medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Evaluation and Improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study

Adding common blood and physiological parameters to NEWS2 can improve risk stratification for severe COVID-19 outcome.

Ewan Carr*[1] (0000-0002-1146-4922), Rebecca Bendayan*[1,2], Daniel Bean[1,3], Matt Stammers[4,5], Wenjuan Wang[6], Huayu Zhang[7], Thomas Searle[1,2], Zeljko Kraljevic[1], Anthony Shek[8], Hang T T Phan[4,5], Walter Muruet[6], Anthony J Shinton[9], Ting Shi[7] , Xin Zhang[10] , Andrew Pickles[1,2], Daniel Stahl[1], Rosita Zakeri[11,12], Kevin O'Gallagher[11,12], Amos Folarin[1,3,13,14], Lukasz Roguski[3,13,14], Florina Borca[4,5,9], James Batchelor[5], Xiaodong Wu[15,16], Jiaxing Sun[15], Ashwin Pinto[9], Bruce Guthrie[7], Cormac Breen[6], Abdel Douiri[6], Honghan Wu[3,13], Vasa Curcin[6], James T Teo[8,11#], Ajay M Shah[9,11#], Richard JB Dobson[1,2,3,13,14#]

* Joint First Author

\# Joint Last Author

Corresponding author: Dr Ewan Carr, Institute of Psychiatry, Psychology & Neuroscience (IoPPN), 16 De Crespigny Park, London, SE5 8AF, +44 (0)20 7848 0304.
Email: ewan.carr@kcl.ac.uk; Telephone: +44 (0)20 7848 0304

[1] Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, U.K.

[2] NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, U.K.

[3] Health Data Research UK London, University College London, London, U.K.

[4] Clinical Informatics Research Unit, University of Southampton, Coxford Rd, Southampton, U.K.

[5] NIHR Biomedical Research Centre at University Hospital Southampton NHS Foundation Trust, Coxford Rd, Southampton, U.K.

[6] School of Population Health and Environmental Sciences, King's College London, London, U.K.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[7] Usher Institute, University of Edinburgh, Edinburgh, U.K.

[8] Department of Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, U.K.

[9] UHS Digital, University Hospital Southampton, Tremona Road, Southampton, U.K.

[10] Department of Pulmonary and Critical Care Medicine, People's Liberation Army Joint Logistic Support Force 920th Hospital, Yunnan, China.

[11] King's College Hospital NHS Foundation Trust, London, U.K.

[12] School of Cardiovascular Medicine & Sciences, King's College London British Heart Foundation Centre of Excellence, London, U.K.

[13] Institute of Health Informatics, University College London, London, U.K.

[14] NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, U.K.

[15] Department of Pulmonary and Critical Care Medicine, Shanghai East Hospital, Tongji University, Shanghai, China.

[16] Department of Pulmonary and Critical Care Medicine, Taikang Tongji Hospital, Wuhan, China.

Word count: 3525 words.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Objectives To evaluate the National Early Warning Score (NEWS2), currently recommended in the UK for risk-stratification of severe COVID-19 outcomes, and subsequently identify and validate a minimal set of common parameters taken at hospital admission that improve the score.

Design Retrospective observational cohort with internal and multi-hospital external validation.

Setting Secondary care.

Interventions Not applicable.

Participants Training and temporal external validation cohorts comprised 1464 patients admitted to King's College Hospital NHS Foundation Trust (KCH) with COVID-19 disease from 1st March to 30th April 2020. External validation cohorts included 3869 patients from two UK NHS Trusts (Guys and St Thomas' Hospitals, GSTT and University Hospitals Southampton, UHS) and two hospitals in Wuhan, China (Wuhan Sixth Hospital and Taikang Tongji Hospital).

Main outcome measures The primary outcome was patient status at 14 days after symptom onset categorised as severe disease (transferred to intensive care unit or death). Age, physiological measures, blood biomarkers, sex, ethnicity and comorbidities (hypertension, diabetes, cardiovascular, respiratory and kidney diseases) were included.

Results NEWS2 score on admission was a weak predictor of severe COVID-19 infection (AUC = 0.628). Adding age and common blood tests (CRP, neutrophil count, estimated GFR and albumin) provided substantial improvements to a risk stratification model, particularly in relation to sensitivity, but performance was only moderate (AUC = 0.753). Improvement over NEWS2 remained robust and generalisable in GSTT (AUC = 0.817), UHS (AUC = 0.835) and Wuhan hospitals (AUC = 0.918).

Conclusions Adding age and a minimal set of blood parameters to NEWS2 improves the detection of patients likely to develop severe COVID-19 outcomes. This finding was replicated across NHS and non-UK hospitals. Adding a few common parameters to a pre-existing acuity score allows rapid and easy implementation of this risk-scoring system.

Keywords: NEWS2 score, Blood parameters, COVID-19, prediction model.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

KEY MESSAGES

- The National Early Warning Score (NEWS2), currently recommended for severe COVID-19 disease in the UK shows overall poor discrimination for severe outcomes (transfer to ICU or death). It can be improved by the addition of a small number of blood and physiological parameters routinely measured at hospital admission.
- The addition of age and a minimal set of common blood tests (C-reactive protein, neutrophil count, estimated GFR and albumin) provided substantial improvements in a risk stratification model.
- Although predictive performance varied from hospital to hospital, the improvement over NEWS2 alone was consistent across different patient cohorts.
- The proposed addition of a limited number of dichotomised parameters is easily derived from a pre-existing acuity score would be substantially easier to implement in a short-time scale compared to novel high-dimensional risk-scoring systems.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

As of 7th June 2020, there have been >6.7 million confirmed cases of COVID-19 disease worldwide[1]. While approximately 80% of infected individuals have mild or no symptoms[2], some develop severe COVID-19 disease requiring hospital admission. Within the subset of those requiring hospitalisation, early identification of those who deteriorate and require transfer to an intensive care unit (ICU) for organ support or may die is vital.

Currently available risk scores for deterioration of acutely-ill patients include (1) widely-used generic ward-based risk indices such as the National Early Warning Score (NEWS2)[3] or modified sequential organ failure assessment (mSOFA)[4]; and (2) the pneumonia-specific risk index, CURB-65[5] which usefully captures a combination of physiological observations with limited blood markers and comorbidities.

NEWS2 is a summary score of six physiological parameters or 'vital signs' (respiratory rate, oxygen saturation, systolic blood pressure, heart rate, level of consciousness, temperature and supplemental oxygen dependency), used to identify patients at risk of early clinical deterioration in the United Kingdom (UK) NHS hospitals[6,7] and primary care. The physiological parameters assessed in the NEWS2 score – particularly patient temperature, oxygen saturations and the supplemental oxygen dependency – have been associated with COVID-19 outcomes[2]; however, little is known about their predictive value for the severity of COVID-19 disease[8]. Additionally, a number of COVID-19-specific risk indices are being developed[9–11] as well as unvalidated online calculators[12] but generalisability is unknown[11]. A Chinese study has suggested a modified version of NEWS2 with addition of age only[13] but without any data on performance. With near universal usage of NEWS2 in UK NHS Trusts since March 2019[14], minor adaptation to NEWS2 is relatively easy to implement.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

As the SARS-Cov2 pandemic has progressed, evidence has emerged regarding potentially useful blood biomarkers[1,15–18]. Although most of these early reports contain data from small numbers of patients, several markers have been found to be associated with severity. These include neutrophilia and lymphopenia, particularly in older adults[10,17,19,20], neutrophil-to-lymphocyte ratio[21], C-reactive Protein (CRP) and lymphocyte-to-CRP ratio[21], markers of liver and cardiac injury such as alanine aminotransferase (ALT), aspartate aminotransferase (AST) and cardiac troponin[22] and elevated D-dimers, ferritin and fibrinogen[2,5,7].

Our aim was to evaluate the NEWS2 score and then identify a minimal combination of clinical and blood biomarkers routinely measured in hospitals to supplement the acuity score to allow medium-term stratification for risk of a severe disease outcome at 14 days from symptom onset. Our specific objectives were:

1. To explore independent associations of routinely measured physiological and blood parameters (including NEWS2 parameters) at or near hospital admission with disease severity (ICU admission or death at 14 days since symptom onset), adjusting for demographics and comorbidities;

2. To examine which minimal combination of these potential determinants of disease severity (physiological and blood parameters, sociodemographics and comorbidities) are the best predictors of disease severity; and

3. To compare the predictive value of the resulting model with a model based on the NEWS2 total score alone.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

METHODS

Study cohorts

The KCH training cohort (n=439) was defined as all adult inpatients testing positive for SARS-Cov2 by reverse transcription polymerase chain reaction (RT-PCR) between 1st March to 31st March 2020 at two acute hospitals (King's College Hospital and Princess Royal University Hospital) in South East London (UK) of Kings College Hospital NHS Foundation Trust (KCH). All patients included in the study had symptoms consistent with COVID-19 (e.g. cough, fever, dyspnoea, myalgia, delirium, diarrhoea). For validation purposes, the following cohorts were used:

1) External temporal hold-out at KCH of 1025 cases (1st April to 30th April 2020)
2) External validation cohort at Guys and St Thomas' Hospital NHS Foundation Trust (GSTT) of 1417 cases (3rd March 2020 to 21st May 2020)
3) External validation cohort at University Hospitals Southampton NHS Foundation Trust (UHS) of 564 cases (7th March to 20th May 2020)
4) External validation cohort at Wuhan Sixth Hospital and Taikang Tongji Hospital totalling 1888 cases (4th February to 30th March 2020)

The KCH component of the project operated under London South East Research Ethics Committee (reference 18/LO/2048) approval granted to the King's Electronic Records Research Interface (KERRI); specific work on COVID-19 research was reviewed with expert patient input on a virtual committee with Caldicott Guardian oversight. The UHS validation was performed as part of a service evaluation agreed with approval from trust research leads and the Caldicott Guardian. Ethical approval for GSTT was granted by The London Bromley Research Ethics Committee (reference 20/HRA/1871) to the King's Health Partners Data Analytics and Modelling COVID-19 Group to collect clinically relevant data points from patient's electronic

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

health records. The Wuhan validation was approved by the Research Ethics Committee of Shanghai Dongfang Hospital and Taikang Tongji Hospital.

Data Processing

King's College Hospital

Data was extracted from the structured and unstructured components of the electronic health record (EHR) using natural language processing (NLP) tools belonging to the CogStack ecosystem[23], namely MedCAT[24] and MedCATTrainer[25]. The CogStack NLP pipeline captures negation, synonyms, and acronyms for medical SNOMED-CT concepts as well as surrounding linguistic context using deep learning and long short-term memory networks. MedCAT produces unsupervised annotations for all SNOMED-CT concepts (Supplementary Table 1) under parent terms Clinical Finding, Disorder, Organism, and Event with disambiguation, pre-trained on MIMIC-III[26].

Starting from our previous model[27], further supervised training improved detection of annotations and meta-annotations such as experiencer (is the concept annotated experienced by the patient or other), negation (is the concept annotated negated or not) and temporality (is the concept annotated in the past or present) with MedCATTrainer. Meta-annotations for hypothetical, historical and experiencer were merged into "Irrelevant" allowing us to exclude any mentions of a concept that do not directly relate to the patient currently. Performance of the NLP pipeline for comorbidities mentioned in the text was evaluated on 4343 annotations in 146 clinical documents by a clinician (JT). F1 scores, precision, and recall are presented in Supplementary Table 2.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Guy's and St Thomas NHS Foundation Trust (GSTT)

Electronic health records from all patients admitted to Guy's and St Thomas NHS Foundation Trust who had a positive COVID-19 test result between the 3rd of March and 21st of May 2020, inclusive, were identified. Data were extracted using structured queries from six complementary platforms and linked using unique patient identifiers. Data processing was performed using Python 3.7. The process and outputs were reviewed by a study clinician.

University Hospitals Southampton (UHS)

Data were extracted from the structured components of the UHS CHARTS EHR system and data warehouse. Data was transformed to the required format for validation purposes using Python 3.7. Diagnosis and comorbidity data of interest were gathered from ICD-10 coded data. No unstructured data extraction was required for validation purposes. The process and outputs were reviewed by an experienced clinician prior to analysis.

Wuhan cohort

Demographic, premorbid conditions, clinical symptoms or signs at presentation, laboratory data, treatment and outcome data were extracted from electronic medical records using a standardised data collection form by a team of experienced respiratory clinicians, with double data checking and involvement of a third reviewer where there was disagreement. Anonymised data was entered into a password-protected computerised database.

Measures

Outcome. For all sites, the primary outcome was patient status at 14 days after the index date, categorised as transfer to ICU/death (WHO-COVID-19 Outcomes Scales 6-8) vs. not ICU/death (Scales 3-5). For KCH, the index date was symptom onset (or hospital admission where symptom onset was missing) and dates of symptom onset, ICU transfer, and death were ascertained and verified manually by a clinician. For GSTT and UHS, the index date was the

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

date of diagnosis (positive SARS-CoV-2 RT-PCR) minus 4 days because symptom onset was not available. For the Wuhan cohort, the index date was the date of admission minus 4 days.

Blood parameters. We focused on biomarkers that were routinely obtained at or shortly after admission and were therefore available for the vast majority of patients. These comprised: albumin (g/L), alanine aminotransferase (ALT; IU/L), creatinine (µmol/L), C-reactive protein (CRP; mg/L), estimated Glomerular Filtration Rate (eGFR; mL/min), Haemoglobin (g/L), lymphocyte count (x 10^9/L), neutrophil count (x 10^9/L), and platelet count (PLT; x| 10^9/L). We also derived the neutrophil-to-lymphocyte ratio (NLR) and the lymphocyte-to-CRP ratio[21]. Troponin-T (ng/L) and Ferritin (ug/L) were included, although these measures were only available for a subset of participants. D-dimers and HbA1c were excluded since they were measured in very few patients at admission and insufficient samples were available for analysis.

Physiological parameters. We included the NEWS2 total score as well as the following parameters measured separately: respiratory rate (breaths per minute), oxygen saturation (%), systolic blood pressure (mmHg), heart rate (beats/min), temperature (°C), and consciousness (Glasgow Coma Scale; GCS). All were measured at or shortly after admission. Diastolic blood pressure was also included.

Demographics and comorbidities. Age, sex, ethnicity and comorbidities were considered. Self-defined ethnicity was categorised as caucasian vs. BAME (Black, Asian and minority ethnic) and patients with ethnicity reported as 'unknown/mixed/other' were excluded. Binary variables were derived for comorbidities: hypertension, diabetes, heart disease (heart failure and ischemic heart disease), respiratory disease (asthma and chronic obstructive pulmonary disease, COPD) and chronic kidney disease.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Statistical analyses

All parameters were scaled (mean 0; standard deviation = 1) to facilitate interpretability and logarithmic or square-root transformations were applied to skewed parameters. Outlying high values for some blood parameters were retained after clinical examination to ascertain plausibility. To explore independent associations of physiological and blood parameters with 14-day death/ICU (Objective 1) we used logistic regression with Firth's bias reduction method[28]. Each parameter was tested independently, adjusted for age and sex (Model 1) and then additionally adjusted for comorbidities (Model 2). *P*-values were adjusted using the Benjamini-Hochberg procedure to keep the False Discovery Rate (FDR) at 5%[29]. These models were conducted with R 4.0[30] using the logistf[31] package.

To investigate which panel of parameters performed best in predicting the 14-day outcome over and above NEWS2 we used a regularized logistic regression with a LASSO (Least Absolute Shrinkage and Selection Operator) estimator which shrinks parameters according to their variance, reduces overfitting and enables automatic variable selection[32]. The optimal degree of regularization was determined by identifying a tuning parameter λ using cross-validation[33]. From an initial model with NEWS2 total score only, sets of features were added in order of (i) age and sex; (ii) blood/physiological parameters; and (iii) comorbidities. To estimate the predictive performance of our model on new cases of the same underlying population, we performed internal nested cross-validation (10 folds/20 repeats for inner loop; 10 folds/100 repeats for outer loop). Overall discrimination was assessed based on the area under the curve (AUC). To impute missing features we used k-Nearest Neighbours imputation (k=5). Scaling and kNN imputation were incorporated within the model development and selection process to avoid data leakage which would otherwise result in optimistic performance measures[34].

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The predictive performance of the derived regularized logistic regression model was then evaluated using temporal external validation[35] on a hold-out sample of 1025 patients admitted to KCH after the training sample (Supplementary Figure 1) as well as external validation samples from GSTT (n=1417), UHS (n=564), and Wuhan (n=1888). Validation used logistic regression models trained on the KCH training sample, with code and pre-trained models shared via GitHub[1]. All procedures were identical to those used during training, including scaling and imputation. Performance was assessed using AUC, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Model calibration was assessed using a calibration plot (model predicted probability vs. true probability). These models were estimated in Python 3.8[36] using Scikit-Learn[37].

For ease-of-use in a clinical setting, two further models were estimated: (1) A model (NEWS2 + CoV) that supplemented NEWS2 with five of the most important features chosen based on internal validation and data completeness; (2) A threshold model, derived by dichotomising each feature from the NEWS2 + CoV model. Thresholds were defined using a decision tree model tuned on the training sample.

To ensure robust results, we explored whether predictive performance could be improved using a more complex approach such as gradient boosted trees as implemented in the XGBoost library[38]. Procedures were identical to those described above, except the imputation step was not required. Additional sensitivity analyses were performed to account for potential demographic variability. Following recent studies suggesting sex differences in COVID outcome[17] we tested interactions between each physiological and blood parameter and sex using likelihood-ratio tests. We also repeated all models with adjustment for ethnicity in the subset of individuals with available data for ethnicity (n=276 in KCH training sample).

---

[1] https://github.com/ewancarr/NEWS2-COVID-19

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

RESULTS

The KCH training cohort comprised 439 inpatients testing positive for COVID-19 (from 1st March to 31st March 2020) of whom 155 (35%) were transferred to ICU or died (COVID-19 WHO Score 6-8) within 14 days of symptom onset. Table 1 describes the clinical characteristics of the cohort. Patients experiencing more severe outcomes were significantly older but there was no evidence of differences by sex or ethnicity. There were some differences between groups in the prevalence of comorbidities but these did not reach statistical significance after correction for multiple testing. At admission, patients who had transferred to ICU or died within 14 days had lower levels of Albumin and estimated GFR; and elevated levels of CRP, creatinine, ferritin, neutrophils, and neutrophil/lymphocyte ratio. Median NEWS2 total scores were significantly higher in patients who transferred to ICU or died (3 vs 2; $p < 0.001$), compared to patients experiencing less severe outcomes. Characteristics of validation cohorts from GSTT, UHS, and Wuhan are described in Supplementary Table 3. There were fewer BAME minorities in UHS (compared to KCH) and a much lower proportion of severe cases in Wuhan hospitals (3% vs. 25-35% in UK sites).

Table 1: Patient characteristics of the training sample at admission to King's College Hospitals (n=439)

	N avail.	**All patients**	**Patients by status at 14-day endpoint**		*P*-value for test between outcome groups[1]
			No ICU/death WHO-COVID-19 Scales 3-5	ICU/death WHO-COVID-19 Scales 6-8	
Days from symptom onset to ICU/death (Median [IQR])	155	-	-	6 [4.5]	-
Age (Median [IQR])	439	69.00 [28.00]	67.00 [30.00]	72.00 [27.00]	0.010
Sex (male) (N (%))	439	242 (55.1%)	152 (53.5%)	90 (58.1%)	0.560
BAME (N (%))	276	115 (41.7%)	70 (40.7%)	45 (43.3%)	0.867
Comorbidities	N avail.		N (%)		
Hypertension	439	239 (54.4%)	146 (51.4%)	93 (60.0%)	0.211

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Diabetes mellitus	439	148 (33.7%)	90 (31.7%)	58 (37.4%)	0.434
Heart Failure	439	51 (11.6%)	33 (11.6%)	18 (11.6%)	0.999
Ischaemic Heart Diseases	439	62 (14.1%)	42 (14.8%)	20 (12.9%)	0.823
COPD	439	46 (10.5%)	26 (9.2%)	20 (12.9%)	0.434
Asthma	439	66 (15.0%)	44 (15.5%)	22 (14.2%)	0.879
Chronic Kidney Disease	439	83 (18.9%)	45 (15.8%)	38 (24.5%)	0.087
Blood biomarkers	N avail.		Median [IQR]		
Albumin	399	37.00 [7.00]	38.00 [6.50]	36.00 [7.00]	<0.001
Alanine aminotransferase (ALT)	183	31.00 [33.00]	29.00 [31.25]	33.00 [34.00]	0.434
C-reactive protein (CRP)	412	69.05 [108.30]	51.80 [84.90]	109.50 [134.45]	<0.001
Creatinine	414	88.00 [48.00]	84.00 [41.25]	99.00 [72.00]	<0.001
Estimated GFR	414	67.50 [41.00]	71.00 [37.00]	59.00 [46.50]	<0.001
Ferritin	122	942.00 [1165.25]	642.50 [1149.75]	1131.00 [902.50]	0.018
Haemoglobin	415	127.00 [30.00]	127.00 [29.00]	126.00 [29.00]	0.683
Lymphocyte count	415	1.00 [0.69]	1.00 [0.70]	0.93 [0.66]	0.683
Neutrophil count	413	4.88 [3.59]	4.53 [3.00]	5.87 [5.21]	<0.001
Neutrophil/lymphocyte ratio	413	4.86 [4.97]	4.50 [4.25]	5.93 [6.22]	<0.001
Lymphocyte/CRP ratio	411	0.01 [0.03]	0.02 [0.04]	0.01 [0.02]	<0.001
Platelet count	415	211.00 [102.00]	209.00 [98.25]	219.00 [106.50]	0.867
Troponin T	138	18.00 [29.50]	16.00 [26.25]	20.00 [38.50]	0.384
Physiological parameters	N avail.		Median [IQR]		
NEWS2 Total Score	400	2.00 [3.00]	2.00 [3.00]	3.00 [4.00]	<0.001
Heart rate	404	85.00 [20.00]	85.00 [19.00]	86.00 [23.00]	0.384
Oxygen saturation	403	96.00 [3.00]	97.00 [2.75]	96.00 [4.00]	0.018
Respiration rate	404	19.00 [2.00]	18.00 [2.00]	19.00 [3.00]	0.019
GCS score	172	15.00 [1.00]	15.00 [1.00]	15.00 [1.00]	0.126
Systolic blood pressure	404	126.00 [29.00]	126.00 [26.50]	125.00 [32.00]	0.881
Diastolic blood pressure	404	72.00 [18.25]	74.00 [17.50]	70.00 [20.00]	0.211
Temperature	404	36.90 [0.90]	37.00 [0.90]	36.90 [1.00]	0.560

Notes. [1]Wilcoxon test for continuous variables; χ^2 test for binary variables. FDR-corrected *P*-values based on the Benjamini-Hochberg correction.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Logistic regression models were used to assess independent associations between each physiological and blood parameter and disease severity (Supplementary Table 4). Transfer to ICU/death was associated with higher NEWS2 total score, CRP, heart rate, neutrophil count, neutrophil/lymphocyte ratio, respiration rate; and lower lymphocyte/CRP ratios, Albumin, estimated GFR, creatinine, and oxygen saturation. These associations remained after adjustment for age, sex and comorbidities. Sensitivity analyses showed no evidence of differences by sex and findings were consistent when additionally adjusting for ethnicity in the subset of individuals with ethnicity data (Supplementary Table 5).

Supplementing NEWS2 with a minimal set of routinely collected blood and physiological parameters to improve NEWS2 to improve prediction of 14-day outcome

We included all predictors in a combined logistic regression model using LASSO regularisation. Internally validated predictive performance is presented in Table 2. The baseline model with only NEWS2 showed poor discrimination with an AUC of 0.628. Adding age, sex, and all blood and physiological parameters increased the AUC by 0.114, to 0.742 (+/- 2SD: 0.735, 0.750). Performance was not improved when comorbidities were included in this model. A final model (NEWS2 + CoV) was estimated including NEWS2 and the top five most important features. This simpler model resulted in a slightly larger AUC of 0.753 (+/- 2SD: 0.739, 0.768) which may indicate some overfitting due to the pre-selection of variables from previous analyses. Results were consistent when repeating these models in the subset of patients with information available on ethnicity (Supplementary Table 6).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2: KCH internally validated predictive performance (n=439) based on nested repeated cross-validation

	Included features	Internally validated AUC			Sensitivity	Specificity	PPV	NPV
		Mean	-2SD	+2SD				
1	NEWS2	0.628	0.623	0.633	0.184	0.948	0.658	0.680
2	NEWS2 + **DBP**	0.742	0.735	0.750	0.415	0.849	0.600	0.727
3	NEWS2 + DBP**C**	0.734	0.723	0.490	0.381	0.848	0.578	0.715
4	NEWS2 **+ CoV** NEWS2 + CRP + neutrophils + eGFR + Albumin + Age	0.753	0.739	0.768	0.433	0.843	0.601	0.731

D = demographic variables age, sex
C = comorbidities (8 features)
B = bloods (10 features)
P = physiological parameters (7 features)

Notes. AUC based on repeated, nested cross-validation (inner loop: 10 folds/20 repeats; outer loop = 10 folds/100 repeats). Missing values imputed at each outer loop with k-Nearest Neighbours (KNN) imputation.

Figure 1 summarises feature importances from LASSO logistic regression models fitted to the KCH training sample. When adding demographic, blood, and physiological parameters to NEWS2 (NEWS2 + DBP) eight features were retained, in order of effect size: NEWS2 total score, CRP, estimated GFR, neutrophils, albumin, age, Troponin T, and oxygen saturation. Notably, comorbid conditions were not retained when added in subsequent models (NEWS2 + DBPC), suggesting most of the variance was already captured by the top five parameters.

FIGURE 1 HERE

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

When these models were repeated using a XGBoost[38] the pattern of results was consistent with those from regularized logistic regression (Supplementary Table 7). The internally validated AUC improved from 0.647 (+/- 2SD: 0.631, 0.663) for a model with NEWS2 alone, to 0.731 (+/- 2SD: 0.706, 0.756) for the NEWS2 + CoV model. Importantly, while the pattern of results was consistent, a more complex machine learning estimator did not improve predictive performance.

Temporal external validation was conducted on a hold-out sample of 1025 KCH patients. This sample was similar to the training sample (Supplementary Table 3) except that the proportion who transferred to ICU or died was lower (35% vs 27%). For the hold-out sample, the AUC for NEWS2 alone was 0.665, and this improved to 0.761 for the simplified MEWS2 + CoV model comprising NEWS2 and the top five features (CRP, neutrophils, estimated GFR, albumin and age) (sensitivity = 0.624; specificity = 0.764) (Table 3). Calibration for these models (Supplementary Figure 2) was acceptable but suggested some overestimation of risk probabilities. Overall, results from the hold-out sample were consistent with those from internal validation.

Multi-site replication

When we replicated our trained models in GSTT (n=1417), UHS (n=564) and Wuhan (n=1888) we found our results to be robust despite changes in cohort characteristics (Table 3). In all sites, the supplemented NEWS2 model (NEWS2 + CoV) showed improved predictive performance over NEWS2 alone, with an average increase in AUC of 0.112. Furthermore, the threshold model showed similar performance across sites (Table 3). Details of the validation cohorts can be found on Supplementary Table 3. The code and pre-trained models used for replications is available at https://github.com/ewancarr/NEWS2-COVID-19 and can be tested in other datasets.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3: Multi-site replications (n=5333) 'THRESHOLDS' refers to the 'NEWS2-CoV' model with dichotomised variables, described above. KCH training performance is included for reference.

Sample	Model	AUC	Sensitivity	Specificity	PPV	NPV
KCH training (n=439)	NEWS2 only	0.628	0.180	0.950	0.664	0.681
	NEWS2 + CoV	0.751	0.415	0.842	0.589	0.727
KCH validation (n=1025)	NEWS2 only	0.665	0.294	0.899	0.523	0.773
	NEWS2 + CoV	0.761	0.624	0.764	0.497	0.844
	THRESHOLDS	0.713	0.556	0.788	0.495	0.826
UHS validation (n=564)	NEWS2 only	0.770	0.357	0.941	0.667	0.816
	NEWS2 + CoV	0.835	0.571	0.899	0.650	0.864
	THRESHOLDS	0.812	0.393	0.934	0.663	0.823
GSTT validation (n=1417)	NEWS2 only	0.692	0.145	0.981	0.726	0.768
	NEWS2 + CoV	0.817	0.510	0.903	0.646	0.841
	THRESHOLDS	0.749	0.526	0.864	0.573	0.840
Wuhan validation (n=1888)	NEWS2 only	0.755	0.385	0.917	0.116	0.981
	NEWS2 + CoV	0.918	0.904	0.796	0.112	0.997
	THRESHOLDS	0.882	0.212	0.985	0.282	0.978

DISCUSSION

Principal findings

To our knowledge, this study is the first to systematically evaluate the UK NEWS2 acuity score in COVID-19 and attempt to improve performance of NEWS2 in multiple NHS Trusts. We found that the NEWS2 score shows overall poor discrimination with high specificity but poor sensitivity for severe outcomes (transfer to ICU or death). However, its value for risk stratification (especially sensitivity) can be improved by adding age and a small number of

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

additional blood parameters (CRP, neutrophils, estimated GFR, and albumin). Several blood measures previously linked with severe outcomes – such as lymphocyte and ALT[13], or transformations of inflammatory markers such as CRP/lymphocyte or neutrophil/lymphocyte ratios – did not provide additional value. Pre-existing comorbidities did not improve risk prediction and were not retained in the final model. This was unexpected but may indicate that the effect of pre-existing health conditions is manifest through some of the included blood or physiological markers.

Comparison with other studies

A systematic review of 10 prediction models for mortality in COVID-19 infection[9] found broad similarities with the features retained in our models, particularly regarding CRP and neutrophil levels. However, existing prediction models suffer several methodological weaknesses including over-fitting, selection bias, and reliance on cross-sectional data without accounting for censoring. Additionally, many existing studies have relied on single centre studies or in ethnically homogenous Chinese cohorts, whereas this study shows validation in multiple organisations and diverse populations indicative of generalisability.

NEWS2 is a summary score derived from six physiological parameters, including oxygen supplementation (yes/no). Lack of evidence for NEWS2 use in COVID-19 especially in primary care has been highlighted[8]. The oxygen saturation component of physiological measurements added value beyond NEWS2 total score and was retained following regularisation (i.e. model NEWS2 + DBP). This suggests some residual association over and above what is captured by the NEWS2 score, and reinforces Royal College of Physicians guidance that the NEWS2 score ceilings with respect to respiratory function[39].

Cardiac disease and myocardial injury have been described in severe COVID-19 cases in China[1,22]. In our model, blood Troponin-T, a marker of myocardial injury, had additional

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

salient signal but was only measured in a subset of our cohort at admission, so it was excluded from our final model. This could be explored further in larger datasets.

Strengths and limitations

Our study provides a risk stratification model for which we obtained generalisable and robust results across national and non-UK sites (total patients=5333) with differing geographical catchment and population characteristics. However, some limitations should be acknowledged. First, there are likely to be other parameters not measured in this study that could substantially improve the risk stratification model (e.g. radiological features or comorbidity load). This could be addressed by future work, but these parameters were not considered in the present study to avoid limiting the real-world implementation of the risk stratification model. Second, we used a 14-day time window from symptom onset to provide a balance between medium-term prognostication and actionable risk stratification at the usual period of deterioration. Longer timeframes may be useful for prognostication but are likely to be harder to generalise due to the greater number of factors affecting outcomes (e.g. institutional, regional or national policies). Since NEWS2 score is optimised for very near-term deterioration at 24 hours[7], a 14-day window was used as a compromise. Thirdly, our models showed better performance in UK secondary care settings with populations with higher rates of severe disease after COVID-19 infection. Further research should be performed to investigate the suitability of our model for primary care settings with higher prevalence mild disease severities[8].

CONCLUSION

The simple addition of age and a limited number of blood parameters to the existing and widely implemented UK NEWS2 system can contribute to improved risk stratification among COVID-19 patients. This type of model can be easily implemented in clinical practice and predicted risk score probabilities of individual patients are easy to communicate. The additional parameters are

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

routinely collected on patients at hospital admission, and with near universal usage of NEWS2 in NHS Trusts since March 2019[14], a minor adaptation to NEWS2 is substantially easier to implement in a variety of health settings than a bespoke risk score.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ACKNOWLEDGMENTS

DMB is funded by a UKRI Innovation Fellowship as part of Health Data Research UK MR/S00310X/1 (https://www.hdruk.ac.uk).

RB is funded in part by grant MR/R016372/1 for the King's College London MRC Skills Development Fellowship programme funded by the UK Medical Research Council (MRC, https://mrc.ukri.org) and by grant IS-BRC-1215-20018 for the National Institute for Health Research (NIHR, https://www.nihr.ac.uk) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London.

RJBD is supported by: (1) NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, U.K. (2) Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust. (3) The BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA; it is chaired by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and ESC. (4) The National Institute for Health Research University College London Hospitals Biomedical Research Centre. (5) National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. (5) The UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare (6) the National Institute for Health Research (NIHR) Applied Research

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust.

KO'G is supported by an MRC Clinical Training Fellowship (MR/R017751/1).

WW is supported by a Health Foundation grant.

AD and VC acknowledge support from National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) South London at King's College Hospital NHS Foundation Trust and the Royal College of Physicians, as well as the support from the NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. VC is additionally supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust.

RZ is supported by a King's Prize Fellowship.

AS is supported by a King's Medical Research Trust studentship.

JTHT is supported by London AI Medical Imaging Centre for Value-Based Healthcare (AI4VBH) and the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

FB and PTTH are funded by National Institute for Health Research (NIHR) Biomedical Research Centre, Data Sciences at University Hospital Southampton NHS Foundation Trust and the Clinical Informatics Research Unit, University of Southampton.

JB is funded by the Clinical Informatics Research Unit, University of Southampton, and part funded by the Global Alliance for Chronic Disease (GDAC).

AP is part funded by UHS Digital, University Hospital Southampton, Tremona Road, Southampton.

AJS is supported by a Digital Health Fellowship through Health Education England (Wessex).

HW and HZ are supported by Medical Research Council and Health Data Research UK Grant (MR/S004149/1), Industrial Strategy Challenge Grant (MC_PC_18029) and Wellcome Institutional Translation Partnership Award (PIII054). XW is supported by National Natural Science Foundation of China (grant number 81700006).

AMS is supported by the British Heart Foundation (CH/1999001/11735), the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's & St Thomas' NHS Foundation Trust and King's College London (IS-BRC-1215-20006), and the Fondation Leducq. AP is partially supported by NIHR NF-SI-0617-10120. This work was supported by the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre (BRC) Clinical and Research Informatics Unit (CRIU), NIHR Health Informatics Collaborative (HIC), and by awards establishing the Institute of Health Informatics at University College London (UCL). This work was also supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and the Wellcome Trust.

This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centres at South London and Maudsley NHS Foundation Trust, London AI Medical Imaging Centre for Value-Based Healthcare, and Guy's & St Thomas' NHS Foundation Trust, both with King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would also like to thank all the clinicians managing the patients, the patient experts of the KERRI committee, Professor Irene Higginson, Professor Alastair Baker, Professor Jules Wendon, Dan Persson and Damian Lewsley for their support.

The authors acknowledge use of the research computing facility at King's College London, Rosalind (https://rosalind.kcl.ac.uk), which is delivered in partnership with the National Institute for Health Research (NIHR) Biomedical Research Centres at South London & Maudsley and Guy's & St. Thomas' NHS Foundation Trusts, and part-funded by capital equipment grants from the Maudsley Charity (award 980) and Guy's & St. Thomas' Charity (TR130505). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, King's College London, or the Department of Health and Social Care.

CONTRIBUTOR AND GUARANTOR INFORMATION

The corresponding author, Dr Ewan Carr, is guarantor of the manuscript.

JT, AS, RD, EC and RB conceived the study design and developed the study objectives. JT, RD, AF, LR, DB, ZK, TS and AS were the leads to develop CogStack platform. DB, ZK, TS, AS

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

were responsible for the data extraction and preparation. EC, RB, AP, DS contributed to the statistical analyses. All authors contributed to the interpretation of the data. AS, JT, KO, RZ provided clinical input. All authors contributed to interpret the data, draft the article and provided final approval of the manuscript. DMB, ZK, AS, TS, JTHT, LR, KN performed data processing and software development; KOG, RZ, JTHT performed data validation.

At GSTT, WW and WM were responsible for the data extraction and preparation. WW performed the model validation. AD and VC contributed to the interpretation of the data.

At UHS, MS and FB were responsible for the data extraction and preparation. MS, HP and AS contributed to the statistical analysis. All authors contributed to the interpretation of the data. MS and AP provided clinical input. MS and HP performed data/model validation.

For the Wuhan cohort, XZ, XW and JS extracted the data from the EHR system. HW and HZ preprocessed the raw data and conducted the prediction model validations. BG, HW, HZ, TS and JS interpreted the data and results.

Disclaimer: The views expressed are those of the authors and not necessarily those of the MRC, NHS, the NIHR or the Department of Health and Social Care. The funders of the study had no role in the study design, data collection, data analysis, data interpretation, writing of the report or the decision to submit the article for publication.

COPYRIGHT/LICENSE FOR PUBLICATION

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or,

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

DATA SHARING

Code and pre-trained models are available at https://github.com/ewancarr/NEWS2-COVID-19 and openly shared for testing in other COVID datasets.

Source text from patient records used at all sites in the study will not be available due to inability to safely fully anonymise up to the Information Commissioner Office (ICO) standards and would be likely to contain strong identifiers (e.g. names, postcodes) and highly sensitive data (e.g. diagnoses).

A subset of the KCH dataset limited to anonymisable information (e.g. only SNOMED codes and aggregated demographics) is available on request to researchers with suitable training in information governance and human confidentiality protocols subject to approval by the King's College Hospital Information Governance committee; applications for research access should be sent to kch-tr.cogstackrequests@nhs.net. This dataset cannot be released publicly due to the risk of re-identification of such granular individual level data, as determined by the King's College Hospital Caldicott Guardian.

The GSTT dataset cannot be released publicly due to the risk of re-identification of such granular individual level data, as determined by the Guy's and St. Thomas's Trust Caldicott Guardian.

The UHS dataset cannot be released publicly due to the risk of re-identification of such granular individual level data, as determined by the University Hospital Southampton Caldicott Guardian.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The Wuhan dataset used in the study will not be available due to inability to fully anonymise in line with ethical requirements. Applications for research access should be sent to TS and details will be made available via https://covid.datahelps.life/prediction/.

PUBLIC AND PATIENT INVOLVEMENT STATEMENT AND ETHICS STATEMENT

This project operated under London South East Research Ethics Committee approval (reference 18/LO/2048) granted to the King's Electronic Records Research Interface (KERRI). Given the current context, specific work on COVID19 research was reviewed with a virtually convened 4-expert patient panel input on a virtual committee with Caldicott Guardian oversight. In addition, the KERRI database used in this study was developed according to KCH use of data for research guidelines (https://www.kch.nhs.uk/research/use-of-data-for-research) and patient engagement groups for this dataset have been conducted in 2018-2019.

DISSEMINATION DECLARATION

Dissemination to study individual participants is not possible to protect anonymity.
Dissemination to patient organisations and stakeholder groups is planned.

TRANSPARENCY DECLARATION

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

ROLE OF THE FUNDING SOURCE

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

COMPETING INTERESTS

JTHT received research support and funding from InnovateUK, Bristol-Myers-Squibb, iRhythm Technologies, and holds shares <£5,000 in Glaxo Smithkline and Biogen.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1 Zhou F, Yu T, Du R, *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* 2020;**395**:1054–62. doi:10.1016/S0140-6736(20)30566-3

2 WHO. WHO COVID-19 Dashboard. 2020. https://who.sprinklr.com/ (accessed 8 June 2020).

3 Scott LJ, Redmond NM, Tavaré A, *et al.* Association between National Early Warning Scores in primary care and clinical outcomes: an observational study in UK primary and secondary care. *Br J Gen Pract* Published Online First: 7 April 2020. doi:10.3399/bjgp20X709337

4 Lambden S, Laterre PF, Levy MM, *et al.* The SOFA score—development, utility and challenges of accurate assessment in clinical trials. *Crit Care* 2019;**23**:374. doi:10.1186/s13054-019-2663-7

5 Lim WS, Eerden MM van der, Laing R, *et al.* Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003;**58**:377–82. doi:10.1136/thorax.58.5.377

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

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

7 Smith GB, Prytherch DR, Meredith P, *et al.* The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* 2013;**84**:465–70. doi:10.1016/j.resuscitation.2012.12.016

8 Greenhalgh T., Treadwell J, *et al.* NEWS (or NEWS2) score when assessing possible COVID-19 patients in primary care? *Oxford COVID-19 Evidence Service* doi: 10.13140/RG.2.2.26433.10089

9 Ji D, Zhang D, Xu J, *et al.* Prediction for Progression Risk in Patients with COVID-19 Pneumonia: the CALL Score. *Clin Infect Dis* doi:10.1093/cid/ciaa414

10 Shi Y, Yu X, Zhao H, *et al.* Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. *Crit Care* 2020;**24**:108. doi:10.1186/s13054-020-2833-7

11 Wynants L, Calster BV, Bonten MMJ, *et al.* Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020;**369**. doi:10.1136/bmj.m1328

12 COVIDAnalytics. https://www.covidanalytics.io/calculator (accessed 21 Apr 2020).

13 Liao X, Wang B, Kang Y. Novel coronavirus infection during the 2019–2020 epidemic: preparing intensive care units—the experience in Sichuan Province, China. *Intensive Care Med* 2020;**46**:357–60. doi:10.1007/s00134-020-05954-2

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

14 NHS England » National Early Warning Score (NEWS). https://www.england.nhs.uk/ourwork/clinical-policy/sepsis/nationalearlywarningscore/ (accessed 23 Apr 2020).

15 Huang C, Wang Y, Li X, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020;**395**:497–506. doi:10.1016/S0140-6736(20)30183-5

16 Li K, Wu J, Wu F, *et al.* The Clinical and Chest CT Features Associated with Severe and Critical COVID-19 Pneumonia. *Invest Radiol* 2020. doi:10.1097/RLI.0000000000000672

17 Xie J, Tong Z, Guan X, *et al.* Clinical Characteristics of Patients Who Died of Coronavirus Disease 2019 in China. *JAMA Netw Open* 2020;**3**:e205619–e205619. doi:10.1001/jamanetworkopen.2020.5619

18 Zhang J-J, Dong X, Cao Y-Y, *et al.* Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy* Published Online First: 19 February 2020. doi:10.1111/all.14238

19 Ruan Q, Yang K, Wang W, *et al.* Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med* Published Online First: 3 March 2020. doi:10.1007/s00134-020-05991-x

20 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* Published Online First: 28 February 2020. doi:10.1056/NEJMoa2002032

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

21 Lagunas-Rangel FA. Neutrophil-to-lymphocyte ratio and lymphocyte-to-C-reactive protein ratio in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis. *J Med Virol* Published Online First: 3 April 2020. doi:10.1002/jmv.25819

22 Guo T, Fan Y, Chen M, *et al.* Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). *JAMA Cardiol* Published Online First: 27 March 2020. doi:10.1001/jamacardio.2020.1017

23 Jackson R, Kartoglu I, Stringer C, *et al.* CogStack - experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital. *BMC Med Inform Decis Mak* 2018;**18**. doi:10.1186/s12911-018-0623-9

24 Kraljevic Z, Bean D, Mascio A, *et al.* MedCAT -- Medical Concept Annotation Tool. *ArXiv191210166 Cs Stat* Published Online First: 18 December 2019.http://arxiv.org/abs/1912.10166 (accessed 17 Apr 2020).

25 Searle T, Kraljevic Z, Bendayan R, *et al.* MedCATTrainer: A Biomedical Free Text Annotation Interface with Active Learning and Research Use Case Specific Customisation. *ArXiv190707322 Cs* Published Online First: 16 July 2019.http://arxiv.org/abs/1907.07322 (accessed 17 Apr 2020).

26 Johnson AEW, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;**3**:1–9. doi:10.1038/sdata.2016.35

27 Bean D, Kraljevic Z, Searle T, *et al.* Treatment with ACE-inhibitors is associated with less severe disease with SARS-Covid-19 infection in a multi-site UK acute Hospital Trust. *medRxiv* 2020;:2020.04.07.20056788. doi:10.1101/2020.04.07.20056788

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

28 Firth D. Bias Reduction of Maximum Likelihood Estimates. *Biometrika* 1993;**80**:27–38. doi:10.2307/2336755

29 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach. *J R Stat Soc Ser B-Methodol* 1995;**57**:289–300.

30 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: 2020. https://www.R-project.org

31 Heinze G, Ploner M. *logistf: Firth's bias-reduced logistic regression*. 2018. https://CRAN.R-project.org/package=logistf

32 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B Methodol* 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x

33 Hastie T, Tibshirani R, Wainwright M. *Statistical Learning with Sparsity: The Lasso and Generalizations*. New York: : CRC Press 2015.

34 Kuhn M, Johnson K. *Applied predictive modeling*. Springer 2013.

35 Steyerberg E. *Clinical prediction models*. Second Edition. Cham, Switzerland: Springer 2019.

36 Van Rossum G, Drake FL. *Python 3 reference manual*. Scotts Valley, CA: CreateSpace 2009.

37 Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;**12**:2825–30.

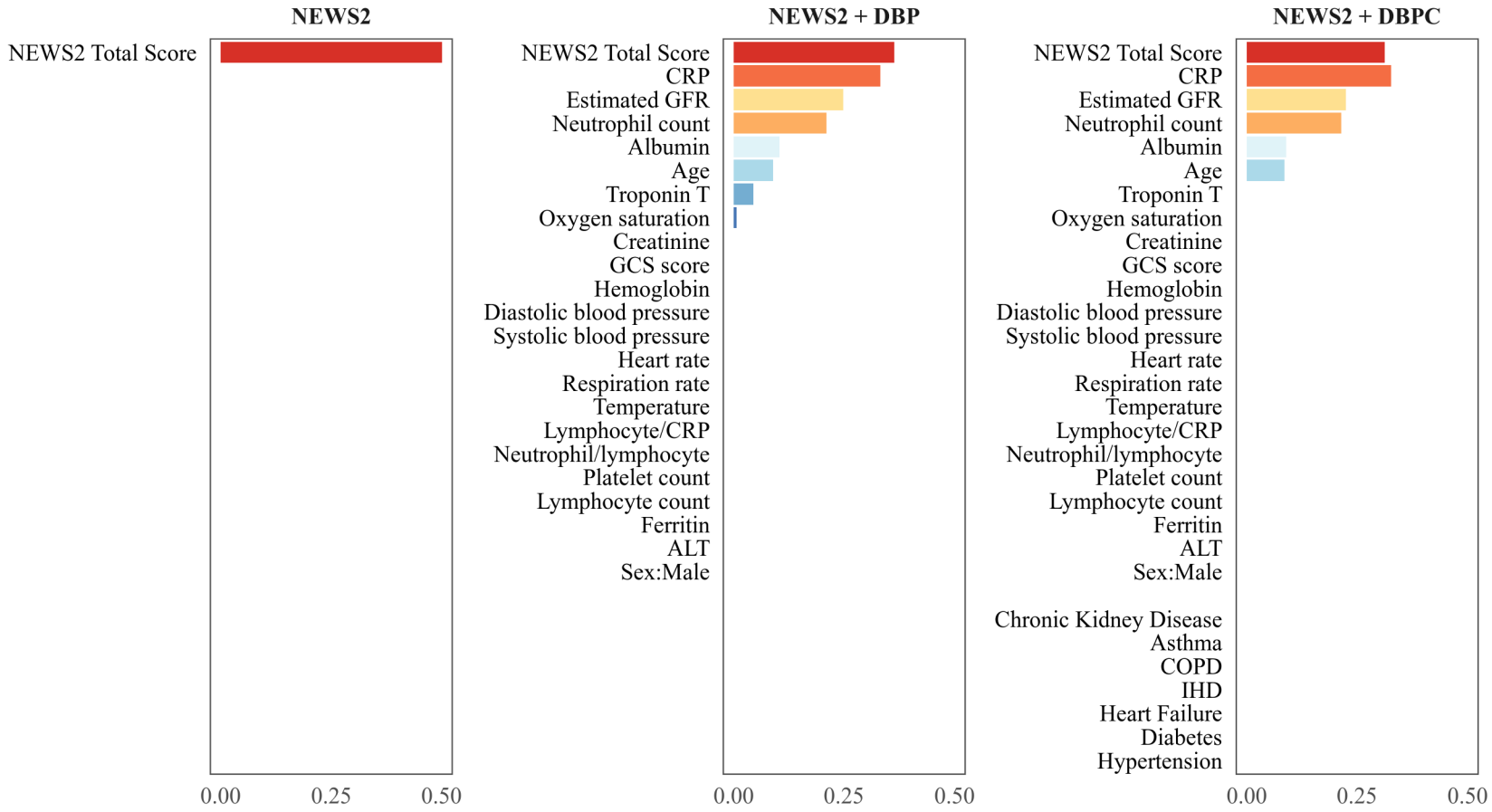
medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078006; this version posted June 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

38 Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. New York, NY, USA: : ACM 2016. 785–94. doi:10.1145/2939672.2939785

39 NEWS2 and deterioration in COVID-19. RCP London. 2020.https://www.rcplondon.ac.uk/news/news2-and-deterioration-covid-19 (accessed 24 Apr 2020).

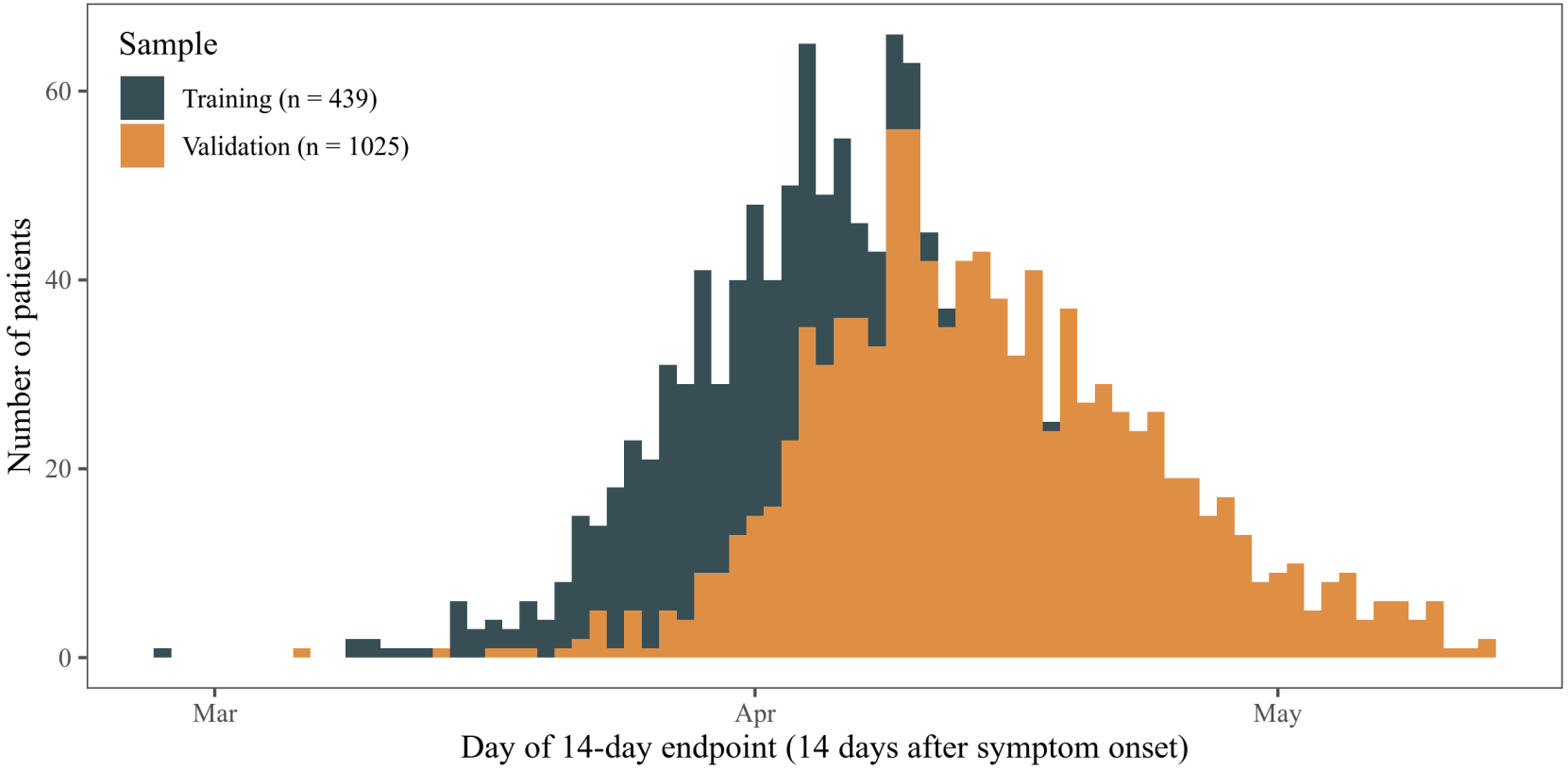
Figure 1: Feature importances from LASSO logistic regression in KCH training sample (n=439)

Notes. Feature importances refer to absolute values of standardised coefficients from logistic regression, sorted by effect size in model 'NEWS2 + DBP'. Where a feature is labelled on the y-axis, it was entered into the model. Features retained following LASSO regularisation are represented by a coloured bar; the absence of a bar indicates that this feature was omitted during regularisation.

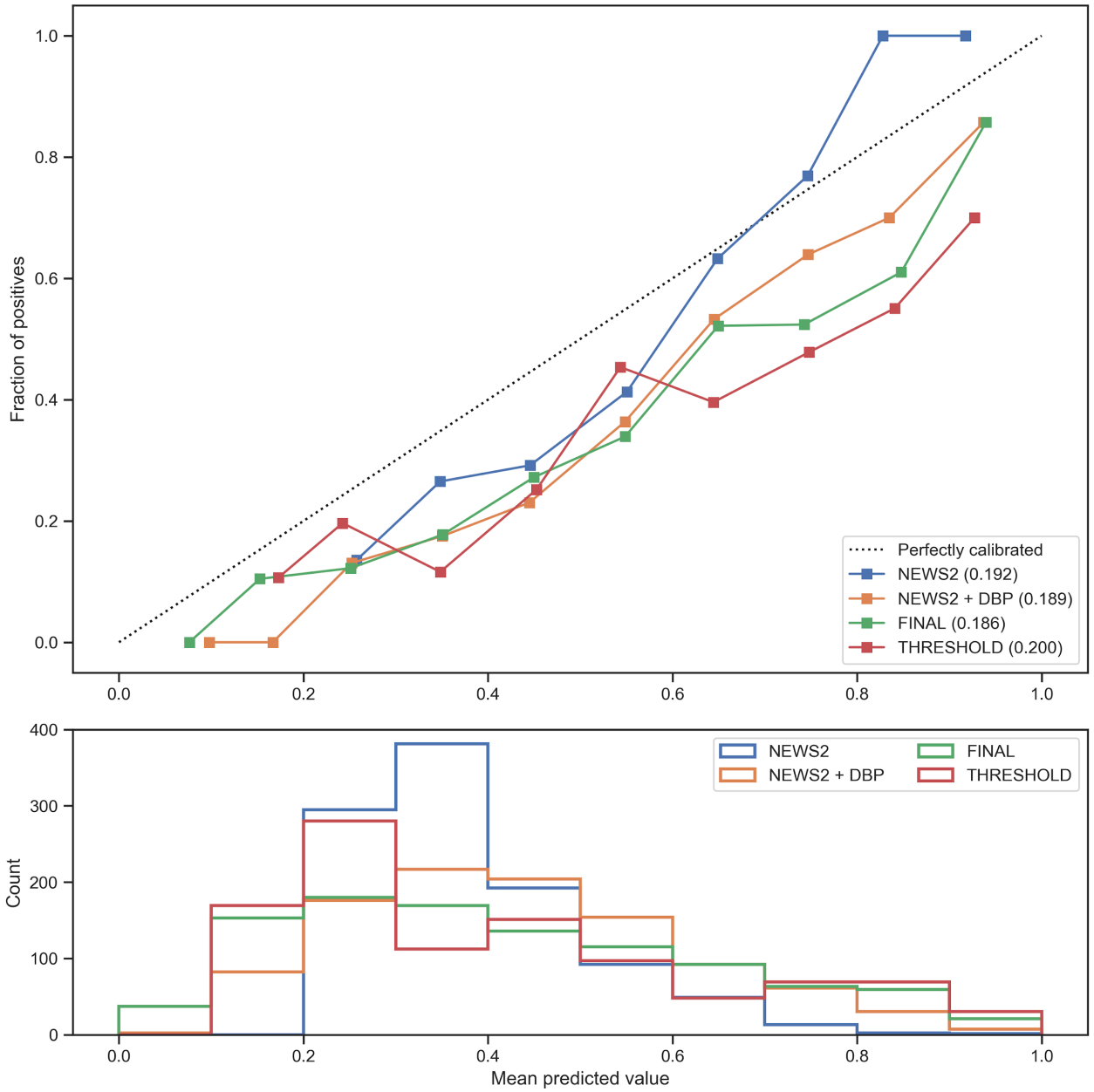


Supplementary Materials

Supplementary Figure 1: Timing of 14-day endpoints for training (n=439; positive test between 1st and 31st March) and hold-out (n=1025; positive test between 1st April and 30th April) samples for KCH.



Supplementary Figure 2: Calibration plot from temporal external validation (KCH; n=1025)



Supplementary Table 1: SNOMED terms

SNOMED concept name	SNOMED concept IDs
Diabetes	S-230572002, S-44054006, S-237599002, S-49455004
Heart Failure	S-42343007, S-426263006, S-48447003, S-418304008, S-10633002
IHD	S-401314000, S-194828000, S-233839009, S-414545008 S-394659003, S-1755008, S-413838009
Hypertension	S-59621000
COPD	S-13645005, S-313297008
Asthma	S-195967001
CKD	S-433144002, S-90688005, S-709044004

Supplementary Table 2: F1, precision and recall for NLP comorbidity detection

MedCATTrainer was used to collect manual annotations for 146 clinical documents totalling 4343 annotations. Each co-morbidity is defined using one or more SNOMED terms. Predicted true positive labels (TP), precision (P), recall (R), F1-score (F1) are shown for these aggregated concepts. These results only consider entity detection and not meta annotation.

	TP	F1	P	R	SNOMED terms
Diabetes mellitus	73	0.936	0.924	0.948	S-230572002, S-44054006, S-237599002, S-49455004
Heart Failure	11	0.893	0.786	1.000	S-42343007, S-426263006 S-48447003, S-418304008 S-10633002
IHD	23	0.979	0.958	1.000	S-401314000, S-194828000 S-233839009, S-414545008 S-394659003, S-1755008 S-413838009
Hypertension	84	0.883	0.988	0.778	S-59621000
COPD	14	0.967	0.933	1.000	S-13645005, S-313297008
Asthma	11	1.000	1.000	1.000	S-195967001
CKD	15	0.938	0.938	0.938	S-433144002, S-90688005, S-709044004

Supplementary Table 3: Comparison of training (n=439) and validation samples (n=5333)

	KCL training sample (n=439)		**KCL validation sample** (n=1025)		**UHS validation sample** (n=564)		**GSTT validation sample** (n=1417)		**Wuhan validation sample** (n=1888)	
14-day outcome[a]	N	N (%)	N	N (%)	N	N (%)	N	N (%)	N	N (%)
COVID-19 WHO Score 6-8 (ICU/death)	439	155 (35.3%)	1025	279 (27.2%)	564	140 (24.8%)	1417	365 (25.8%)	1888	52 (2.8%)
Demographics										
Age (median [IQR])	439	69.0 [28.0]	1025	71.5 [24.6]	564	68.9 [29.0]	1417	58.0 [29.0]	1888	60.0 [19.0]
Sex (male) (N (%))	439	242 (55.1%)	1025	602 (58.7%)	564	319 (56.6%)	1417	787 (55.5%)	1888	938 (49.7%)
BAME (N (%))	276	115 (41.7%)	817	327 (40.0%)	564	117 (20.7%)	1417	538 (38.0%)	-	-
Comorbidities						N (%)				
Hypertension	439	239 (54.4%)	1025	557 (54.3%)	564	214 (37.9%)	1417	471 (33.2%)	-	-
Diabetes	439	148 (33.7%)	1025	366 (35.7%)	564	63 (11.2%)	1417	319 (22.5%)	-	-
Heart Failure	439	51 (11.6%)	1025	76 (7.4%)	564	127 (22.5%)[c]	1417	70 (4.9%)	-	-
Ischaemic Heart Diseases	439	62 (14.1%)	1025	150 (14.6%)			1417	289 (20.4%)	-	-
COPD	439	46 (10.5%)	1025	112 (10.9%)	564	104 (18.4%)[d]	1417	78 (5.5%)	-	-
Asthma	439	66 (15.0%)	1025	129 (12.6%)	564	103 (18.3%)	1417	79 (5.6%)	-	-
Chronic Kidney Disease	439	83 (18.9%)	1025	175 (17.1%)	564	100 (17.7%)	1417	168 (11.9%)	-	-
Blood biomarkers						Median [IQR]				
Albumin	399	37.0 [7.0]	857	36.0 [7.0]	321	27.0 [7.0]	914	35.0 [9.0]	1888	38.7 [5.1]
ALT	183	31.0 [33.0]	521	32.0 [31.0]	116	43.0 [53.0][e]	824	27.0 [30.0]	-	-
CRP	412	69.0 [108.3]	889	82.0 [111.0]	383	109 [111.5]	1100	68.0 [118.0]	1888	1.8 [5.1]
Creatinine	414	88.0 [48.0]	883	93.0 [69.0]	390	75.0 [55.8]	1093	83.0 [52.0]	1888	64.2 [20.7]
Estimated GFR	414	67.5 [41.0]	895	64.0 [45.5]	386	88.5 [35.0]	1077	74.0 [53.0]	1888	103.2 [29.4]
Ferritin	122	942.0 [1165.2]	390	908.0 [1118.5]	218	544.5 [863.5]	691	750.0 [1207.5]	-	-

Haemoglobin	415	127.0 [30.0]	886	127.0 [29.0]	389	118.0 [31.0]	1127	122.0 [32.0]	-	-
Lymphocyte count	415	1.0 [0.7]	886	1.0 [0.6]	389	1.0 [0.7]	1127	0.9 [0.7]	-	-
Neutrophil count	413	4.9 [3.6]	887	5.7 [4.1]	382	5.6 [4.4]	1126	4.8 [4.2]	1888	3.4 [1.7]
Neutrophil/lymphocyte ratio	413	4.9 [5.0]	886	5.6 [6.1]	-	-	1126	5.2 [6.62]	-	-
Lymphocyte/CRP ratio	411	0.0 [0.0]	878	0.0 [0.0]	-	-	1089	0.0 [0.0]	-	-
Platelet count	415	211.0 [102.0]	886	214.5 [115.5]	389	249.0 [153.0]	1121	209.0 [121.0]	-	-
Troponin T	138	18.0 [29.5]	384	23.0 [49.2]	205	13.0 [36.0][f]	-	-	-	-
Physiological parameters						Median [IQR]				
NEWS2 total score	400	2.0 [3.0]	980	3.0 [3.0]	466	5.0 [4.0]	863	2.0 [3.0]	1888	1.0 [3.0]
Heart rate	404	85.0 [20.0]	984	84.0 [19.0]	-	-	878	86.0 [20.0]	1888	86.0 [18.0]
Oxygen saturation	403	96.0 [3.0]	984	96.0 [3.0]	-	-	822	96.0 [2.0]	-	-
Respiration rate	404	19.0 [2.0]	984	20.0 [4.0]	504	24.0 [9.0]	881	19.0 [3.0]	1888	20.0 [2.0]
GCS score	172	15.0 [1.0]	412	15.0 [1.0]	-	-	260	15.0 [0.0]	-	-
Systolic blood pressure	404	126.0 [29.0]	984	125.0 [27.0]	-	-	873	126.0 [27.0]	1888	130.0 [21.0]
Diastolic blood pressure	404	72.0 [18.2]	984	71.0 [17.0]	-	-	873	74.0 [16.0]	-	-
Temperature	404	36.9 [0.9]	984	36.8 [0.7]	-	-	873	37.0 [1.0]	1888	36.5 [0.3]

Notes. Some measures were unavailable for validation cohorts, but all sites measured NEWS2 plus the five supplementary measures (age, CRP, neutrophils, eGFR, albumin) and thus were able to evaluate the 'NEWS + CoV' and 'THRESHOLD' models.

[a] Outcome in all sites was ICU/death (WHO-COVID-19 Outcomes Scales 6-8) within 14 days of index date. Index date at KCH was symptom onset or hospital admission if symptom onset missing. For GSTT and UHS, index date was the date of diagnosis (positive SARS-CoV-2 RT-PCR) minus 4 days. For Wuhan cohort, index date was the date of admission minus 4 days.

[b] Wilcoxon test for continuous variables; X^2 test for binary variables. FDR-adjusted.

[c] UHS used combined comorbidity of 'Chronic Heart Disease'.

[d] UHS used comorbidity of 'Chronic Respiratory Disease'

[e] UHS used aspartate aminotransferase (AST) instead of ALT.

[f] UHS used High Sensitivity Troponin I, rather than Troponin T measured at other sites.

Supplementary Table 4: Logistic regression models for each blood and physiological measure tested separately in the KCH training sample

	N	**Model 1: Age, sex only**[1]		**Model 2: + comorbidities**[2]	
		Odds Ratio [95% C.I.]	*P*-value[2]	Odds Ratio [95% C.I.]	*P*-value[3]
CRP	412	2.02 [1.62, 2.54]	<0.001	2.01 [1.61, 2.54]	<0.001
NEWS2 Total Score	400	1.80 [1.44, 2.27]	<0.001	1.83 [1.46, 2.32]	<0.001
Lymphocyte/CRP ratio	411	0.56 [0.44, 0.71]	<0.001	0.57 [0.45, 0.72]	<0.001
Neutrophil count	413	1.67 [1.34, 2.11]	<0.001	1.72 [1.38, 2.18]	<0.001
Troponin T	138	1.48 [1.00, 2.25]	0.162	1.70 [1.08, 2.80]	0.065
Ferritin	122	1.55 [1.05, 2.40]	0.098	1.63 [1.09, 2.57]	0.055
Albumin	399	0.68 [0.54, 0.85]	0.004	0.67 [0.53, 0.84]	0.004
Estimated GFR	414	0.69 [0.56, 0.86]	0.005	0.69 [0.54, 0.90]	0.023
Respiration rate	404	1.44 [1.17, 1.79]	0.004	1.43 [1.16, 1.78]	0.005
Creatinine	414	1.41 [1.14, 1.76]	0.008	1.41 [1.09, 1.84]	0.035
Neutrophil/lymphocyte ratio	413	1.36 [1.09, 1.72]	0.022	1.37 [1.10, 1.73]	0.023
Oxygen saturation	403	0.74 [0.59, 0.91]	0.019	0.73 [0.58, 0.91]	0.023
Heart rate	404	1.30 [1.05, 1.62]	0.068	1.33 [1.07, 1.66]	0.038
ALT	183	1.16 [0.85, 1.59]	0.932	1.21 [0.88, 1.68]	0.711
Diastolic blood pressure	404	0.89 [0.72, 1.09]	0.783	0.89 [0.72, 1.10]	0.749
Platelet count	415	0.93 [0.75, 1.14]	0.999	0.93 [0.75, 1.14]	0.999
Temperature	404	1.07 [0.86, 1.33]	0.999	1.07 [0.85, 1.32]	0.999
Lymphocyte count	415	1.04 [0.85, 1.27]	0.999	1.05 [0.86, 1.29]	0.999
Systolic blood pressure	404	0.96 [0.77, 1.18]	0.999	0.95 [0.76, 1.19]	0.999
GCS score	172	0.95 [0.70, 1.31]	0.999	0.98 [0.72, 1.36]	0.999
Hemoglobin	415	0.96 [0.78, 1.18]	0.999	1.01 [0.82, 1.25]	0.999

Notes. Odds ratios for one standard deviation change in the respective parameter, each tested in a separate model. [1]Adjusted for age and sex only. [2]Additionally adjusted for comorbidities (hypertension, diabetes, heart diseases, respiratory diseases and chronic kidney disease). [3]FDR-corrected *P*-values based on the Benjamini-Hochberg correction.

Supplementary Table 5: Logistic regression models for each blood/physiological parameter tested separated in the KCH training sample, adjusted for ethnicity, in the subset of patients with information on ethnicity (n=276)

Parameter	N	**Model 1: Age, sex, ethnicity**		**Model 2: + all comorbidities**	
		OR [95% C.I.]	FDR-adjusted *P*-value	OR [95% C.I.]	FDR-adjusted *P*-value
CRP	261	2.17 [1.64, 2.94]	<0.001	2.25 [1.68, 3.07]	<0.001
NEWS2 Total Score	249	2.01 [1.52, 2.72]	<0.001	2.02 [1.52, 2.74]	<0.001
Troponin T	82	1.57 [0.90, 2.90]	0.391	1.93 [1.01, 3.91]	0.192
Lymphocyte/CRP ratio	260	0.56 [0.41, 0.75]	<0.001	0.56 [0.40, 0.75]	<0.001
Neutrophil count	261	1.59 [1.21, 2.16]	0.006	1.62 [1.23, 2.20]	0.004
Albumin	251	0.71 [0.53, 0.93]	0.075	0.67 [0.50, 0.90]	0.040
Heart rate	252	1.45 [1.12, 1.92]	0.031	1.47 [1.13, 1.96]	0.034
Oxygen saturation	251	0.67 [0.50, 0.88]	0.026	0.69 [0.51, 0.91]	0.045
Creatinine	262	1.33 [1.01, 1.77]	0.202	1.42 [1.02, 2.00]	0.165
Respiration rate	252	1.42 [1.12, 1.84]	0.026	1.39 [1.10, 1.80]	0.040
GCS score	109	0.70 [0.43, 1.11]	0.406	0.73 [0.43, 1.19]	0.656
Estimated GFR	262	0.79 [0.60, 1.03]	0.339	0.76 [0.55, 1.04]	0.344
Neutrophil/lymphocyte ratio	261	1.26 [0.97, 1.68]	0.339	1.25 [0.96, 1.68]	0.344
ALT	130	1.13 [0.72, 1.79]	0.999	1.22 [0.77, 1.97]	0.985
Ferritin	81	1.08 [0.64, 1.81]	0.999	1.20 [0.67, 2.15]	0.999
Platelet count	263	0.86 [0.64, 1.11]	0.725	0.86 [0.64, 1.11]	0.765
Diastolic blood pressure	252	0.86 [0.65, 1.14]	0.774	0.88 [0.66, 1.16]	0.946
Temperature	252	1.15 [0.90, 1.49]	0.739	1.14 [0.88, 1.47]	0.894
Systolic blood pressure	252	0.88 [0.68, 1.14]	0.822	0.90 [0.69, 1.17]	0.993
Hemoglobin	263	1.05 [0.81, 1.36]	0.999	1.06 [0.82, 1.38]	0.999
Lymphocyte count	263	1.05 [0.83, 1.33]	0.999	1.06 [0.83, 1.35]	0.999

Supplementary Table 6: Internally validated predictive performance, tested in the KCH training sample, adjusted for ethnicity for patients with information on ethnicity (n=276)

	Included features	Internally validated AUC			Sensitivity	Specificity	PPV	NPV
		Mean	-2SD	+2SD				
1	NEWS2	0.649	0.623	0.676	0.266	0.887	0.590	0.666
2	NEWS2 + **DBP**	0.719	0.695	0.743	0.408	0.805	0.560	0.692
3	NEWS2 + DBP**C**	0.708	0.688	0.728	0.382	0.788	0.528	0.682
4	NEWS2 + **CoV** NEWS2 + CRP + Neutrophil + eGFR + Albumin + Age	0.747	0.732	0.763	0.447	0.794	0.569	0.703

D = Age, sex
C = comorbidities (8 features)
B = bloods (10 features)
P = physiological parameters (7 features)

Supplementary Table 7: Internally validated predictive performance using XGBoost (Gradient Boosting Trees) (n=439)

	Included features	**Internally validated AUC**			**Sensitivity**	**Specificity**	**PPV**	**NPV**
		Mean	-2SD	+2SD				
1	NEWS2	0.647	0.631	0.663	0.336	0.897	0.641	0.712
2	NEWS2 + **DBP**	0.729	0.703	0.753	0.454	0.838	0.603	0.741
3	NEWS2 + **DBPC**	0.721	0.695	0.748	0.430	0.837	0.589	0.732
4	NEWS2 + **CoV** NEWS2 + CRP + Neutrophil + eGFR + Albumin + Age	0.731	0.706	0.756	0.458	0.863	0.646	0.745

Notes.
AUC based on repeated, nested cross-validation (inner loop: 10 fold/20 repeats; outer: 10 fold/20 repeats).
D = Age, sex
C = comorbidities (8 features)
B = bloods (10 features)
P = physiological parameters (7 features)