 Table S2 in appendix for characteristic of emergency admissions with incomplete data).

The in-hospital mortality was 5.7% (1766/30996) in NH and 6.5% (1703/26247) 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.

Figures S2 to S5 (see appendix) 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 (figures S2 & S3 in appendix) 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 the appendix (figures S4 & S5 in appendix) 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 the appendix (see 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 Figure S6 in the appendix for ROC plots comparing CARM versus NEWS). The c-statistic was high in the development dataset 0.87 [95% CI 0.86 - 0.88] and the external validation dataset 0.86 [95% CI 0.85 - 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 appendix Figure S7). The final CARM model, which is not intended for paper-based use, is shown in the appendix with accompanying internal and external validation plots (see appendix 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 (see Figure S8 appendix for corresponding ROC plots).

Table 4 shows the sensitivity, specificity and positive & 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 (more than 80% in development and validation datasets – see table 4 and figure S9 in appendix).

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.

Whilst several previous studies [1] have used blood test results [21–28] or patient physiology [29,30] to predict the risk of in-hospital mortality, few studies have combined these two data sources [31–

34] 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 [35].

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 [1,18] and its usefulness in clinical practice remains to be seen. We found that up to about ¼ 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 was 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 [36]. 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 [37]. 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 becomes 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) [38,39]. 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. Whilst 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 [38,39]. For example, whilst 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, whilst 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.

Acknowledgements

The authors would like to express their gratitude to Sue Rushbrook and Gary Hardcastle at York Hospital for their support in obtaining the data frame, and to Laura Lamming at the University of Bradford for her support with the Ethics application.

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.

Contributorship

MAM & DR had the original idea for this work. NJ was overall study coordinator with JeD as local NLAG coordinator. MF, AS and MAM undertook the statistical analyses. JuD, CM & 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.

Competing Interests: The authors declare no conflicts of interest.

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

Characteristic	Development dataset	Validation dataset
	N (%)	N (%)
Total emergency medical admissions	37100	36751
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)	10504 (28.6)
Total included (%)	30996 (83.5)	26247 (71.4)

Table 2 Characteristics of emergency admissions for development and validation datasets.

Characteristic	Development dataset (NH)		Validation dataset (YH)	
	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 [umol/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 minute] (SD)	18 (3.5)	20.1 (4.8)	18.6 (4.6)	21.7 (6.8)
Mean Temperature [^{o}C] (SD)	36.5 (0.7)	36.3 (0.8)	36.3 (0.8)	36.1 (1.1)
Mean Systolic pressure [mmHg] (SD)	129.6 (22.7)	119.8 (24.8)	136.1 (27.2)	128.5 (30.3)
Mean Diastolic pressure [mmHg] (SD)	75 (14.8)	69.5 (15.8)	75.4 (15.5)	71.3 (17.7)
Mean Pulse rate [beats per minute] (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)

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]

NB:

† mean predicted risk difference between who discharged died and discharged alive.
‡ Died in-hospital following emergency admission
corrected optimism (original = 0.874, and corrected=0.873).

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 > cutoff	%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)

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

Figure 1 Area under the Receiver Operating Characteristic curve for development dataset (0.87) and validation dataset (0.86).

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. doi:10.1186/1757-7241-18-8

2 Silke B, Kellett J, Rooney T, *et al.* An improved medical admissions risk system using multivariable fractional polynomial logistic regression modelling. *QJM An Int J Med* 2010;**103**:23–32. doi:10.1093/qjmed/hcp149

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. doi:10.1016/j.resuscitation.2006.03.008

6 Edwards M, McKay H, Van Leuvan C, *et al.* Modified Early Warning Scores: inaccurate summation or inaccurate assignment of score? *Crit Care* 2010;**14**:P257. doi:10.1186/cc8489

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. doi:10.1016/j.resuscitation.2005.12.002

8 Mohammed M, Hayton R, Clements G, *et al.* Improving accuracy and efficiency of early warning scores in acute care. *Br J Nurs*;**18**:18–24. doi:10.12968/bjon.2009.18.1.32072

9 NHS England: Acute Kidney Injury (AKI) Programme. http://www.thinkkidneys.nhs.uk, 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 StatCorp. Stata: Release 13. Statistical Software. College Station, TX: StataCorp LP; 2014.

12 Venables W, Ripley B. *Modern Applied Statistics with S*. Fourth Edi. New York: : Springer 2002.

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

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

15 Harrell FE. rms: Regression Modeling Strategies http://cran.r-project.org/package=rms. 2015.

16 Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;**45**:562–5. doi:10.1093/biomet/45.3-4.562

17 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. doi:10.1097/EDE.0b013e3181c30fb2

18 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36. doi:10.1148/radiology.143.1.7063747

19 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. doi:10.1186/1471-2105-12-77

20 Moons KGM, 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. doi:10.7326/M14-0698

21 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. doi:10.1002/bjs.4274

22 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. doi:10.1016/j.resuscitation.2005.02.011

23 Pine M, Jones B, Lou YB. Laboratory values improve predictions of hospital mortality. *Int J Qual Heal care J Int Soc Qual Heal Care* 1998;**10**:491–501.

24 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. doi:10.1515/CCLM.2005.093

25 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. doi:10.1093/bja/aei122

26 Froom P, Shimoni Z. Prediction of Hospital Mortality Rates by Admission Laboratory Tests. *Clin Chem* 2005;**52**:325–8. doi:10.1373/clinchem.2005.059030

27 Asadollahi K, Hastings IM, Gill G V, *et al.* Prediction of hospital mortality from admission laboratory data and patient age: A simple model. *Emerg Med Australas* 2011;**23**:354–63. doi:10.1111/j.1742-6723.2011.01410.x

28 ten Boekel E, Vroonhof K, Huisman A, *et al.* Clinical laboratory findings associated with in-hospital mortality. *Clin Chim Acta* 2006;**372**:1–13. doi:10.1016/J.CCA.2006.03.024

29 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†‡. *BJA Br J Anaesth* 2007;**98**:769–74. doi:10.1093/bja/aem097

30 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.

31 Prytherch DR, Briggs JS, Weaver PC, *et al.* Measuring clinical performance using routinely collected clinical data. *Med Inform Internet Med* 2005;**30**:151–6. doi:10.1080/14639230500298966

32 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. doi:10.1093/qjmed/hcp149

33 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.

34 Mohammed M a., 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**. doi:10.1371/journal.pone.0064340

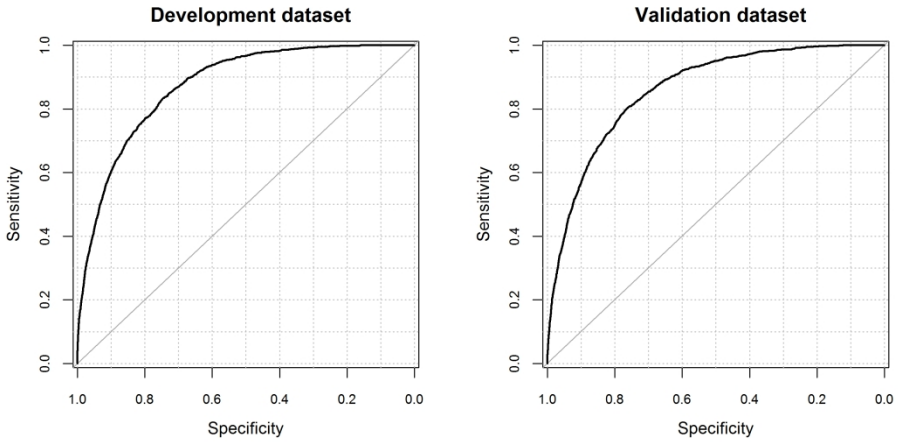
35 Faisal M, Scally A, Richardson D, *et al.* Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test Results. *Crit Care Med* 2018;**46**:612–8. doi:10.1097/CCM.0000000000002967

36 Algorithm for detecting Acute Kidney Injury (AKI) based on serum creatinine changes with time https://www.england.nhs.uk/wp-content/uploads/2014/06/psa-aki-alg.pdf. 2014.

37 Silcock DJ, Corfield AR, Gowens PA, *et al.* Validation of the National Early Warning Score in the prehospital setting. *Resuscitation* 2015;**89**:31–5. doi:10.1016/j.resuscitation.2014.12.029

38 Escobar GJ, Dellinger RP. Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med* 2016;**11**:S5–10. doi:10.1002/jhm.2653

39 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. doi:10.1002/jhm.2652



Area under the Receiver Operating Characteristic curve for development dataset (0.87) and validation dataset (0.86).

254x127mm (300 x 300 DPI)

Supplementary Material

The NEWS [https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news] is based on a scoring system in which a score is allocated to vital signs physiological measurements already undertaken when patients present to, or are being monitored in hospital. Six physiological parameters form the basis of the scoring system:

Physiological Parameters	3	2	1	0	1	2	3
Respiration Rate	≤8		9 - 11	12 - 20		21 - 24	≥25
Oxygen Saturations	≤91	92 - 93	94 - 95	≥96			
Any Supplemental Oxygen		Yes		No			
Temperature	≤35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥39.1	
Systolic BP	≤90	91 - 100	101 - 110	111 - 219			≥220
Heart Rate	≤40		41 - 50	51-90	91 - 110	111 - 130	≥131
Level of Consciousness				Alert			Voice, Pain, or Unconscious

A score is allocated to each as they are measured, the magnitude of the score reflecting how extreme the parameter varies from the norm. This score is then aggregated, and uplifted for people requiring oxygen.

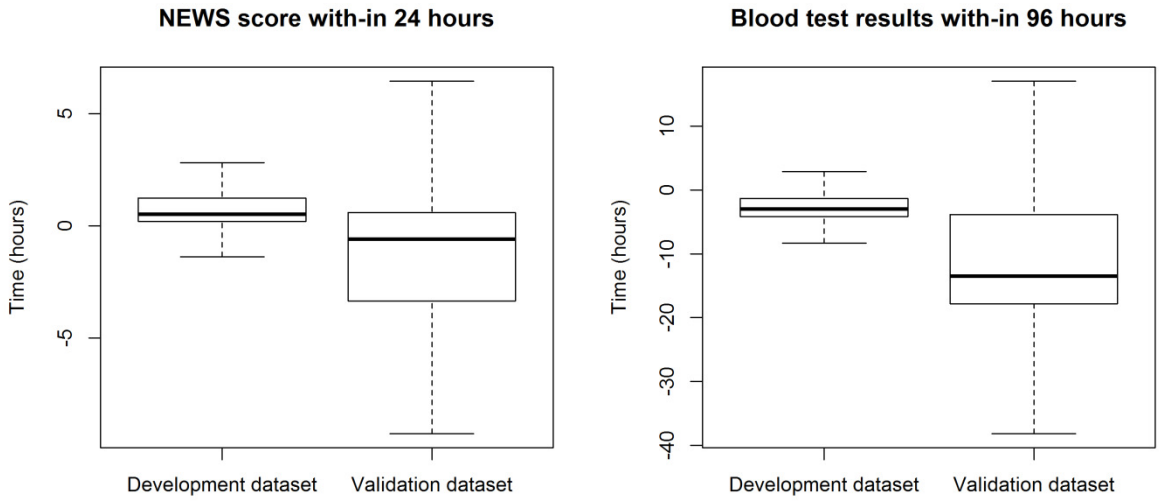


Figure S1 Distribution of time to first NEWS score and Blood test results for development and validation datasets.

Blood tests results recorded	Development dataset N(%)	Validation dataset N(%)
within 24 hours	29255 (94.4)	24341 (92.7)
within 48 hours	894 (2.9)	1098 (4.2)
within 72 hours	512 (1.6)	495 (1.9)
within 96 hours	335 (1.1)	313 (1.2)

Table S1 Distribution of time to the first set of Blood test results recorded within 4 days for development and validation datasets.

Characteristic	Development dataset	Validation dataset
N	6104	10504
Male (%)	3008 (49.3)	4875 (46.4)
Mean Age (SD)	61.4 (20.2)	64.7 (21.4)
Median Length of Stay (days) (IQR)	1.1 (4.0)	1.4 (4.4)
In-hospital mortality (%)	405 (6.6)	434 (4.1)

Table S2 Characteristics of emergency admissions with incomplete data in development and validation datasets.

		Example 1 (Discharged alive)			Example 2 (Discharged dead)		
Variable Name	Coefficient	Values	Transformed	Log(Odds)	Values	Transformed	Log(Odds)
Intercept	-3.22	1	1	-3.22	1	1	-3.22
Male	0.14	1	1	0.14	0	0	0
Age	0.077	51	51	3.927	44	44	3.388
Albumin	-0.104	36	36	-3.744	11	11	-1.144
1/sqrt(Creatinine)	9.883	86	0.107832773	1.065711	153	0.080845	0.798993
Haemoglobin	0.002	148	148	0.296	48	48	0.096
Log(Potassium)	-0.024	3.4	1.223775354	-0.02937	6.9	1.931521	-0.04636
Sodium	-0.023	111	111	-2.553	130	130	-2.99
Log(White Blood Count)	1.167	12	2.484906493	2.899886	34.7	3.546739	4.139045
Log(Urea)	1.211	7.8	2.054123604	2.487544	28.6	3.353407	4.060975
AKI (reference 0)	0	0	0	0	0	0	0
AKI stage 1	0.131	0	0	0	0	0	0
AKI stage 2	0.443	0	0	0	1	1	0.443
AKI stage 3	-0.388	0	0	0	0	0	0
NEWS	0.093	2	2	0.186	9	9	0.837
Log(Respiratory)	0.569	16	2.772588547	1.577603	22	3.091042	1.758803
Temperature	-0.145	36	36	-5.22	36.5	36.5	-5.2925
Log(Systolic)	-0.919	133	4.89034882	-4.49423	112	4.718499	-4.3363
Log(Diastolic)	0.777	79	4.369447577	3.395061	74	4.304065	3.344258
Log(Pulse)	0.511	63	4.143134465	2.117142	149	5.003946	2.557016
Oxygen Saturation	-0.016	94	94	-1.504	95	95	-1.52
Oxygen supplementation	0.606	0	0	0	0	0	0
Alert	0	0	0	0	0	0	0
Pain	0.716	0	0	0	0	0	0
Voice	0.395	0	0	0	0	0	0
Unconscious	1.925	0	0	0	1	1	1.925
Age * Log(White Cell Count)	-0.015	-	126.7302311	-1.90095	-	156.0565	-2.34085
1/sqrt(Creatinine)* Log(White Cell Count)	1.481	-	0.267954358	0.39684	-	0.286737	0.424657
AKI stage 3 * 1/sqrt(Creatinine)	15.551	-	0	0	-	0	0
Sum of Log(Odds)		-	-	**-4.17677**	-	-	**2.882744**
Probability of dying		-	-	**0.015116**	-	-	**0.946987**

Table S3 Coefficient of CARM model to predict in-hospital mortality with two examples (one discharged alive and one discharged died).

We accounted a baseline difference in risk of death in the external validation data by adding 0.52 to the CARM logit model using an iterative procedure described elsewhere[1].

1 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. doi:10.1093/qjmed/hcw110

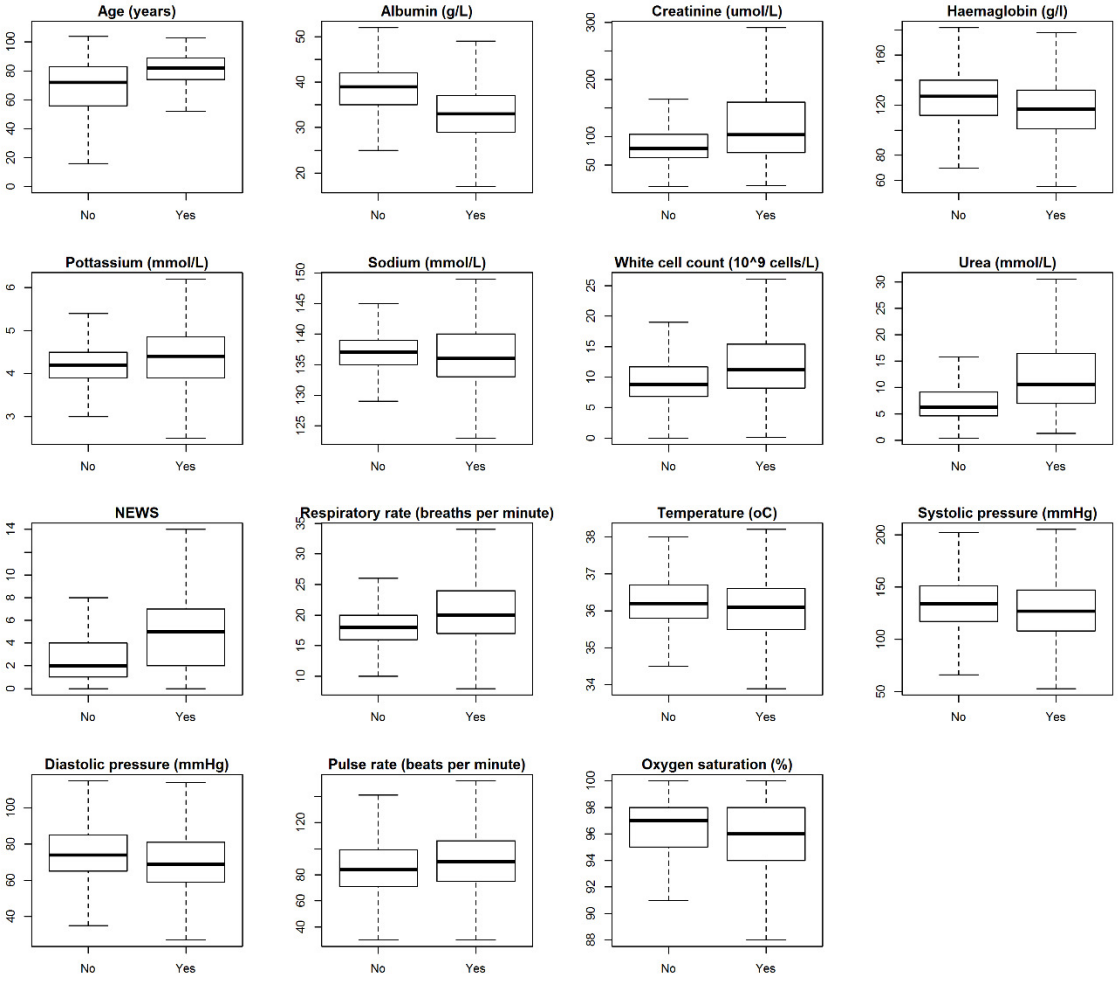


Figure S2 Boxplot without outliers for continuous covariates with respect to patient's discharge status (Alive/Died) for NLAG hospitals

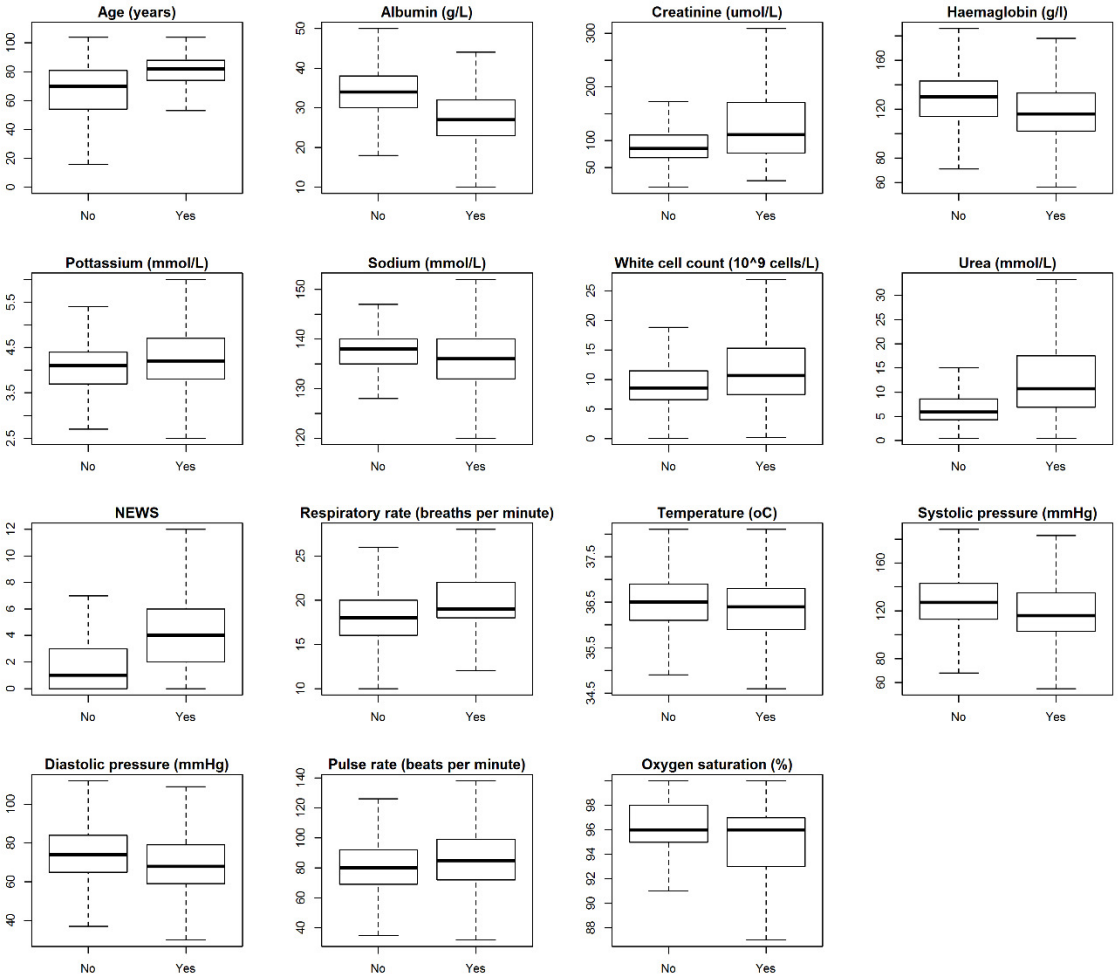


Figure S3 Boxplot without outliers for continuous covariates with respect to patient's discharge status (Alive/Died) for York hospital

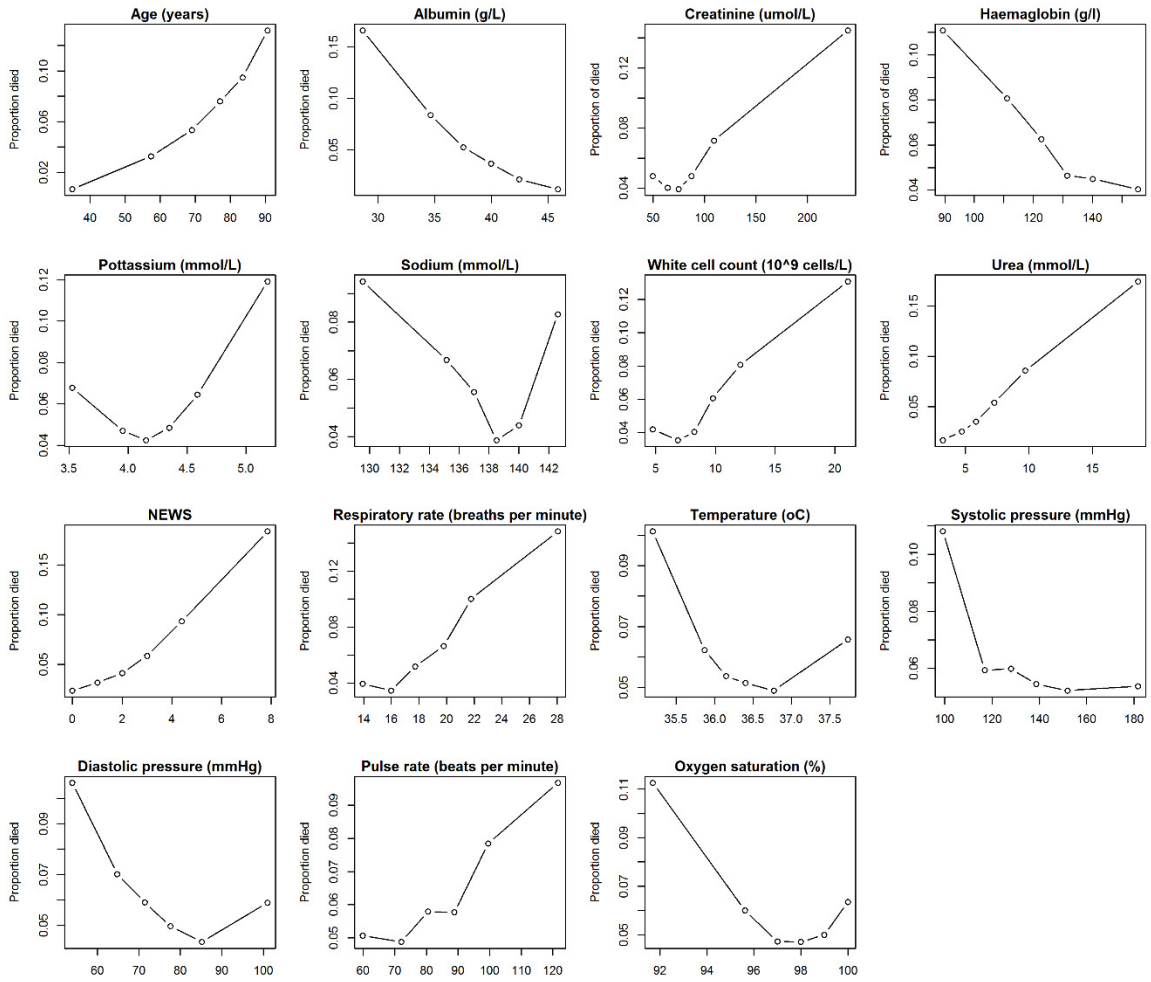


Figure S4 Line plots showing the observed risk of death with continuous covariates for NLAG hospitals

NB: y-axis range changes in each plot.

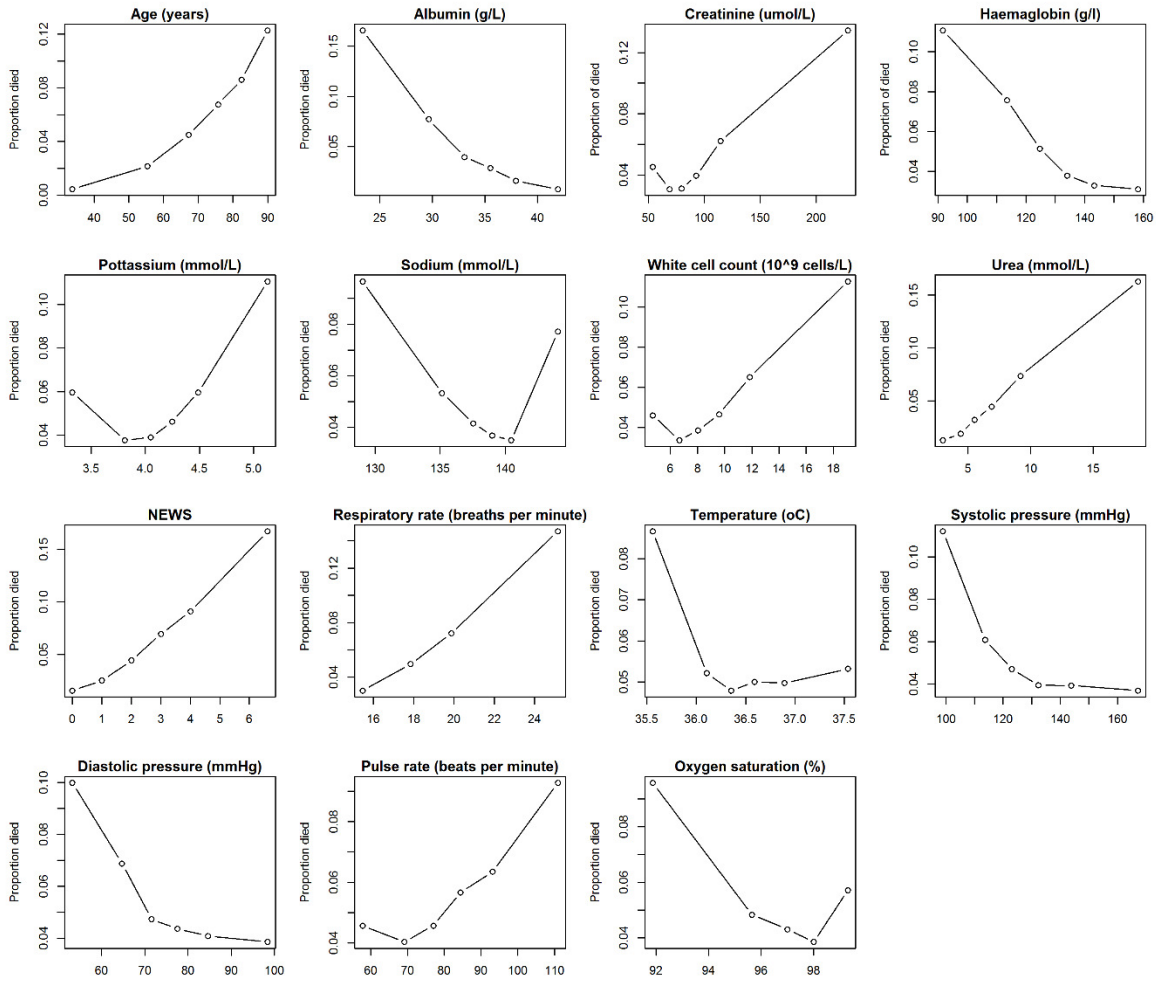


Figure S5 Line plots showing the observed risk of death with continuous covariates for York hospital.

NB: y-axis range changes in each plot.

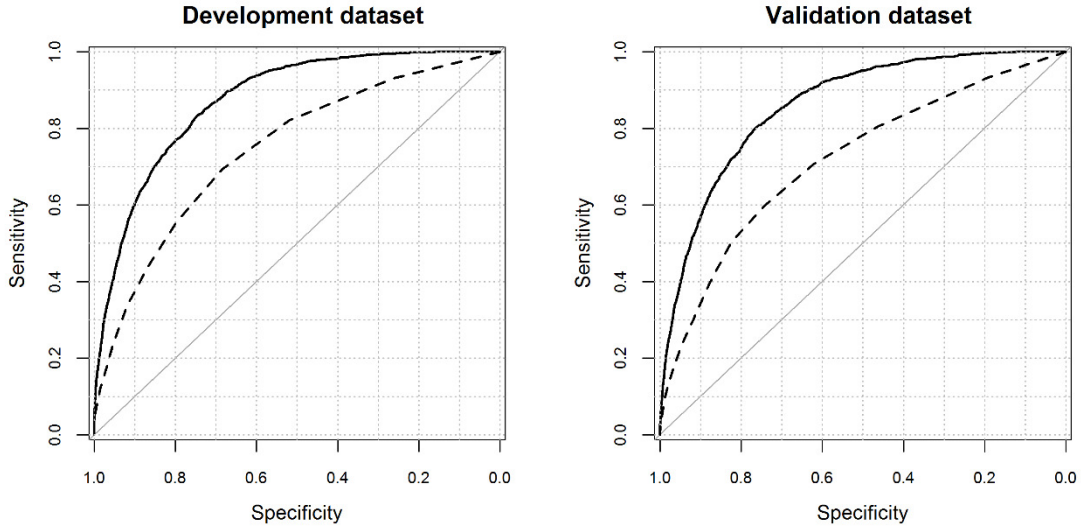


Figure S6 Area under the Receiver Operating Characteristic curve for development dataset (NEWS =0.75, CARM=0.87) and validation dataset (NEWS=0.72, CARM=0.86).

Black dashed line for NEWS and black solid line for CARM

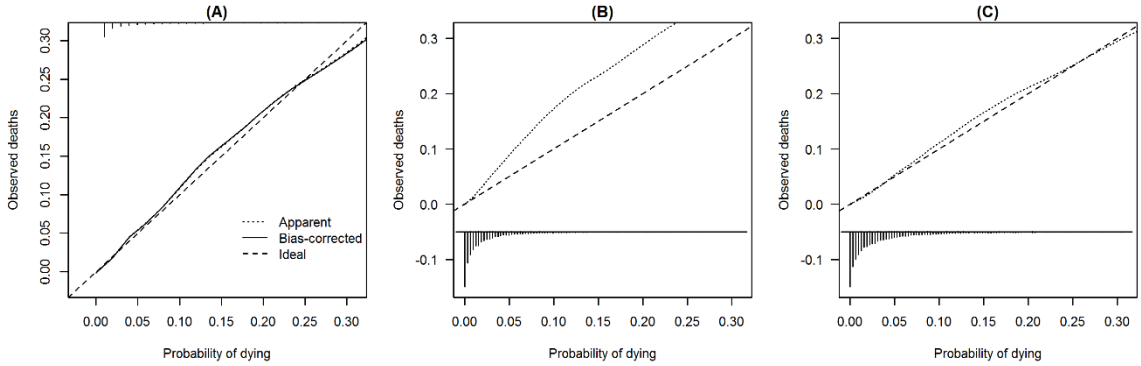


Figure S7 Internal and external validation with and without recalibration of CARM model

(A) Internal validation (B) external validation before recalibration (C) external validation after recalibration

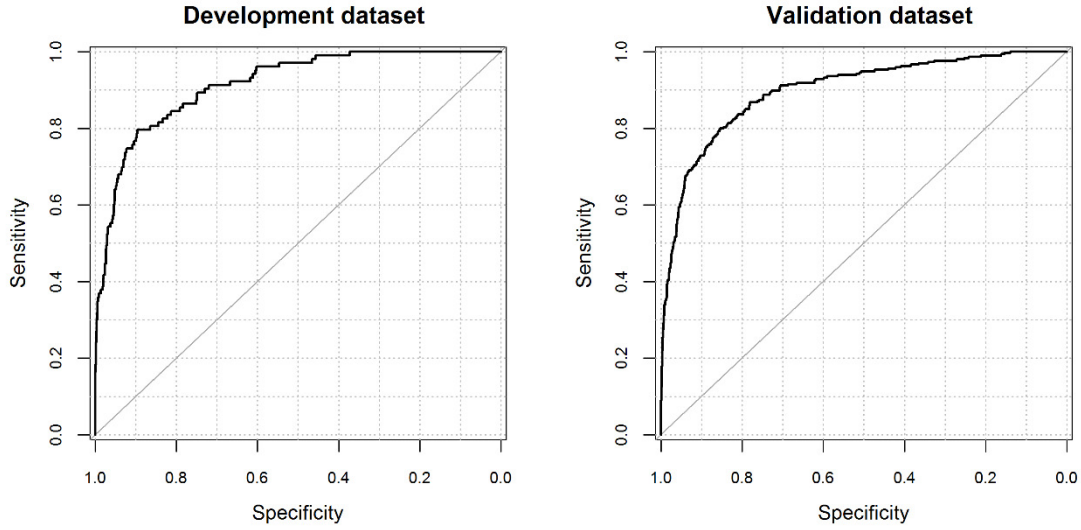


Figure S8 Receiver Operating Characteristic curve of (median) imputed blood tests results on development dataset and validation dataset.

NB: patients with imputed values were omitted during model development and validation.

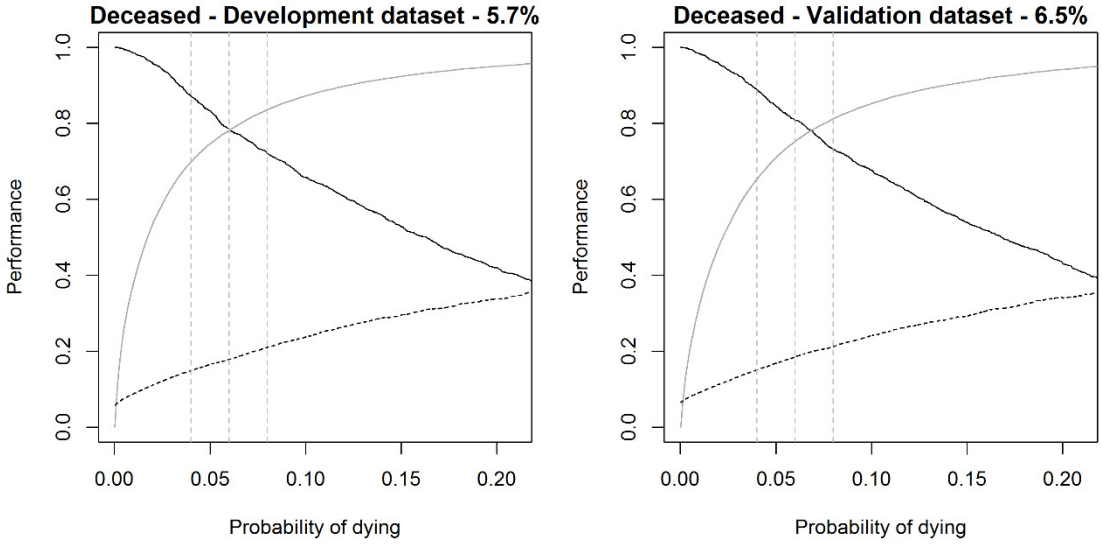


Figure S9: Sensitivity analysis of CARM model at various thresholds of probability of dying (0.0, 0.01, …, 0.20) on development dataset and validation dataset.

Black solid line is for sensitivity and black dashed line is for positive predictive value (PPV). Grey solid line is specificity and grey dashed vertical lines are at thresholds (0.04, 0.06, and 0.08).

NB: We selected thresholds exclusively based on development dataset.

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	6
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	6
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7
	5b	D;V	Describe eligibility criteria for participants.	7
	5c	D;V	Give details of treatments received, if relevant.	
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	8
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	D;V	Explain how the study size was arrived at.	7
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	7
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	8
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	8
	10c	V	For validation, describe how the predictions were calculated.	8
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	8,9
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	suppl
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	8,9
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	10,11
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	11
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	10
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	suppl
	15b	D	Explain how to the use the prediction model.	suppl
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	11
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	12
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	12
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	12,13
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	13
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	suppl
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	14

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-022939.R2
Article Type:	Research
Date Submitted by the Author:	24-Sep-2018
Complete List of Authors:	Faisal, Muhammad; University of Bradford; Bradford Institute for Health Research Scally, Andrew ; University College Cork Jackson, Natalie; Bradford Institute for Health Research Richardson, Donald; York Teaching Hospital NHS Foundation Trust Beatson, Kevin; York Teaching Hospital NHS Foundation Trust Howes, Robin; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Speed, Kevin; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Menon, Madhav; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Daws, Jeremey; Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Dyson, Judith; University of Hull, Department of Community Health and Social Care Marsh, Claire; Bradford Institute for Health Research Mohammed, Mohammed ; University of Bradford Faculty of Health Studies; Bradford Institute for Health Research
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health informatics
Keywords:	computer aided risk score, hospital mortality, vital signs and blood test, national early warning score, emergency admission

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

Muhammad Faisal
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

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

Natalie Jackson
Health Informatics Senior Project Manager
Improvement Academy, Bradford Institute for Health Research
E-mail: Natalie.jackson@yhahsn.nhs.uk

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

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

Robin Howes
Operational Manager for Electronic Patient Records
Department of Strategy & Planning
Northern Lincolnshire and Goole Hospitals
E-mail: robin.howes@nhs.net

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

Madhav Menon
Consultant in respiratory medicine
Northern Lincolnshire and Goole Hospitals

E-mail: madhav.menon1@nhs.net

Jeremy Daws
Head of Quality Assurance
Northern Lincolnshire and Goole Hospitals
E-mail: Jeremy.daws@nhs.net

Judith Dyson
Senior Lecturer Mental Health
University Of Hull
E-mail: J.Dyson@hull.ac.uk

Claire Marsh
Senior Research Fellow/ Patient & Public Engagement Lead
Improvement Academy, Bradford Institute for Health Research
E-mail: claire.marsh@bthft.nhs.uk

Mohammed A Mohammed
Professor of Healthcare Quality & Effectiveness
Faculty of Health Studies, University of Bradford, Bradford, UK
Deputy Director of the Bradford Institute for Health Research
Academic Director to the Yorkshire & Humberside Academic Health Sciences Network
E-mail: M.A.Mohammed5@Bradford.ac.uk

Corresponding author: 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 (NH – model development data; YH – external validation data).

Participants: Adult (>=16 years) medical admissions discharged over a 24 month period with electronic NEWS and blood test results recorded on admission.

Results: The risk of in-hospital mortality following emergency medical admission was 5.7% (NH: 1766/30996) and 6.5% (YH: 1703/26247). 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 tests 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.

Key words: computer aided risk score, hospital mortality, vital signs and blood test, national early warning score, emergency admission

Article Summary

- 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 NEWS 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 data sets – 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 United Kingdom (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 Royal College of Physicians (London) [3]. 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 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 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 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 National Health Service (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 & 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). 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, YH respectively) 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 (age≥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 tests results recorded within 4 days (96 hours) of admission (>90% of blood test results were within ±24 hours of admission - see table S1 in appendix).

For model development purposes, we were unable to consider emergency admissions without complete blood test results and NEWS recorded – this constituted 16.5% (6104/37100) of records in NH and 28.6% (10504/36751) 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 figure S1 in appendix).

Development of a Computer Aided Risk of Mortality (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 Acute Kidney Injury (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 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* [11]), 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 variable continious covariate and chose the following transformations:- (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 [12] in *R* [13].

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 [14], using data from another hospital (YH) (the external validation dataset). The data from YH is 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 [15] in R to estimate statistical optimism [14,15].

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 [14]. 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 0–100%. 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' [16]. 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 [17].

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 [14]. 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, whilst a perfect model has a c-statistic of 1. The higher the c-statistic, the better the model. In general, values less than 0.7 are considered to show poor discrimination, values of 0.7–0.8 can be described as reasonable, and values above 0.8 suggest good discrimination [18]. The 95% confidence interval for the c-statistic was derived using DeLong's method as implemented in the *pROC* library [19] in *R* [13]. 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 [14]. We followed the TRIPOD guidelines for model development and validation [20]. All analyses were carried using *R* [13] and *Stata* [11].

Ethical 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.

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 five years. The patient and public voice continued to be included throughout the project with three patient representatives invited to sit on 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.

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.

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% (6104/37100) in NH and 28.6% (10504/36751) 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, 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 Table S2 in appendix for characteristic of emergency admissions with incomplete data).

The in-hospital mortality was 5.7% (1766/30996) in NH and 6.5% (1703/26247) 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.

Figures S2 to S5 (see appendix) 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 (figures S2 & S3 in appendix) 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 the appendix (figures S4 & S5 in appendix) 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 the appendix (see 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 Figure S6 in the appendix for ROC plots comparing CARM versus NEWS). The c-statistic was high in the development dataset 0.87 [95% CI 0.86 - 0.88] and the external validation dataset 0.86 [95% CI 0.85 - 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 appendix Figure S7). The final CARM model, which is not intended for paper-based use, is shown in the appendix with accompanying internal and external validation plots (see appendix 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 (see Figure S8 appendix for corresponding ROC plots).

Table 4 shows the sensitivity, specificity and positive & 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 (more than 80% in development and validation datasets – see table 4 and figure S9 in appendix). Furthermore, the CARM model performance is good in each hospital in various subgroups such as, by sex, age, seasons, longer vs. shorter length of stay admissions, day of the week, and 16 CCI disease groups (see table S4 in appendix).

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.

Whilst several previous studies [1] have used blood test results [21–28] or patient physiology [29,30] to predict the risk of in-hospital mortality, few studies have combined these two data sources [31–34] 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 [35].

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 [1,18] and

its usefulness in clinical practice remains to be seen. We found that up to about ¼ 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 was 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 [36]. 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 [37]. 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 becomes 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) [38,39]. 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. Whilst 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 [38,39]. For example, whilst 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, whilst 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.

Acknowledgements

The authors would like to express their gratitude to Sue Rushbrook and Gary Hardcastle at York Hospital for their support in obtaining the data frame, and to Laura Lamming at the University of Bradford for her support with the Ethics application.

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.

Contributorship

MAM & DR had the original idea for this work. NJ was overall study coordinator with JeD as local NLAG coordinator. MF, AS and MAM undertook the statistical analyses. JuD, CM & 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.

Competing Interests: The authors declare no conflicts of interest.

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

Characteristic	Development dataset	Validation dataset
	N (%)	N (%)
Total emergency medical admissions	37100	36751
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)	10504 (28.6)
Total included (%)	30996 (83.5)	26247 (71.4)

Table 2 Characteristics of emergency admissions for development and validation datasets.

Characteristic	Development dataset (NH)		Validation dataset (YH)	
	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 [umol/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 minute] (SD)	18 (3.5)	20.1 (4.8)	18.6 (4.6)	21.7 (6.8)
Mean Temperature [^{o}C] (SD)	36.5 (0.7)	36.3 (0.8)	36.3 (0.8)	36.1 (1.1)
Mean Systolic pressure [mmHg] (SD)	129.6 (22.7)	119.8 (24.8)	136.1 (27.2)	128.5 (30.3)
Mean Diastolic pressure [mmHg] (SD)	75 (14.8)	69.5 (15.8)	75.4 (15.5)	71.3 (17.7)
Mean Pulse rate [beats per minute] (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)

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]

NB:

† mean predicted risk difference between who discharged died and discharged alive.
‡ Died in-hospital following emergency admission
corrected optimism (original = 0.874, and corrected=0.873).

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 > cutoff	%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)

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

Figure 1 Area under the Receiver Operating Characteristic curve for development dataset (0.87) and validation dataset (0.86).

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. doi:10.1186/1757-7241-18-8

2 Silke B, Kellett J, Rooney T, *et al.* An improved medical admissions risk system using multivariable fractional polynomial logistic regression modelling. *QJM An Int J Med* 2010;**103**:23–32. doi:10.1093/qjmed/hcp149

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. doi:10.1016/j.resuscitation.2006.03.008

6 Edwards M, McKay H, Van Leuvan C, *et al.* Modified Early Warning Scores: inaccurate summation or inaccurate assignment of score? *Crit Care* 2010;**14**:P257. doi:10.1186/cc8489

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. doi:10.1016/j.resuscitation.2005.12.002

8 Mohammed M, Hayton R, Clements G, *et al.* Improving accuracy and efficiency of early warning scores in acute care. *Br J Nurs*;**18**:18–24. doi:10.12968/bjon.2009.18.1.32072

9 NHS England: Acute Kidney Injury (AKI) Programme. http://www.thinkkidneys.nhs.uk, 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 StatCorp. Stata: Release 13. Statistical Software. College Station, TX: StataCorp LP; 2014.

12 Venables W, Ripley B. *Modern Applied Statistics with S*. Fourth Edi. New York: : Springer 2002.

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

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

15 Harrell FE. rms: Regression Modeling Strategies http://cran.r-project.org/package=rms. 2015.

16 Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;**45**:562–5. doi:10.1093/biomet/45.3-4.562

17 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. doi:10.1097/EDE.0b013e3181c30fb2

18 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36. doi:10.1148/radiology.143.1.7063747

19 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. doi:10.1186/1471-2105-12-77

20 Moons KGM, 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. doi:10.7326/M14-0698

21 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. doi:10.1002/bjs.4274

22 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. doi:10.1016/j.resuscitation.2005.02.011

23 Pine M, Jones B, Lou YB. Laboratory values improve predictions of hospital mortality. *Int J Qual Heal care J Int Soc Qual Heal Care* 1998;**10**:491–501.

24 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. doi:10.1515/CCLM.2005.093

25 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. doi:10.1093/bja/aei122

26 Froom P, Shimoni Z. Prediction of Hospital Mortality Rates by Admission Laboratory Tests. *Clin Chem* 2005;**52**:325–8. doi:10.1373/clinchem.2005.059030

27 Asadollahi K, Hastings IM, Gill G V, *et al.* Prediction of hospital mortality from admission laboratory data and patient age: A simple model. *Emerg Med Australas* 2011;**23**:354–63. doi:10.1111/j.1742-6723.2011.01410.x

28 ten Boekel E, Vroonhof K, Huisman A, *et al.* Clinical laboratory findings associated with in-hospital mortality. *Clin Chim Acta* 2006;**372**:1–13. doi:10.1016/J.CCA.2006.03.024

29 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†‡. *BJA Br J Anaesth* 2007;**98**:769–74. doi:10.1093/bja/aem097

30 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.

31 Prytherch DR, Briggs JS, Weaver PC, *et al.* Measuring clinical performance using routinely collected clinical data. *Med Inform Internet Med* 2005;**30**:151–6. doi:10.1080/14639230500298966

32 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. doi:10.1093/qjmed/hcp149

33 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.

34 Mohammed M a., 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**. doi:10.1371/journal.pone.0064340

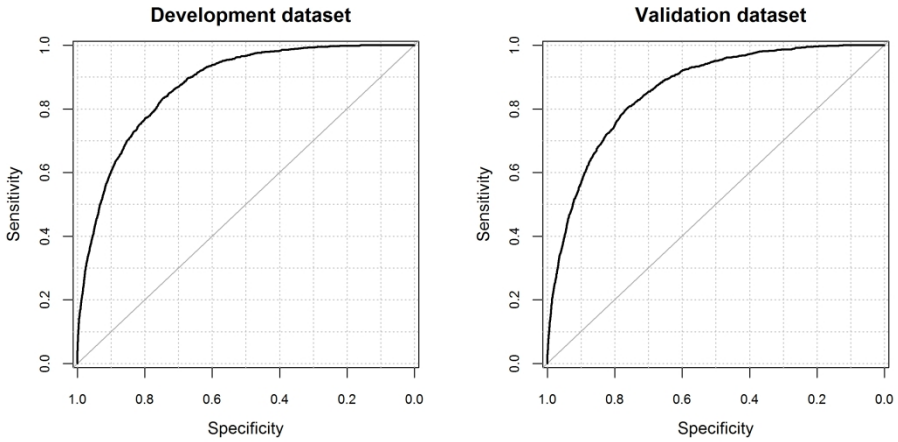
35 Faisal M, Scally A, Richardson D, *et al.* Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test Results. *Crit Care Med* 2018;**46**:612–8. doi:10.1097/CCM.0000000000002967

36 Algorithm for detecting Acute Kidney Injury (AKI) based on serum creatinine changes with time https://www.england.nhs.uk/wp-content/uploads/2014/06/psa-aki-alg.pdf. 2014.

37 Silcock DJ, Corfield AR, Gowens PA, *et al.* Validation of the National Early Warning Score in the prehospital setting. *Resuscitation* 2015;**89**:31–5. doi:10.1016/j.resuscitation.2014.12.029

38 Escobar GJ, Dellinger RP. Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med* 2016;**11**:S5–10. doi:10.1002/jhm.2653

39 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. doi:10.1002/jhm.2652



Area under the Receiver Operating Characteristic curve for development dataset (0.87) and validation dataset (0.86).

254x127mm (300 x 300 DPI)

Supplementary Material

The NEWS [https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news] is based on a scoring system in which a score is allocated to vital signs physiological measurements already undertaken when patients present to, or are being monitored in hospital. Six physiological parameters form the basis of the scoring system:

Physiological Parameters	3	2	1	0	1	2	3
Respiration Rate	≤8		9 - 11	12 - 20		21 - 24	≥25
Oxygen Saturations	≤91	92 - 93	94 - 95	≥96			
Any Supplemental Oxygen		Yes		No			
Temperature	≤35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥39.1	
Systolic BP	≤90	91 - 100	101 - 110	111 - 219			≥220
Heart Rate	≤40		41 - 50	51-90	91 - 110	111 - 130	≥131
Level of Consciousness				Alert			Voice, Pain, or Unconscious

A score is allocated to each as they are measured, the magnitude of the score reflecting how extreme the parameter varies from the norm. This score is then aggregated, and uplifted for people requiring oxygen.

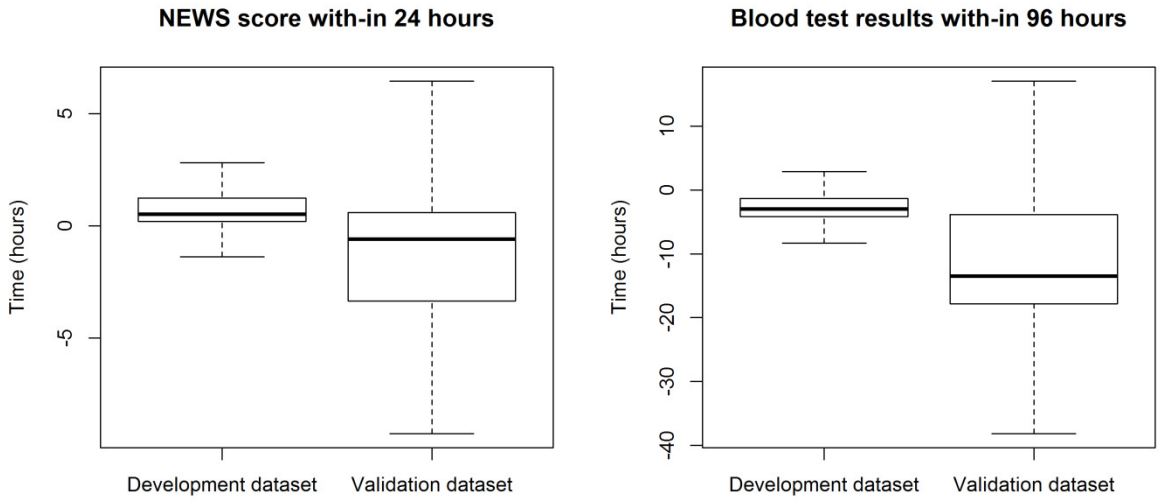


Figure S1 Distribution of time to first NEWS score and Blood test results for development and validation datasets.

Blood tests results recorded	Development dataset N(%)	Validation dataset N(%)
within 24 hours	29255 (94.4)	24341 (92.7)
within 48 hours	894 (2.9)	1098 (4.2)
within 72 hours	512 (1.6)	495 (1.9)
within 96 hours	335 (1.1)	313 (1.2)

Table S1 Distribution of time to the first set of Blood test results recorded within 4 days for development and validation datasets.

Characteristic	Development dataset	Validation dataset
N	6104	10504
Male (%)	3008 (49.3)	4875 (46.4)
Mean Age (SD)	61.4 (20.2)	64.7 (21.4)
Median Length of Stay (days) (IQR)	1.1 (4.0)	1.4 (4.4)
In-hospital mortality (%)	405 (6.6)	434 (4.1)

Table S2 Characteristics of emergency admissions with incomplete data in development and validation datasets.

		Example 1 (Discharged alive)			Example 2 (Discharged dead)		
Variable Name	Coefficient	Values	Transformed	Log(Odds)	Values	Transformed	Log(Odds)
Intercept	-3.22	1	1	-3.22	1	1	-3.22
Male	0.14	1	1	0.14	0	0	0
Age	0.077	51	51	3.927	44	44	3.388
Albumin	-0.104	36	36	-3.744	11	11	-1.144
1/sqrt(Creatinine)	9.883	86	0.107832773	1.065711	153	0.080845	0.798993
Haemoglobin	0.002	148	148	0.296	48	48	0.096
Log(Potassium)	-0.024	3.4	1.223775354	-0.02937	6.9	1.931521	-0.04636
Sodium	-0.023	111	111	-2.553	130	130	-2.99
Log(White Blood Count)	1.167	12	2.484906493	2.899886	34.7	3.546739	4.139045
Log(Urea)	1.211	7.8	2.054123604	2.487544	28.6	3.353407	4.060975
AKI (reference 0)	0	0	0	0	0	0	0
AKI stage 1	0.131	0	0	0	0	0	0
AKI stage 2	0.443	0	0	0	1	1	0.443
AKI stage 3	-0.388	0	0	0	0	0	0
NEWS	0.093	2	2	0.186	9	9	0.837
Log(Respiratory)	0.569	16	2.772588547	1.577603	22	3.091042	1.758803
Temperature	-0.145	36	36	-5.22	36.5	36.5	-5.2925
Log(Systolic)	-0.919	133	4.89034882	-4.49423	112	4.718499	-4.3363
Log(Diastolic)	0.777	79	4.369447577	3.395061	74	4.304065	3.344258
Log(Pulse)	0.511	63	4.143134465	2.117142	149	5.003946	2.557016
Oxygen Saturation	-0.016	94	94	-1.504	95	95	-1.52
Oxygen supplementation	0.606	0	0	0	0	0	0
Alert	0	0	0	0	0	0	0
Pain	0.716	0	0	0	0	0	0
Voice	0.395	0	0	0	0	0	0
Unconscious	1.925	0	0	0	1	1	1.925
Age * Log(White Cell Count)	-0.015	-	126.7302311	-1.90095	-	156.0565	-2.34085
1/sqrt(Creatinine)* Log(White Cell Count)	1.481	-	0.267954358	0.39684	-	0.286737	0.424657
AKI stage 3 * 1/sqrt(Creatinine)	15.551	-	0	0	-	0	0
Sum of Log(Odds)		-	-	**-4.17677**	-	-	**2.882744**
Probability of dying		-	-	**0.015116**	-	-	**0.946987**

Table S3 Coefficient of CARM model to predict in-hospital mortality with two examples (one discharged alive and one discharged died).

We accounted a baseline difference in risk of death in the external validation data by adding 0.52 to the CARM logit model using an iterative procedure described elsewhere[1].

1 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. doi:10.1093/qjmed/hcw110

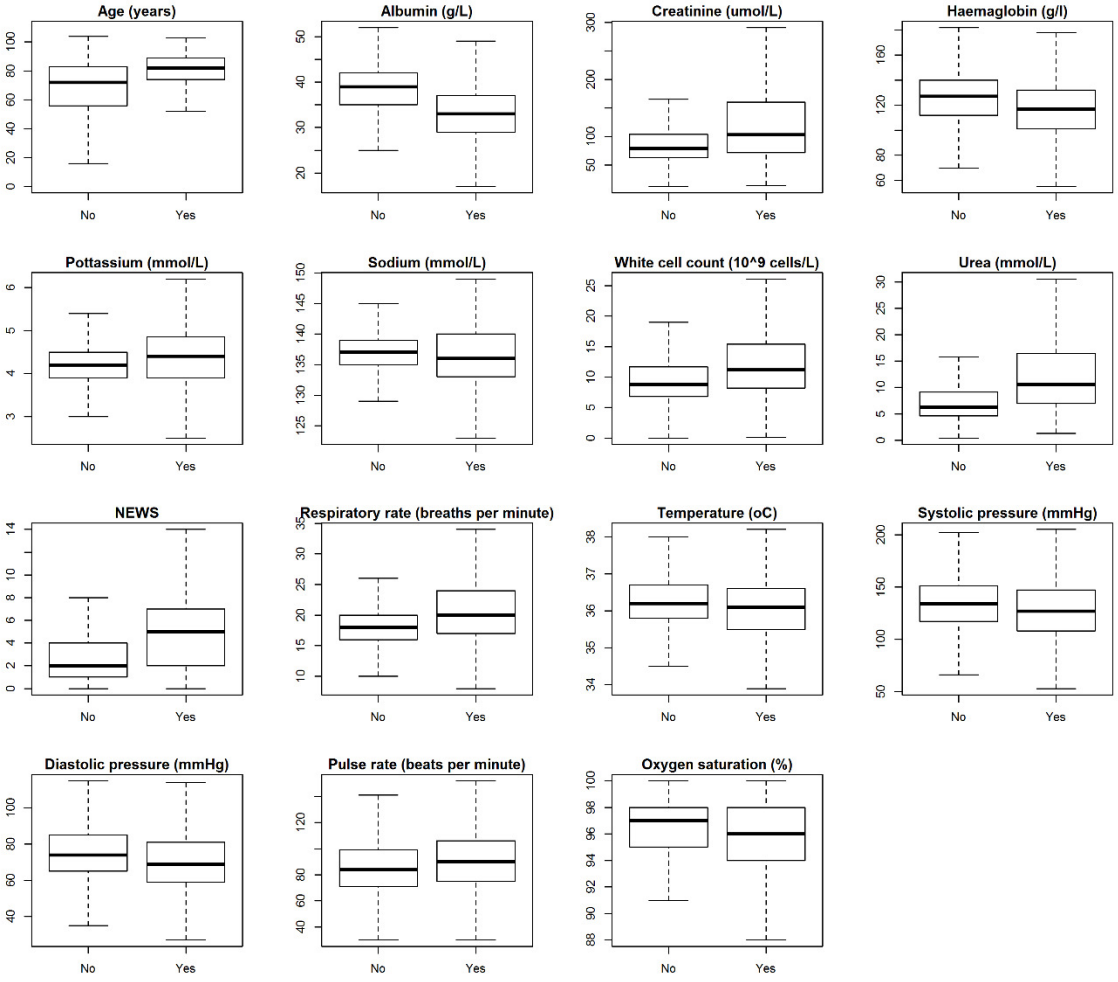


Figure S2 Boxplot without outliers for continuous covariates with respect to patient's discharge status (Alive/Died) for NLAG hospitals

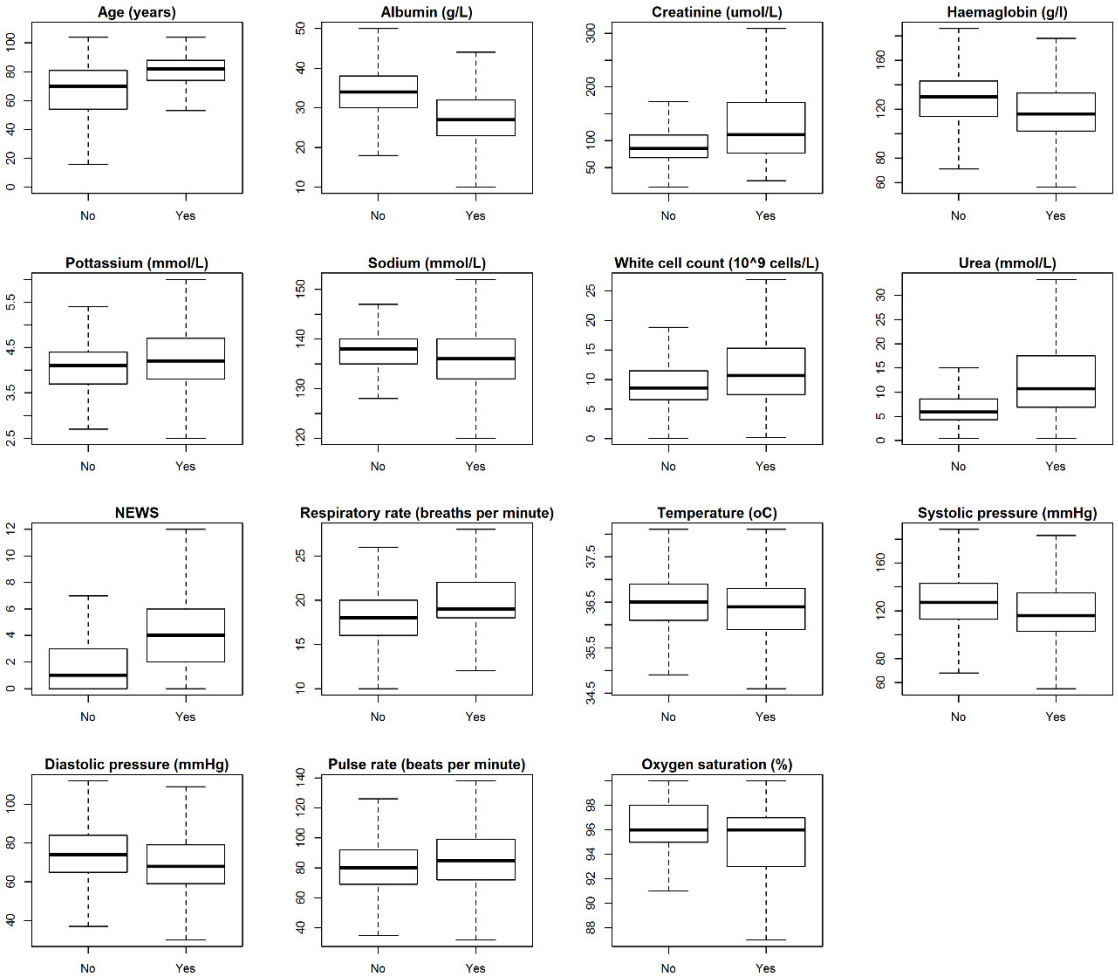


Figure S3 Boxplot without outliers for continuous covariates with respect to patient's discharge status (Alive/Died) for York hospital

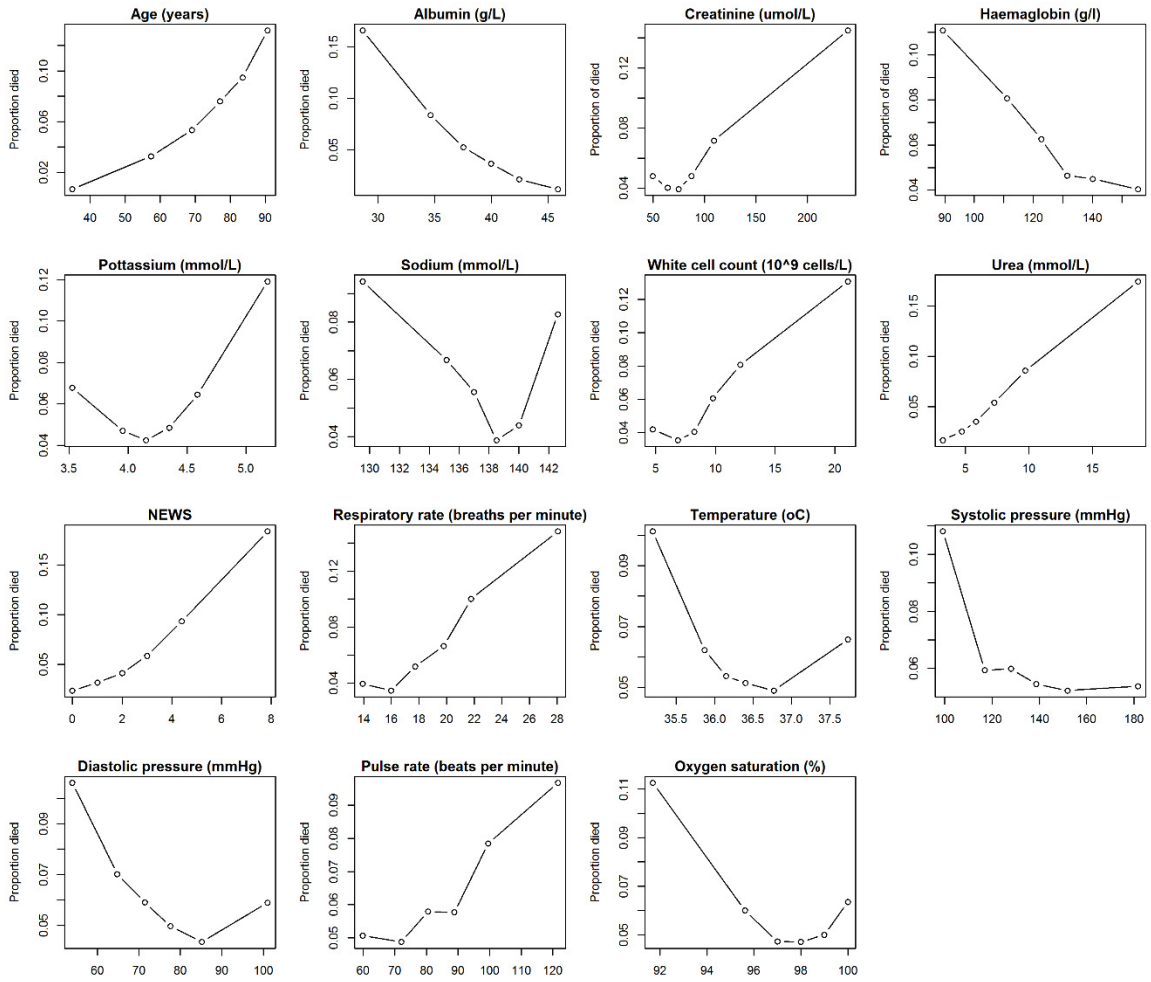


Figure S4 Line plots showing the observed risk of death with continuous covariates for NLAG hospitals

NB: y-axis range changes in each plot.

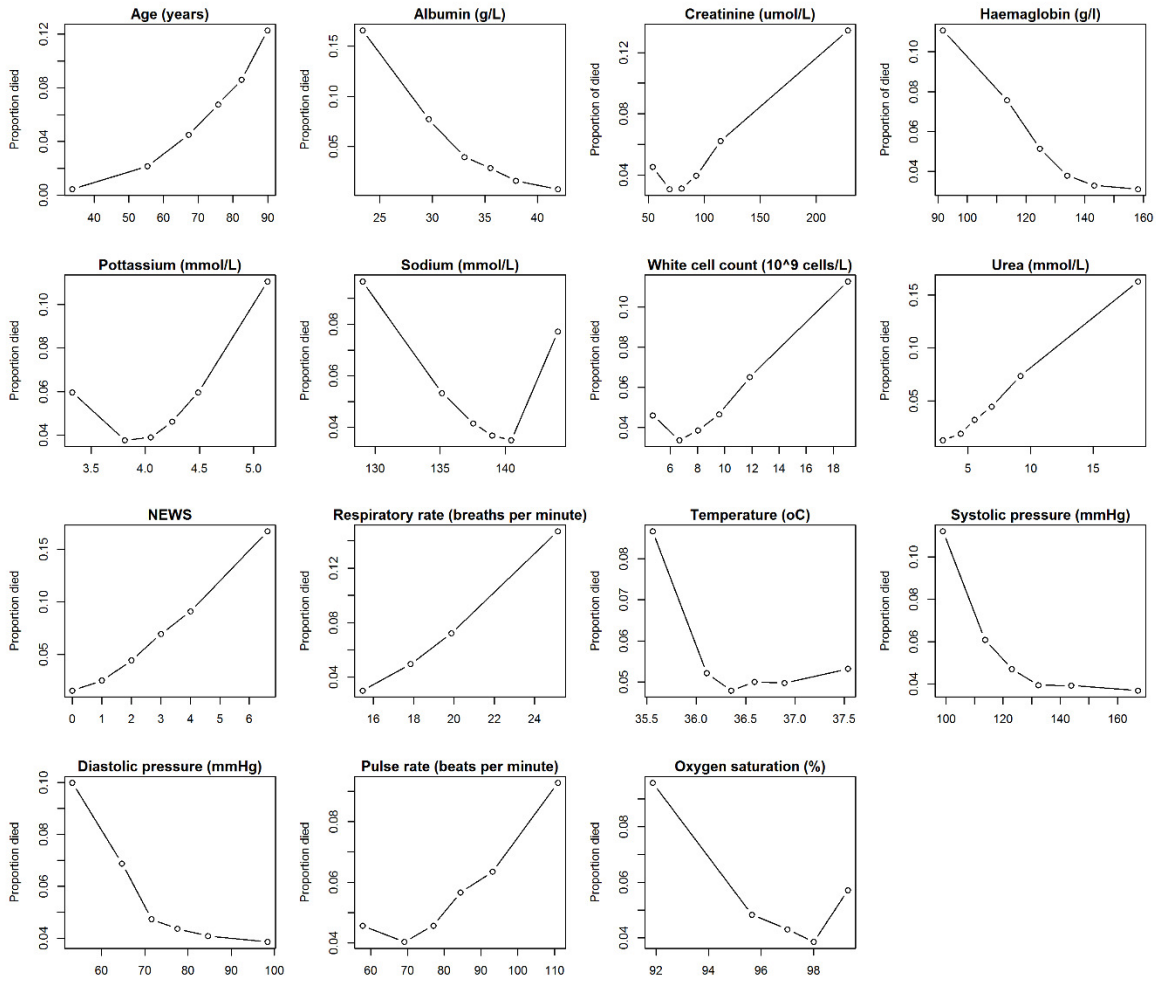


Figure S5 Line plots showing the observed risk of death with continuous covariates for York hospital.

NB: y-axis range changes in each plot.

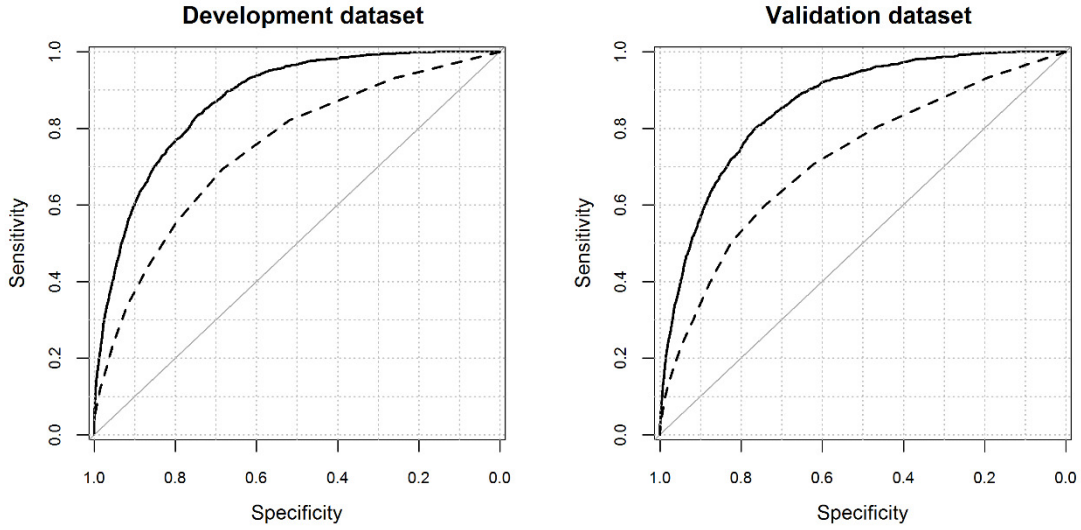


Figure S6 Area under the Receiver Operating Characteristic curve for development dataset (NEWS =0.75, CARM=0.87) and validation dataset (NEWS=0.72, CARM=0.86).

Black dashed line for NEWS and black solid line for CARM

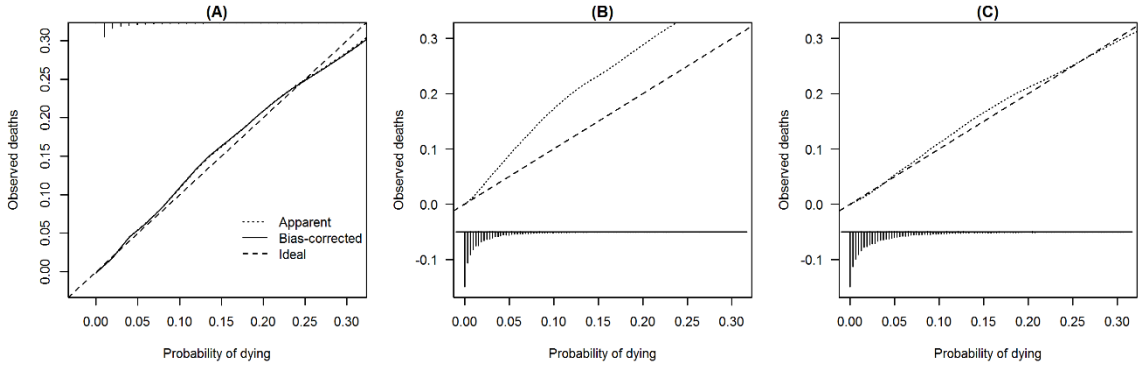


Figure S7 Internal and external validation with and without re-calibration of CARM model

(A) Internal validation of development dataset only using bootstrap method (B) performance of CARM model on external validation dataset without any re-calibration (C) performance of CARM model on external validation dataset after correcting for baseline mortality difference (5.7% vs 6.5%).

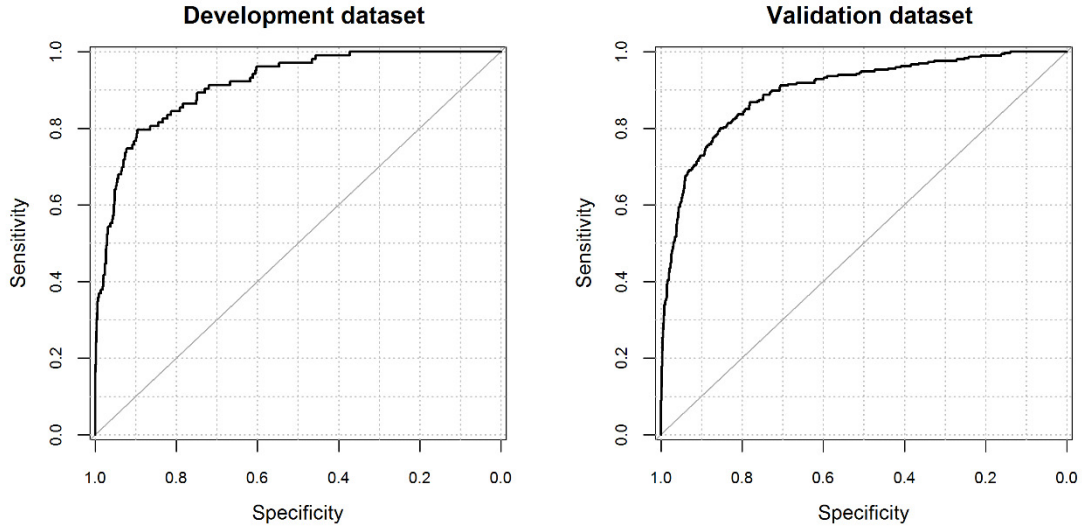


Figure S8 Receiver Operating Characteristic curve of (median) imputed blood tests results on development dataset and validation dataset.

NB: patients with imputed values were omitted during model development and validation.

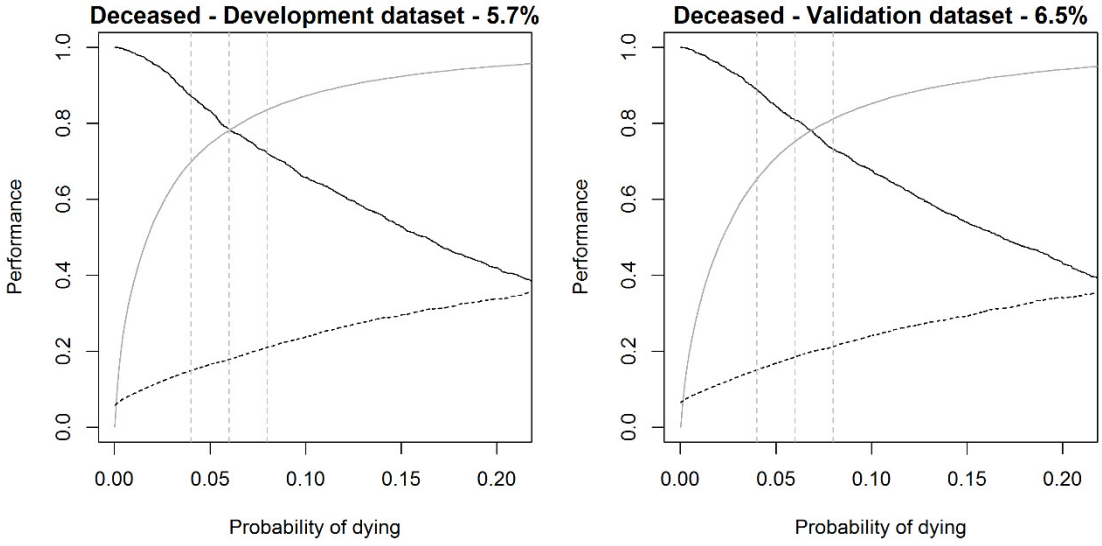


Figure S9: Sensitivity analysis of CARM model at various thresholds of probability of dying (0.0, 0.01, …, 0.20) on development dataset and validation dataset.

Black solid line is for sensitivity and black dashed line is for positive predictive value (PPV). Grey solid line is specificity and grey dashed vertical lines are at thresholds (0.04, 0.06, and 0.08).

NB: We selected thresholds exclusively based on development dataset.

Characteristics	NLAG AUC (95% CI)	York AUC (95% CI)
Overall	0.87 (0.87 to 0.88)	0.86 (0.85 to 0.87)
Sex		
Male	0.87 (0.86 to 0.88)	0.85 (0.84 to 0.86)
Female	0.88 (0.87 to 0.89)	0.87 (0.86 to 0.88)
Age		
Age>=75	0.81 (0.80 to 0.82)	0.81 (0.79 to 0.82)
Age<75	0.91 (0.90 to 0.92)	0.89 (0.88 to 0.91)
Seasons		
Spring	0.87 (0.86 to 0.89)	0.86 (0.84 to 0.88)
Summer	0.88 (0.87 to 0.90)	0.85 (0.83 to 0.87)
Autumn	0.87 (0.86 to 0.89)	0.87 (0.86 to 0.89)
Winter	0.87 (0.85 to 0.88)	0.86 (0.84 to 0.87)
Length of Stay (LoS)		
LoS>=5 days	0.77 (0.76 to 0.79)	0.75 (0.73 to 0.76)
Los<5 days	0.95 (0.94 to 0.96)	0.94 (0.93 to 0.95)
Day of the Week		
Sunday	0.93 (0.91 to 0.94)	0.93 (0.92 to 0.95)
Monday	0.87 (0.85 to 0.89)	0.87 (0.84 to 0.89)
Tuesday	0.87 (0.85 to 0.89)	0.83 (0.81 to 0.85)
Wednesday	0.86 (0.84 to 0.88)	0.84 (0.82 to 0.86)
Thursday	0.86 (0.84 to 0.88)	0.86 (0.84 to 0.88)
Friday	0.88 (0.86 to 0.90)	0.87 (0.85 to 0.89)
Saturday	0.90 (0.88 to 0.92)	0.91 (0.89 to 0.92)
Charlson Comorbidity Index		
Acute Myocardial	0.84 (0.82 to 0.86)	0.81 (0.78 to 0.84)
Congestive Heart	0.77 (0.75 to 0.79)	0.75 (0.73 to 0.78)
Peripheral Vascular	0.78 (0.74 to 0.82)	0.82 (0.79 to 0.86)
Cerebrovascular	0.78 (0.73 to 0.83)	0.78 (0.75 to 0.81)
Dementia	0.78 (0.75 to 0.81)	0.77 (0.73 to 0.80)
COPD	0.85 (0.84 to 0.87)	0.85 (0.83 to 0.86)
Rheumatoid Disease	0.87 (0.84 to 0.90)	0.83 (0.79 to 0.88)
Peptic Ulcer	0.88 (0.83 to 0.93)	0.83 (0.74 to 0.91)
Mild LD (Liver)	0.86 (0.82 to 0.90)	0.83 (0.76 to 0.90)
Diabetes	0.85 (0.83 to 0.87)	0.84 (0.82 to 0.86)
Diabetes+Complications	0.79 (0.69 to 0.89)	0.90 (0.85 to 0.95)
Hemiplegia/Paraplegia	0.80 (0.75 to 0.85)	0.74 (0.67 to 0.81)
RD (Renal)	0.80 (0.78 to 0.82)	0.81 (0.78 to 0.83)
Cancer	0.80 (0.78 to 0.83)	0.81 (0.78 to 0.84)
Moderate/Severe LD (Liver)	0.80 (0.73 to 0.87)	0.78 (0.71 to 0.84)
Metastatic Cancer	0.77 (0.74 to 0.80)	0.77 (0.74 to 0.81)

Table S4: The c-statistics (95% CI) is showing for CARM model in each hospital by Sex, Age, Seasons, Longer vs. shorter length of stay subjects, Day of the week, and 16 CCI disease groups.

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	6
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	6
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7
	5b	D;V	Describe eligibility criteria for participants.	7
	5c	D;V	Give details of treatments received, if relevant.	
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	8
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	D;V	Explain how the study size was arrived at.	7
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	7
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	8
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	8
	10c	V	For validation, describe how the predictions were calculated.	8
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	8,9
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	suppl
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	8,9
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	10,11
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	11
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	10
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	suppl
	15b	D	Explain how to the use the prediction model.	suppl
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	11
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	12
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	12
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	12,13
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	13
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	suppl
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	14

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.