medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 .

Title

Development and validation of a dynamic 48-hour in-hospital prognostic risk stratification for COVID-19 in a UK teaching hospital: a retrospective cohort study

Martin Wiegand, Sarah L. Cowan, Claire S. Waddington, David J. Halsall, Victoria L. Keevil, Brian D. M. Tom, Vince Taylor, Effrossyni Gkrania-Klotsas, Jacobus Preller*, Robert J. B. Goudie*

* Joint last

Martin Wiegand, PhD, MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, UK

Sarah L. Cowan, BM BCh, MRCP (UK), FFICM, Addenbrooke's Hospital, Cambridge, UK https://orcid.org/0000-0001-9149-3530

Claire S. Waddington, BMedSci, BMBS, MSc, DPhil, MRCP (UK), Department of Medicine, University of Cambridge, UK

David J. Halsall PhD, FRCPath, Cambridge University Hospitals Trust, Cambridge, UK.

Victoria L. Keevil, BMBCh, MA, MPhil, PhD, MRCP (London), Department of Medicine for the Elderly, Addenbrooke's Hospital, Cambridge, UK; Department of Medicine, University of Cambridge https://orcid.org/0000-0001-6148-0640

Brian D. M. Tom, PhD, MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, UK https://orcid.org/0000-0002-3335-9322

Vince Taylor, Cancer Research UK, Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK

Effrossyni Gkrania-Klotsas, MD, MPH, FRCP, PhD, Department of Infectious Diseases, Cambridge University Hospitals, Cambridge, UK https://orcid.org/0000-0002-0930-8330

Jacobus Preller, MBChB, MRCP, EDIC, DICM, FFICM, Cambridge University Hospitals, Cambridge, UK

Robert J. B. Goudie, PhD, MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, UK https://orcid.org/0000-0001-9554-1499

Corresponding author:

Martin Wiegand, MRC Biostatistics Unit, School of Clinical Medicine, East Forvie Building, University Forvie Site, CB2 0SR, University of Cambridge, UK martin.wiegand@mrc-bsu.cam.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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

We propose a prognostic dynamic risk stratification for 48-hour in-hospital mortality in patients with COVID-19, using demographics and routinely-collected observations and laboratory tests: age, Clinical Frailty Scale score, heart rate, respiratory rate, SpO2/FiO2 ratio, white cell count, acidosis (pH < 7.35) and Interleukin-6. We train and validate the model using data from a UK teaching hospital, adopting a landmarking approach that accounts for competing risks and informative missingness. Internal validation of the model on the first wave of patients presenting between March 1 and September 12, 2020 achieves an AUROC of 0.90 (95% CI 0.87–0.93). Temporal validation on patients presenting between September 13, 2020 and January 1, 2021 gives an AUROC of 0.91 (95% CI 0.88-0.95). The resulting mortality stratification tool has the potential to provide physicians with an assessment of a patient's evolving prognosis throughout the course of active hospital treatment.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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

SARS-CoV-2 virus infection, the cause of COVID-19, results in a spectrum of disease ranging from asymptomatic infection through to life threatening disease requiring critical care, and even death. For patients admitted to hospital, it is essential to identify who is at risk of deterioration and death to enable timely targeted interventions (such as immune modulation and mechanical ventilation), to facilitate appropriate resource allocation and patient flow, and to inform discussions with patients and families.

Most existing disease severity prediction models for COVID-19 use only data that are available at the time of admission to hospital. Numerous such models have been proposed for both mortality and composite escalation/mortality outcomes, including new and re-purposed severity and early warning scores [1–7] and time-to-event models [8–13]. Most, however, perform only moderately well and are at high risk of bias [14].

While some markers of severity, such as sex and age can be assumed constant for the duration of the hospital visit, others, such as clinical observations and blood test results, can change markedly over the course of admission. COVID-19 is a dynamic disease in which patients can deteriorate over a short time period or suffer acute complications e.g. thromboembolism [15–16]. This may have a significant effect on a patient's prognosis that cannot be foreseen by a point-of-admission model.

Dynamic models that assimilate clinical data as it accrues may provide more accurate and clinically useful prediction of a patient's clinical course and prognosis over the subsequent days than that of point-of-admission models. Predictive models that incorporate post-admission information are limited in number and scope. Some models using information from the first four or five days after admission to predict mortality or deterioration have been proposed, but do not continue beyond the first few days of admission [17–18]. Other more recent models have made use of additional post-admission data, using a time-varying Cox model for mortality and escalation [19] or a machine learning model for mortality [20]. While indicating promising discrimination, these models use clinically unjustifiable or unclear methods for handling missing data and censoring, and do not account for informative missingness or consider the effect of treatments. Informative missingness describes the fact that in routinely-collected data the availability (or absence) of a result or observation may be related to the probability of the outcome. For example, a more extensive panel of investigations may be sent in patients thought

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

more likely to benefit from Intensive Care Unit (ICU) admission. While often ignored, such effects can be strong in Electronic Health Record (EHR) data [21–22].

We propose a prognostic risk stratification score for hospital patients with COVID-19, based on prediction of mortality in the subsequent 48 hours, using routinely-collected clinical data. The model is based upon a principled statistical approach called landmarking [23] that allows inclusion of any time-varying clinical parameters recorded prior to the time of prediction, whilst appropriately accounting for censoring and changes in the set of patients at risk. Our model further accounts for informative missingness and competing risks, which arise when there are two or more mutually exclusive outcomes: for example, once a patient is discharged, the risk of in-hospital mortality (during that admission) is removed, and therefore discharge is a "competing risk".

Methods

Study design

This is a retrospective cohort study of all patients presenting to Cambridge University Hospitals, a regional, tertiary care, university hospital in the East of England, between March 1, 2020 and January 1, 2021. This hospital is the sole admission hospital for patients in its immediate catchment population with COVID-19, and is a regional referral centre for a wide range of specialist services (not including ECMO).

The model was trained using data from patients presenting between March 1, 2020 and September 12, 2020 (hereafter “Wave 1”). A temporal validation of the model was then performed using patients presenting between September 13, 2020 and January 1, 2021 (hereafter “Wave 2”).

We report our findings according to the Transparent Reporting of multivariate prediction models for Individual Prediction Or Diagnosis (TRIPOD) reporting guidelines [24].

Study Population

All adults (>= 18 years of age) presenting to hospital during the study period and diagnosed with COVID-19 were included. Diagnosis was based on either a positive diagnostic SARS-CoV-2 test during or up to 14 days prior to the hospital visit, or a clinical diagnosis of COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Diagnostic testing used either a real-time reverse transcription polymerase chain reaction (RT-PCR) of the RdRp gene from a nasopharyngeal swab, or the SAMBA II point-of-care test used at the hospital [25]. Patients with clinically diagnosed COVID-19 (based on symptoms, and in the opinion of the treating clinician) were included because diagnostic testing was limited during the early stages of the pandemic [26]. Clinical diagnosis of COVID-19 was identified using International Classification of Diseases 10th Edition (ICD-10) codes in the EHR.

We include only the first hospital visit for each patient after the positive test; any re-admissions were excluded. Nosocomial infection was defined as a first positive SARS-CoV-2 test or diagnosis more than 10 days after hospital admission. Since the first prediction by our model is made at 6 hours (to allow time for laboratory investigations), patients who died, were discharged or were classified as end of life within 6 hours of presentation to hospital were excluded.

All patients were treated as per standard treatment approaches, as guided by detailed local guidance in use in the hospital at the time. Patients were also eligible for inclusion in relevant clinical trials running at the hospital during the study period. These studies were the TACTIC-E and TACTIC-R trials (ISRCTN11188345 https://doi.org/10.1186/ISRCTN11188345), the REMAP-CAP platform trial for intensive care patients (ISRCTN67000769 https://doi.org/10.1186/ISRCTN67000769), and the RECOVERY trial (ISRCTN50189673 https://doi.org/10.1186/ISRCTN50189673).

Outcomes

Throughout each patient's hospital visit, we aim to predict all-cause in-hospital mortality during the next 48 hours, a time period that we refer to as the "prediction horizon". We also consider two other outcomes, which are competing risks for the primary outcome: transfer to a tertiary Intensive Care Unit (ICU) for ECMO; and discharge from the hospital due to clinical improvement. Patients were followed up until January 3, 2021.

Ethics

The study was approved by a UK Health Research Authority ethics committee (20/WM/0125). Patient consent was waived because the de-identified data presented here were collected during routine clinical practice; there was no requirement for informed consent.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Data collection

The study used routinely collected data, extracted in anonymised form from the hospital EHR system, Epic (Epic Systems Corporation, Verona, Wisconsin).

We selected a list of 59 candidate clinical parameters (Table S1) that have been included in existing point-of-admission prediction models or were clinically judged to be likely predictors. These are divided into five categories: demographics; comorbidities; observations; laboratory tests; and treatments, interventions and level of care.

Demographics

Basic patient demographics were extracted from the hospital EHR: age, sex, ethnicity, and Body Mass Index (BMI).

Comorbidities

Twelve comorbidities that have previously been associated with COVID-19 [27] were identified by the presence of the corresponding ICD-10 codes entered in the EHR prior to the time at which the prediction is made (either before or during the hospital visit). Table S2 provides the ICD-10 codes used to define each comorbidity. In addition to specific comorbidities, frailty amongst patients over 65 years old was assessed by the Clinical Frailty Scale (CFS) score [28]. For patients for whom a CFS score had not been recorded by the treating team, a consultant or specialist registrar in Geriatric Medicine reviewed the clinical records and assigned a CFS score using only information recorded at the time of admission [29]. This approach has been shown to have good agreement with CFS scores assigned after face to face assessment (inter-rater reliability kappa 0.84) [30].

Observations

We included the following observations that are regularly recorded in the EHR: heart rate (HR), mean arterial pressure, temperature and respiratory rate (RR). We summarised the measurements recorded over the previous 24-hour period as follows: mean, minimum and maximum value. We also calculated the trend as the difference between the median value for the last 24 hours, and the median value for the 24 hours prior to this. The Glasgow Coma Score (GCS) was extracted from the EHR; patients without a recorded GCS were assumed to have a GCS >= 12. PaO2/FiO2 (P/F) and SpO2/FiO2 (S/F) ratios were calculated to indicate the

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

severity of hypoxia [31–32]. SpO2 itself was not included as a potential predictor as our exploratory work suggested that, without accounting for FiO2, this largely reflected a patient's assigned oxygen saturation targets, and therefore acted as a proxy for underlying respiratory disease. Where only oxygen flow rate was available, FiO2 was estimated according to the EPIC II conversion tables [33].

Laboratory tests

For each of the 31 laboratory tests we considered, we included results up to 48 hours prior to the time at which the prediction was made. Where more than one result was available, we used the most recent result. In addition, for 7 of the most frequently measured blood tests (C-reactive protein (CRP), white cell count (WCC), platelets, haemoglobin, creatinine, sodium, potassium), we included the trend, defined as the difference between the median value for the last 24 hours, and the median value for the 24 hours prior to that. The neutrophil/lymphocyte and Interleukin-6/Interleukin-10 (serum IL-6/IL-10) ratios have previously been identified as prognostic, therefore we also considered these as potential predictors [9, 18, 34]. For blood markers where both abnormally low and abnormally high results could potentially be associated with poor prognosis (sodium and blood gas pH), we included the maximum deviation below and above the standard range in the previous 24 hours. We adjusted venous pH results by adding 0.03 to approximate arterial pH results [35].

Treatments, interventions and level of care

We included 5 indicators of treatments, interventions and levels of care. The level of care of the patient was summarised by whether the patient had been in an ICU bed in the previous 24 hours. Mechanical ventilation was defined as patients receiving invasive ventilation during the previous 24 hours, either via endotracheal tube or tracheostomy. The use of renal replacement therapy during the last 24 hours was identified from the EHR. Cardiovascular support was defined as the administration of any vasopressors or inotropes in the last 24 hours. We also included in-hospital steroid administration prior to the prediction time, given that steroids can decrease the risk of death in patients with COVID-19 and that clinical trials operated in our hospital for part of the study time [36-37].

Model development

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

We use the landmarking approach to dynamic prediction [23, 38-39]. Landmarking fits a series of models at a sequence of fixed time origins called landmark times. These are the time points at which we extract the data used to develop the proposed model. At each landmark time, a time-to-event model is used to describe the likelihood of the event occurring within the next 48 hours (the "prediction horizon"). Only data recorded before (or at) the landmark time are used in each model; no data from the future can be used. If the event of interest happens after the prediction horizon then the event is treated as censored at this landmark. Patients who have had any event prior to the landmark time are excluded, since these patients are no longer at risk. We use the supermodel approach in which the time-to-event model is assumed constant across landmark times [40].

Our landmark times are every 24 hours during the hospital visit, starting 6 hours after presentation to the hospital, to allow sufficient clinical information to accrue for prediction at the initial landmark to be feasible. We only use data at each landmark time from patients being actively treated for COVID-19 at that point in time. Landmark times after a patient transitions to end of life care were excluded, meaning that once patients transition to end-of-life care no further predictions are made, but any events occurring within the existing prediction horizon are still included (i.e. transitions to end-of-life care are not treated as censoring events). For patients with nosocomial infection only landmark times subsequent to the diagnosis of COVID-19 are included.

We use a Fine-Gray competing risk model [41–42] to account for the competing risks of in-hospital death, which are transferral for ECMO and hospital discharge.

We handle missing data using the missingness indicator approach because the recording in the EHR of a clinical parameter, regardless of the value, is often indicative of the treating health professional's contemporaneous view of the patient's prognosis [43–44]. To do this we augment the set of potential predictors with binary variables that indicate whether, during the window of time we consider, any measurement of the corresponding parameter is available for that patient. The missing indicator approach avoids the need to make the missing at random (MAR) assumption that is unlikely to hold in these data [45].

To select the most predictive parameters into the model we used standard penalised variable selection, specifically Smoothly Clipped Absolute Deviations (SCAD), with the tuning parameter

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

chosen to minimise the Bayesian Information Criterion [46]. We paired parameters together with their corresponding missingness indicator to prevent inclusion of an incompletely-observed parameter without its missingness indicator, using the group SCAD [47]. We also allowed missingness indicators to be included by themselves, in case the presence of a parameter is informative irrespective of its value.

All analyses were conducted in R version 3.6 [48].

Model assessment

The discriminative ability of our model was assessed visually using the Receiver Operator Characteristic (ROC) curve, showing the false positive rate (FPR, also equal to 1-specificity) against true positive rate (TPR, also equal to sensitivity); and the Precision-Recall (PR) curve, showing precision (positive predictive value, PPV) against recall (sensitivity). PPV is the key metric for a dynamic predictive model such as we propose, because the low incidence of the primary outcome leads to a strong imbalance of events and non-events that is not accounted for by the ROC curve [49]. We also assess the Number Needed to Evaluate (NNE), defined as the number of patients needed to evaluate in order to identify one in-hospital death within the prediction horizon, calculated as 1/PPV.

Quantitative assessment of discrimination was performed using the Area Under the ROC (AUROC) curve, in which 0.5 indicates no discrimination and 1.0 indicates perfect discrimination. For validation of the performance of the model on the training data, in addition to the unadjusted AUROC, we also performed repeated 5-fold cross-validation to account for uncertainty and over-optimism due to the complete model building process (including variable selection) [50]. We also calculate the Area Under the PR Curve (AUPRC) since it provides a clearer performance summary than AUROC when the primary outcome has low incidence, as here [51].

We assessed calibration visually using a calibration plot of predicted risk against observed mortality rate. We also quantitatively assessed the calibration slope, testing whether beta=1 in the logistic regression model logit(probability of death) = alpha + beta * logit(p), where p is the risk score; and calibration-in-the-large, testing whether intercept alpha=0 when fixing beta=1 [52].

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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

Baseline characteristics and outcomes

519 patients presented to hospital with COVID-19 during Wave 1 (March 1, 2020 and September 12, 2020), of whom 46 were excluded due to discharge (34), death (2) or transition to end of life care (10) prior to the first landmark time (i.e. within 6 hours of presentation). 473 patients were therefore included in the development of the model.

Table 1 shows baseline characteristics, comorbidities and severity markers of all patients. The median age was 69 years. 196 (41.4%) patients were female. 99 patients (20.9%) died during their hospital admission, 5 patients (1.1%) were transferred for extracorporeal membrane oxygenation (ECMO) and 369 (78.0%) were discharged from hospital alive. The median length of hospital stay was 10.8 days. 103 patients (21.8%) were admitted to ICU of whom 82 required mechanical ventilation, 86 required cardiovascular support and 32 required renal replacement therapy. 36 (7.6%) patients had nosocomial infection and compared to the rest of the cohort were slightly younger (median age 65) and more likely to be male (66.6%) but had a similar mortality rate (22.2%).

Development of prediction model

In total we include 6846 landmark times for training the model, with a median of 9 (IQR 3–17) landmark times per patient. In the 48-hour prediction horizon following these landmark times, there were 119 in-hospital death events (1.7% of landmarks), 658 hospital discharge events (9.6%) and 10 transfers for ECMO (0.1%). Note that, since landmarks occur every 24 hour and the prediction horizon is 48 hours, patient events will usually occur within the prediction horizon of two adjacent landmark times. Table S1 reports summary statistics, missingness and the number of measurements available per landmark time for each predictor. No patients were excluded due to missing data.

Model results

Our proposed model for 48-hour in-hospital mortality includes age, Clinical Frailty Scale (CFS) score, heart rate (HR), respiratory rate (RR), oxygen saturation/fraction of inspired oxygen (SpO2/FiO2) ratio, white cell count (WCC), acidosis (pH < 7.35) and Interleukin-6 (IL-6). The

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

estimated coefficients are shown in Table 2, and the estimated hourly baseline cumulative subdistribution hazards are in Table S3.

Internal performance assessment

Figure 1 shows the internal performance metrics. The unadjusted internal area under receiver-operating characteristic curve (AUROC) was 0.90 (95% CI 0.87–0.93) and the median cross-validation AUROC was 0.87, both indicating good discrimination (Figure 1A). The precision-recall (PR) curve (Figure 1B) also shows good discrimination, with an area under precision-recall curve (AUPRC) 0.31, in a population with 48 hour in-hospital mortality of 0.017 (1.7%), and the number needed to evaluate (NNE) < 10 for sensitivity less than 0.75 (Figure 1C). Figure 1D shows the calibration plot. The calibration intercept alpha = -0.02 (95% CI -0.22–0.17) indicates that the mean predicted probabilities match the mean observed mortality, while the calibration slope beta = 1.16 (95% CI 1.02–1.31) suggests that the observed mortality in high predicted risk patients slightly exceeds the predicted mortality risk.

Temporal validation of prediction model

We assessed the performance of the model by applying it to held-out data corresponding to visits during Wave 2 (September 13, 2020 and January 1, 2021). 405 patients presented to the study hospital during this period. In total we include 3086 landmark times for training the model, with a median of 4 (IQR 1–10) landmark times per patient. In the 48-hour prediction horizon following these landmark times, there were 55 in-hospital death events (1.8% of landmarks), 441 hospital discharge events (14.3%) and 0 transfers for ECMO (0%). Characteristics are summarised in Table 1. Of note, compared to Wave 1, patients presenting in Wave 2 were younger (median age 61 vs 69) and more likely to be female (50.1% vs 41.4%). Note that because follow-up continued until January 3, 2021, the outcome during the 48-hour prediction horizon was known for all landmark times up to January 1, 2021, and so patients remaining in hospital at the study end-date could be included in the validation.

Figure 2 shows the temporal validation performance metrics, obtained by applying the trained model to the Wave 2 patients. The receiver-operating characteristic (ROC) curve (Figure 2A) shows the model continues to discriminate well, with AUROC 0.91 (95% CI 0.88-0.95). The PR curve (Figure 2B) shows that the positive-predictive value (PPV) is consistently well above the 48-hour in-hospital mortality incidence of 0.018 (1.8%) in the Wave 2 cohort across all

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

sensitivities, with AUPRC 0.21, and NNE < 10 for sensitivities between 0.01 and 0.87 (Figure 2C). Figure 2D shows the calibration plot, which shows a tendency of the model to underpredict risk in the higher risk patients: calibration-in-the-large was 0.33 (95% CI 0.04-0.60), suggesting the mean of the predicted probabilities is lower than the mean observed mortality, and calibration slope 1.15 (95% CI 0.95-1.37), indicating that the spread of predicted risk corresponds reasonably well with the spread of observed mortality.

Discussion

SARS-CoV-2 causes a wide spectrum of disease, from mild and short lived disease, through to more severe disease that can evolve over weeks, and may necessitate critical care management and even result in death. There is a pressing clinical need to be able to anticipate both disease severity and the trajectory of illness in order to facilitate patient management, resource allocation, and inform discussions with both patients and families.

Although risk factors for severe disease, such as age and male sex have been widely recognised, these static risk factors provide little nuance or discrimination at the individual patient level. Disease severity models have been proposed that provide a more accurate assessment but these have focused on a single time point such as admission to hospital, and do not respond to changes in the clinical picture as the disease evolves. The model described herein uniquely incorporates both static risk factors (age and CFS) and evolving clinical and laboratory data, providing a dynamic 48-hour risk prediction model that can adapt to both sudden and gradual changes in an individual patient's clinical condition. Our model is further strengthened by the competing-risk landmarking approach we adopt, allowing the model to account for events other than death that remove the patient from the population at risk.

The data used in the model were routinely collected demographic and clinical data from during the patient's hospitalisation, and were automatically extracted from patient EHRs. As such, this model could be readily incorporated into routine clinical care, providing invaluable information to clinicians on the ground who are managing patients with COVID-19. More broadly, our model highlights the potential utility of EHR data to inform our understanding of disease by making it feasible to collect and analyse the detailed data accrued during routine care, in contrast to the limited information that can be gathered via manual data collection.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Several methodological aspects of our approach further strengthened our model. Firstly, we account for competing risks in the model. Competing risks refers to the situation whereby the outcome (risk) of interest (in this case in-hospital mortality) can only happen whilst the patient is in hospital, and therefore the outcome of interest is 'competing' against the risk of transfer to another hospital and/or discharge from hospital. Allowance for this has been shown to be important in predictive modeling [53]. Secondly, allowance was made for the potential of the availability of observations and investigations to in itself be a reflection of disease severity. For example, the fact that an arterial blood gas has been taken may reflect the clinical impression that a patient is deteriorating. While multiple imputation is often used in clinical prediction models because it gives unbiased estimates under the missing at random (MAR) assumption, it is unlikely that the MAR assumption holds in the routinely-collected EHR data that we use [45]. The missing indicator method that we adopt does not rely on the MAR assumption and has been found to lead to improved predictive performance in EHR data [43-45]. Furthermore, we do not seek to make prognostic predictions for patients after clinicians have identified them as entering the last few hours or days of life. Such patients are commonly transitioned to a symptom management approach, in which observations, investigations and active treatment are discontinued: in our data, no vital sign observations were recorded on 43% of days during end-of-life care, compared to 0% of days during active treatment. This extreme missingness is therefore a strong predictor of mortality, and including these end-of-life time periods would distort the model for other patients by exaggerating the importance of missing data. In addition, predicting end-of-life after it is clinically apparent would have little clinical utility.

Several predictors of disease severity selected by our model have also been identified by models that aimed to assess severity of disease at the point of admission to hospital, and in epidemiological studies of risk factors for severe disease. Increasing age is widely recognised as being the strongest predictor of poor outcome from COVID-19 [3, 11, 27]. Frailty has similarly been shown to be a strong independent predictor of mortality in hospitalised older adults [54], including those with COVID-19 [29,55], and it is therefore unsurprising that the frailty score was selected in the model.

The deleterious effect of SARS-CoV-2 infection on respiratory function is one of the commonest and often most severe effects of illness, and frequently precipitates hospital admission and the need for critical care. Markers of respiratory function, including respiratory rate [3, 4, 12] and SpO2/FiO2 (S/F) ratio [56] have been included in previous point-of-admission models; similarly

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 recent ISARIC 4C deterioration model includes SpO2 and the need for supplemental oxygen [3]. The S/F ratio, as selected by our model may be a more informative measure of respiratory compromise as it allows a fully quantitative rather than dichotomous measure of the need for additional oxygen, as well as allowing for the confounding effect of variation in the target oxygen saturations in different patient groups. Clinicians often set a lower target SpO2 in patients with pre-existing respiratory disease, and thus oxygen therapy is not initiated until a much lower SpO2 is reached compared to otherwise healthy patients.

Our model selected two markers of infection and inflammation: WCC and IL-6. This is consistent with other findings [11, 57–58]. IL-6 was included in the routine COVID-19 panel of blood tests at the study hospital but we recognise that this may be less commonly requested in other hospitals. To assess whether C-reactive protein (CRP) could serve as a proxy for IL-6 in our model when it is not available, we refitted the model with CRP in place of IL-6 (Tables S4–S5). The AUROC is slightly lower on both training (0.89, 95% CI 0.85–0.93) and test (0.83, 95% CI 0.78–0.89) data, yielding a slightly weaker but potentially more broadly applicable model. The preference of the model for IL-6 over CRP may reflect the fact that IL-6 is responsible for the production of CRP and, as such, is an earlier and more dynamic marker of the inflammatory response [59].

The inclusion of acidosis in our model is more novel, although it has previously been noted as a marker of disease severity [60]. Acidosis frequently complicates respiratory, renal and advanced circulatory failure in patients, and is therefore frequently observed in patients with severe disease. The separate inclusion of the severity of acidosis and alkalosis in our set of candidate predictors removed the need to assume a linear effect of pH, allowing pH changes in either direction to be accounted for. This would avoid, for example, a minor negative effect of alkalosis from masking a more major effect of acidosis.

There are several limitations to our study. We have not incorporated imaging data that have been used as a proxy for disease severity in some clinical trials, although we have included an extensive set of relevant clinical information. We also chose to include only laboratory results up to 48 hours and vital signs up to 24 hours before the landmark time. Exploiting older data in addition might improve the predictive ability of our model, but would also be likely to increase the model's complexity considerably, and therefore decrease its real-world utility. Our data were gathered from a single centre, and therefore the generalisability of our findings to other centres

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

and populations are uncertain. Further, our model was generated from a relatively modest sample size due to the relatively low prevalence of COVID-19 patients in the catchment population of the hospital, particularly during the early months of the pandemic. One advantage of using this single dataset from a large, tertiary hospital was that the hospital never became overwhelmed with patients, and therefore it is considered that patients received care according to what was felt to be clinically appropriate rather than according to what resources were available. It is also important to note that as the pandemic has evolved in the UK, there have been changes in both the clinical care of patients (notably with the routine inclusion of steroid therapy for patients requiring oxygen) and in the strains of the virus circulating. It is encouraging that the model continued to perform well in the Wave 2 validation data, but over time, changes such as these may influence the clinical picture of the disease, its severity and the risk factors for disease. It is likely therefore that the model will need to be updated as the pandemic evolves and the utilisation of routinely available data in this model makes this relatively simple to do.

The ongoing COVID-19 pandemic continues to affect the lives of many and to put extreme pressure on clinical services. We have developed and validated a dynamic prediction model with high discrimination of 48-hour in-hospital mortality in COVID-19 using routine clinical data that updates over the course of the illness and patient admission. This represents a significant advance on existing point-of-admission COVID-19 prediction models: it has the potential to inform patient management and resource planning and allocation in real-world settings. In doing so, this model could considerably improve both individual patient care, and the ability to manage pandemic response on a hospital wide scale.

Data availability

The de-identified data that support the findings of this study are available from Cambridge University Hospitals but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request with permission of Cambridge University Hospitals.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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] Carr E, Bendayan R, Bean D, et al. Evaluation and Improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. *BMC Med* 2021; 19(23) https://doi.org/10.1186/s12916-020-01893-3.

[2] Fan G, Tu C, Zhou F, et al. Comparison of Severity Scores for COVID-19 Patients with Pneumonia: A Retrospective Study. *Eur Respir J* 2020; 56(3): 2002113. https://doi.org/10.1183/13993003.02113-2020.

[3] Gupta RK, Harrison EM, Ho A, et al. Development and Validation of the ISARIC 4C Deterioration Model for Adults Hospitalised with COVID-19: A Prospective Cohort Study. *Lancet Respir Med* 2021; 2: e592. http://doi.org/10.1016/S2213-2600(20)30559-2.

[4] Haimovich AD, Ravindra NG, Stoytchev S, et al. Development and Validation of the Quick COVID-19 Severity Index: A Prognostic Tool for Early Clinical Decompensation. *Ann Emerg Med* 2020; 76(4): 442–453. http://doi.org/10.1016/j.annemergmed.2020.07.022.

[5] Mei J, Hu W, Chen Q, et al. Development and External Validation of a COVID-19 Mortality Risk Prediction Algorithm: A Multicentre Retrospective Cohort Study. BMJ Open 2020; 10: e044028. https://doi.org/10.1136/bmjopen-2020-044028

[6] Zhang K, Liu X, Shen J, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* 2020; 181(6): 1423–1433.e11. http://doi.org/10.1016/j.cell.2020.04.045.

[7] Shang Y, Liu T, Wei Y, et al. Scoring Systems for Predicting Mortality for Severe Patients with COVID-19. *EClinicalMedicine* 2020; 24: 100426. https://doi.org/10.1016/j.eclinm.2020.100426

[8] Foy BH, Carlson JCT, Reinertsen E, et al. Association of Red Blood Cell Distribution Width with Mortality Risk in Hospitalized Adults with SARS-CoV-2 Infection. *JAMA Netw Open* 2020; 3(9): e2022058. https://doi.org/10.1001/jamanetworkopen.2020.22058.

[9] Ye W, Chen G, Li X, et al. Dynamic Changes of D-Dimer and Neutrophil-Lymphocyte Count Ratio as Prognostic Biomarkers in COVID-19. *Respir Res* 2020; 21(1): 1–7. http://doi.org/10.1186/s12931-020-01428-7.

[10] Sun H, Ning R, Tao Y, et al. Risk Factors for Mortality in 244 Older Adults with COVID-19 in Wuhan, China: A Retrospective Study. *J Am Geriatics Soc* 2020; 68:E19-E23. https://doi.org/10.1111/jgs.16533

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

[11] Cummings MJ, Baldwin MR, Abrams D, et al. Epidemiology, Clinical Course, and Outcomes of Critically Ill Adults with COVID-19 in New York City: a prospective cohort study. *Lancet* 2020; 395: 1763–70. https://doi.org/10.1016/S0140-6736(20)31189-2

[12] Cecconi M, Piovani D, Brunetta E, et al. Early Predictors of Clinical Deterioration in a Cohort of 239 Patients Hospitalized for Covid-19 Infection in Lombardy, Italy. *J Clin Med* 2020; 9(5): 1548. http://doi.org/10.3390/jcm9051548.

[13] Liang W, Yao J, Chen A, et al. Early Triage of Critically Ill COVID-19 Patients Using Deep Learning. *Nat Commun* 2020; 11(1): 1–7. http://doi.org/10.1038/s41467-020-17280-8.

[14] Wynants L, Van Calster B, Collins GS, et al. Prediction Models for Diagnosis and Prognosis of Covid-19: Systematic Review and Critical Appraisal. *Br Med J* 2020; 369(6): 327–331. http://doi.org/10.1136/bmj.m1328.

[15] Klok FA, Kruip MJHA, van der Meer NJM, et al. Incidence of Thrombotic Complications in Critically Ill ICU Patients with COVID-19. *Thromb Res* 2020; 191: 145–147. http://doi.org/10.1016/j.thromres.2020.04.013.

[16] Poissy J, Goutay J, Caplan M, et al. Pulmonary Embolism in Patients with COVID-19. *Circulation* 2020; 142(2): 184–186. http://doi.org/10.1161/CIRCULATIONAHA.120.047430.

[17] Berzuini C, Hannan C, King A, et al. Value of Dynamic Clinical and Biomarker Data for Mortality Risk Prediction in COVID-19: A Multicentre Retrospective Cohort Study. *BMJ Open* 2020; 10(9): e041983. http://doi.org/10.1136/bmjopen-2020-041983.

[18] McElvaney OJ, Hobbs BD, Qiao D, et al. A Linear Prognostic Score Based on the Ratio of Interleukin-6 to Interleukin-10 Predicts Outcomes in COVID-19. *EBioMedicine* 2020; 61: 103026. https://doi.org/10.1016/j.ebiom.2020.103026

[19] Crooks CJ, West J, Fogarty A, et al. Predicting the Need for Escalation of Care or Death from Repeated Daily Clinical Observations and Laboratory Results in Patients with SARS-CoV-2 During 2020: A Retrospective Population-Based Cohort Study from the United Kingdom. *medRxiv* 2020:12.14.20248181. http://doi.org/10.1101/2020.12.14.20248181.

[20] Wanyan T, Vaid A, De Freitas JK, et al., Relational Learning Improves Prediction of Mortality in COVID-19 in the Intensive Care Unit. *IEEE Trans on Big Data* 2021: in press. https://doi.org/10.1109/TBDATA.2020.3048644.

[21] Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data: A Systematic Review. *J Am Med Inform Assoc* 2017; 24(1): 198-208. https://doi.org/10.1093/jamia/ocw042.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

[22] Agniel D, Kohane IS, Weber GM. Biases in Electronic Health Record Data Due to Processes Within the Healthcare System: Retrospective Observational Study. *Br Med J* 2018; 361: k1479. https://doi.org/10.1136/bmj.k1479.

[23] van Houwelingen HC, Putter H. Dynamic Prediction in Clinical Survival Analysis. Boca Raton: CRC Press, 2012. https://doi.org/10.1201/b11311

[24] Collins GS, Reitsma J B, Altman DG, et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015: 162(1): 55. http://doi.org/10.7326/M14-0697.

[25] Assennato SM, Ritchie AV, Nadala C, et al. Performance Evaluation of the SAMBA II SARS-CoV-2 Test for Point-of-Care Detection of SARS-CoV-2. *J Clin Microbiol* 2020; 59(1): e01262-20. http://doi.org/10.1128/JCM.01262-20.

[26] Iacobucci G. Covid-19: What is the UK's Testing Strategy? *Br Med J* 2020; 368: m1222. http://doi.org/10.1136/bmj.m1222.

[27] Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors Associated with COVID-19-Related Death Using OpenSAFELY. *Nature* 2020; *584*(7821): 430–436. http://doi.org/10.1038/s41586-020-2521-4

[28] Rockwood K, Song X, MacKnight C, et al. A Global Clinical Measure of Fitness and Frailty in Elderly People. *Can Med Assoc J* 2005; 173(5): 489-95. https://doi.org/10.1503/cmaj.050051.

[29] Osuafor, CN, Davidson C, Mackett AJ, et al. Clinical Features, Inpatient Trajectories and Frailty in Older Inpatients with COVID-19: A Retrospective Observational Study. *Geriatrics* 2021; 6(1): 11. http://doi.org/10.3390/geriatrics6010011

[30] Marincowitz C, Turner V, Allgar V, et al. Can Patient Frailty Be Estimated from Inpatient Records? A Prospective Cohort Study. *Adv Geriatr Med Res*. 2020; 2(1): e200004. https://doi.org/10.20900/agmr20200004

[31] Rice TW, Wheeler AP, Bernard GR, et al. Comparison of the Spo2/Fio2 Ratio and the PaO2/FiO2 Ratio in Patients with Acute Lung Injury or ARDS. *Chest* 2007; 132(2): 410–417. http://doi.org/10.1378/chest.07-0617

[32] Pandharipande PP, Shintani AK, Hagerman HE, et al. Derivation and Validation of SpO2/FiO2 Ratio to Impute for PaO2/FiO2 Ratio in the Respiratory Component of the Sequential Organ Failure Assessment Score*. *Crit Care Med* 2009; 37(4): 1317–1321. http://doi.org/10.1097/CCM.0b013e31819cefa9

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

[33] Vincent JL, Rello J, Marshall J, et al. International Study of the Prevalence and Outcomes of Infection in Intensive Care Units. *JAMA* 2009; 302(21): 2323–2329. http://doi.org/10.1001/jama.2009.1754

[34] Lin HA, Lin SF, Chang HW, et al. Clinical Impact of Monocyte Distribution Width and Neutrophil-to-Lymphocyte Ratio for Distinguishing COVID-19 and Influenza from Other Upper Respiratory Tract Infections: A Pilot Study. *PLoS One* 2020; 15(11): e0241262. https://doi.org/10.1371/journal.pone.0241262

[35] Byrne AL, Bennett M, Chatterji R, et al. Peripheral Venous and Arterial Blood Gas Analysis in Adults: Are They Comparable? A Systematic Review and Meta-Analysis. *Respirology* 2014; 19(2): 168–175. http://doi.org/10.1111/resp.12225

[36] The RECOVERY Collaborative Group. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *N Engl J Med* 2020; NEJMoa2021436. http://doi.org/10.1056/NEJMoa2021436

[37] The WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-Analysis. *JAMA* 2020; *324*(13): 1330–1341. http://doi.org/10.1001/jama.2020.17023

[38] Wells BJ, Chagin KM, Li L, et al. Using the Landmark Method for Creating Prediction Models in Large Datasets Derived from Electronic Health Records. *Health Care Manag Sci* 2015; 18(1): 86–92. http://doi.org/10.1007/s10729-014-9281-3

[39] Paige E, Barrett J, Stevens D, et al. Landmark Models for Optimizing the Use of Repeated Measurements of Risk Factors in Electronic Health Records to Predict Future Disease Risk. *Am J Epidemiol* 2018; 187(7): 1530–1538. http://doi.org/10.1093/aje/kwy018

[40] van Houwelingen HC. Dynamic Prediction by Landmarking in Event History Analysis. *Scand J Stat* 2007; 34: 70–5. https://doi.org/10.1111/j.1467-9469.2006.00529.x

[41] Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc* 1999; 94(446): 496–509. http://doi.org/10.1080/01621459.1999.10474144

[42] Wu C, Li L, Li R. Dynamic Prediction of Competing Risk Events Using Landmark Sub-Distribution Hazard Models with Multiple Longitudinal Biomarkers. *Stat Methods Med Res* 2020; 29(11): 3179–3191. http://doi.org/10.1177/0962280220921553

[43] Sisk R, Lin L, Sperrin M, et al. Informative Presence and Observation in Routine Health Data: A Review of Methodology for Clinical Risk Prediction, *J Am Med Inform Assoc* 2021; 28(1): 155–166, http://doi.org/10.1093/jamia/ocaa242

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

[44] Sharafoddini A, Dubin JA, Maslove DM, et al. New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study. *JMIR Med Inform* 2019; 7(1): e11605. https://doi.org/10.2196/11605

[45] Sperrin M, Martin GP, Sisk R, et al. Missing Data Should be Handled Differently for Prediction Than for Description or Causal Explanation. *J Clin Epidemiol* 2020; 125: 183-187. http://doi.org/10.1016/j.jclinepi.2020.03.028.

[46] Fu Z, Parikh CR, Zhou B. Penalized Variable Selection in Competing Risks Regression. *Lifetime Data Anal* 2017; 23(3): 353-376. https://doi.org/10.1007/s10985-016-9362-3

[47] Wang L, Chen G, Li H. Group SCAD Regression Analysis for Microarray Time Course Gene Expression Data. *Bioinformatics* 2007; 23(12): 1486–1494. http://doi.org/10.1093/bioinformatics/btm125.

[48] R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.r-project.org/

[49] Romero-Brufau S, Huddleston JM, Escobar GJ, et al. Why the C-Statistic is Not Informative to Evaluate Early Warning Scores and What Metrics to Use. *Critical Care* 2015; 19(1): 1–6. http://doi.org/10.1186/s13054-015-0999-1

[50] Castaldi PJ, Dahabreh IJ, Ioannidis JPA. An Empirical Assessment of Validation Practices for Molecular Classifiers. *Brief Bioinform* 2011; 12(3): 189–202. http://doi.org/10.1093/bib/bbq073

[51] Ozenne B, Subtil F, Maucort-Boulch D. The Precision–Recall Curve Overcame the Optimism of the Receiver Operating Characteristic Curve in Rare Diseases. *J Clin Epidemiol* 2015; 68(8): 855-859. https://doi.org/10.1016/j.jclinepi.2015.02.010

[52] Steyerberg EW. Clinical Prediction Models: A Practical Approach To Development, Validation and Updating. Springer, Cham, 2019. https://doi.org/10.1007/978-0-387-77244-8

[53] Wolbers M, Koller MT, Witteman J, et al. Prognostic Models with Competing Risks: Methods and Application to Coronary Risk Prediction. *Epidemiology* 2009; 20(4): 555–561. http://doi.org/10.1097/EDE.0b013e3181a39056.

[56] Romero-Ortuno R, Wallis S, Biram R, et al. Clinical Frailty Adds to Acute Illness Severity in Predicting Mortality in Hospitalized Older Adults: An Observational Study. *Eur J Intern Med* 2016; 35(10): 24–34. http://doi.org/10.1016/j.ejim.2016.08.033

[54] Hewitt J, Carter B, Vilches-Moraga A, et al. The Effect of Frailty on Survival in Patients with COVID-19 (COPE): A Multicentre, European, Observational Cohort Study. *Lancet Public Health* 2020; 5(8): e444–e451. http://doi.org/10.1016/S2468-2667(20)30146-8.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

[55] Garibaldi BT, Fiksel J, Muschelli J, et al. Patient Trajectories Among Persons Hospitalized for COVID-19 : A Cohort Study. *Ann Intern Med* 2021; 174(1): 33–41. http://doi.org/10.7326/M20-3905.

[57] Laguna-Goya R, Utrero-Rico A, Talayero P, et al. IL-6–Based Mortality Risk Model for Hospitalized Patients with COVID-19. *J Allergy Clin Immunol* 2020; 146(4): 799–807.e9. http://doi.org/10.1016/j.jaci.2020.07.009.

[58] Wu C, Chen X, Cai Y, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* 2020; 180(7): 934-943. http://doi.org/10.1001/jamainternmed.2020.0994.

[59] Ma L, Zhang H, Yin Y, et al. Role of Interleukin-6 to Differentiate Sepsis from Non-Infectious Systemic Inflammatory Response Syndrome. *Cytokine* 2016; 88: 126–135. https://doi.org/10.1016/j.cyto.2016.08.033

[60] Chen T, Wu D, Chen H, et al. Clinical Characteristics of 113 Deceased Patients with Coronavirus Disease 2019: Retrospective Study. *Br Med J* 2020; 368: m1091. https://doi.org/10.1136/bmj.m1091

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Acknowledgements

Martin Wiegand was funded by the NIHR Cambridge Biomedical Research Centre [BRC-1215-20014]. Victoria L. Keevil was funded by the MRC/NIHR Clinical Academic Research Partnership Grant (CARP) [grant code MR/T023902/1]. Vince Taylor was funded by the Cancer Research UK Cambridge Centre. Effrossyni Gkrania-Klotsas was supported by the NIHR Clinical Research Network (CRN) Greenshoots Award. Brian D. M. Tom and Robert J. B. Goudie were funded by the UKRI Medical Research Council (MRC) [programme code MC_UU_00002/2] and supported by the NIHR Cambridge Biomedical Research Centre [BRC-1215-20014].

We thank Christopher Osuafor, Catriona Davidson, Alistair Mackett, Marie Goujon and Lelane Van Der Poel for assisting with the frailty sub-study.

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the MRC, or the Department of Health and Social Care.

Author contributions

Conception: JP, RJBG, SLC, EGK, CW and MW conceived the research project and designed the study.

Data collection: VT, MW, SLC and RJBG extracted and curated the dataset. VLK and DJH collected data.

Analysis tools & Data interpretation: MW, RJBG and BDMT analysed and contributed to the selection and creation of analysis tools. JP, CSW, EGK, DJH, VLK and BDMT interpreted the data and results.

Implementation: MW and RJBG performed the implementation and analysis of the proposed method.

Draft writing: MW, CSW, EGK, SLC and RJBG wrote the initial draft. All authors contributed to substantively revising the article for important intellectual content.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

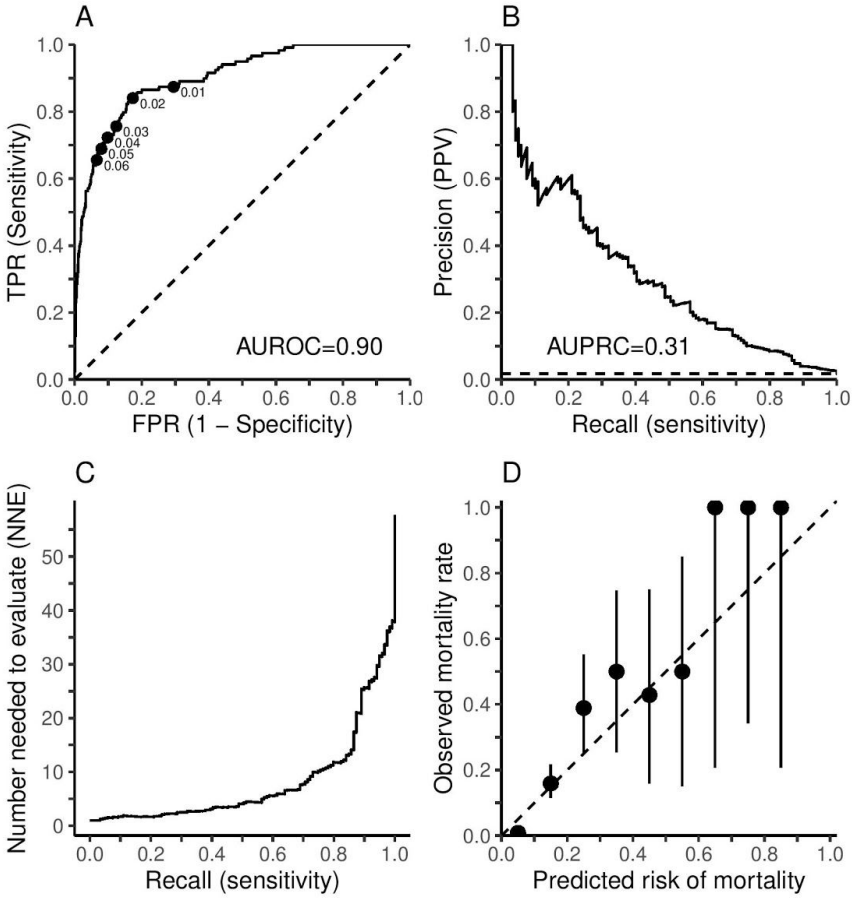
Approval of final submission

All authors approved of the final version of this script to be published, and are accountable for the work presented.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

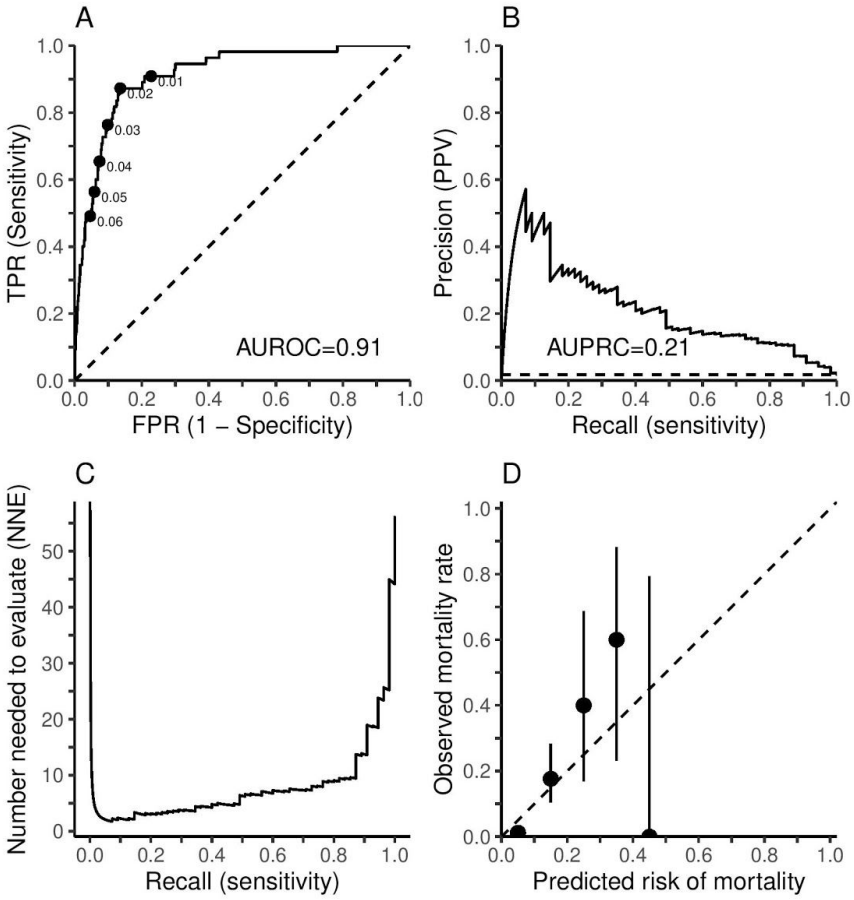
Figures

Figure 1. Performance metrics for in-hospital mortality in the training dataset. (A) Receiver operator characteristic plot, with labels indicating the corresponding threshold and the dashed line indicating the line of no discrimination. (B) Precision-recall plot, with the 1.7% observed incidence indicated by the dashed line. (C) Number needed to evaluate against sensitivity. (D) Calibration plot (with 95% CI), by tenths of predicted risk, with the dashed line indicating perfect calibration.



medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Figure 2. Performance metrics for in-hospital mortality in the test dataset. (A) Receiver operator characteristic plot, with labels indicating the corresponding threshold and the dashed line indicating the line of no discrimination. (B) Precision-recall plot, with the 1.8% observed incidence indicated by the dashed line. (C) Number needed to evaluate against sensitivity. (D) Calibration plot (with 95% CI), by tenths of predicted risk, with the dashed line indicating perfect calibration.



medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Tables

Table 1: Cohort demographics and clinical features in Wave 1 and Wave 2.

Characteristic	**Wave 1 (training data set)**	**Wave 2 (test data set)**
Admission dates	Mar 1, 2020 – Sep 12, 2020	Sep 13, 2020 – Jan 1, 2021
Number of patients	473	405
Female, n (%)	196 (41.4%)	203 (50.1%)
Age at admission, median [IQR], years	69 [55, 81]	61 [47, 77]
Admission BMI, median [IQR], kg/m [a]	25.7 [22.1, 29.9]	27.4 [23.4, 32]
Clinical Frailty Score at admission for over 65 year olds, median [IQR]	5 [3, 6]	5 [4,6]
Nosocomial infection, n (%)	36 (7.6%)	37 (9.2%)
Length of stay, median [IQR], days	10.8 [3.9, 19.8]	3.8 [1.0,12.1] [b]
Ethnicity, number (%)		
White	354 (74.8%)	255 (63.3%)
Asian	24 (5.1%)	26 (6.5%)
Black	10 (2.1%)	6 (1.5%)
Other	11 (2.3%)	23 (5.7%)
Prefer not to say/ Not recorded	74 (15.6%)	93 (23.1%)
Outcomes, number (%)		
Deceased in-hospital [c]	99 (20.9%)	41 (10.1%)
Transferred for ECMO	5 (1.1%)	0 (0%)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Discharged alive	369 (78.0%)	277 (68.0%)
Remain in hospital on Jan 3, 2021	0 (0%)	87 (21.4%)
Support / treatments received during hospital stay, number (%)		
ICU admission	103 (21.8%)	69 (17.0%)
Invasive mechanical ventilation	82 (17.3%)	48 (11.9%)
Cardiovascular support	86 (18.2%)	47 (11.6%)
Renal replacement therapy	32 (6.8%)	10 (2.5%)
Steroids	125 (26.4%)	185 (45.7%)
Admission observations [a], median [IQR]		
Mean arterial pressure, mmHg	86 [76, 96]	87 [79, 98]
Heart rate, beats/min	83 [72, 93]	82 [72, 92]
Temperature, degrees celsius	37.2 [36.7, 37.8]	37 [36.6, 37.4]
Respiratory rate, breaths/min	19 [17, 22]	18 [17, 20]
Oxygen saturation (SpO2), %	96 [94, 97]	96 [94, 98]
SpO2/FiO2 ratio	448 [337, 457]	452 [346, 462]
Admission blood results [a], median [IQR]		
Urea, mmol/L	7.8 [5.0, 11.8]	6.3 [5.0, 10.1]
Creatinine, μmol/L	75 [61, 106]	70 [55, 93]
Sodium, mmol/L	137 [134, 140]	137 [134, 139]
CRP, mg/L	87 [39, 178]	53 [24, 94]
WCC, 10^9/L	6.9 [4.9, 9.3]	6.3 [4.7, 9.5]

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Neutrophils, 10^9/L	5.4 [3.6, 7.8]	4.8 [3.3, 7.8]
Lymphocytes, 10^9/L	0.8 [0.5, 1.2]	0.8 [0.5, 1.1]
D-dimer, ng/ml	334 [177, 682]	275 [168, 661]
Troponin, ng/L	20.0 [8.3, 63.6]	11.0 [4.0, 47.0]
pH	7.43 [7.37, 7.46]	7.44 [7.40, 7.46]
IL-6, pg/ml	15.1 [5.3, 40.2]	17.0 [5.2, 30.2]

[a] First after positive SARS-CoV-2 test result for hospital-acquired COVID-19 cases

[b] For Wave 2, includes stays completed by Jan 3, 2021 only. Note that because follow-up continued until January 3 2021, the outcome during the 48-hour prediction horizon was known for all landmark times up to January 1 2021, and so patients remaining in hospital at the study end-date could be included in the validation.

[c] Total in-hospital mortality including those patients who were classified as “end of life” prior to death.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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. Final model coefficients

Predictor	**Coefficients**	
	When recorded	**If unrecorded**
Age <75 years, at admission	-0.516	–
Age <80 years, at admission	-0.245	–
Clinical Frailty Score, at admission	0.0678	0.0514
Heart rate, beats/min, mean during last 24h	0.00282	–
Respiratory rate, breaths/min, minimum during last 24h	0.102	–
S/F ratio, minimum during last 24h	-0.0116	–
WCC, 10^9/L, most recent measurement during last 48h	0.00651	-0.00151
Acidosis, 7.35 - (lowest pH during last 24h), or 0 if all above 7.35	3.18	0.220
IL-6, pg/ml, most recent measurement during last 48h	0.000166	-0.0218

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Supplementary Information

Table S1. Complete list of candidate predictors, summary statistics and missingness for the development dataset.

Marker	Unit	Summary measure/coding	Summary across landmark times [a]	Missingness across landmark times [b]
Demographics				
Age at admission	Years		69 [55,81]	– (0%)
		<45	517(7.6%)	–
		<50	853 (12.5%)	–
		<55	1386 (20.2%)	–
		<60	2164 (31.6%)	–
		<65	2779 (40.6%)	–
		<70	3371 (49.2%)	–
		<75	4284 (62.6%)	–
		<80	5236 (76.5%)	–
		<85	5857 (85.6%)	–
		<90	6315 (92.3%)	–
		<95	6670 (97.4%)	–
Sex	Female/ Male		41.4% / 58.6%	– (0%)
Ethnicity		Not included		
White	yes/no	White British/ White Irish/ Other white background	5215 (76.2%)	– (16.3%)
Asian	yes/no	Asian Indian/ Asian Pakistani/ Asian Bangladeshi/ Other Chinese/ Other Asian background	225 (3.3%)	– (16.3%)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Black	yes/no	Black Carribean/ Black African/ Other Black background	155 (2.3%)	– (16.3%)
Other	yes/no	Other ethnic group/ Mixed white and black carribean/ Mixed white and black African/ Mixed White and Asian/ Other mixed background	137 (2.0%)	– (16.3%)
Body Mass Index	kg/m^2	Not included	27.3 [22.5, 30.3]	– (6.1%)
Underweight	kg/m^2	18.5 - (Most recent BMI), or 0 if most recent BMI above 18.5	93.5% = 0 After excluding zero: 1.4 [0.5, 1.9]	– (6.1%)
Overweight	kg/m^2	(Most recent BMI) - 25, or 0 if most recent BMI below 25	38% = 0 After excluding zero: 4.4 [2.3, 9.1]	– (6.1%)
Clinical Frailty Scale [c]		Value	5 [3, 6]	– (54.5%)
Comorbidities				
Asthma	yes/no	Documented history of	780 (11.4%)	-
Dementia	yes/no	Documented history of	353 (5.2%)	-
Diabetes	yes/no	Documented history of	1233 (18.0%)	-
Chronic heart	yes/no	Documented history of	1260 (18.4%)	-

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

disease				
Hypertension	yes/no	Documented history of	2193 (32.0%)	-
Immunocompromised	yes/no	Documented history of	80 (1.1%)	-
Chronic liver disease	yes/no	Documented history of	640 (9.4%)	-
Non-haematological malignancy	yes/no	Documented history of	576 (8.4%)	-
Haematological malignancy	yes/no	Documented history of	284 (4.1%)	-
Chronic kidney disease	yes/no	Documented history of	502 (7.3%)	-
Respiratory disease	yes/no	Documented history of	833 (12.2%)	-
Stroke	yes/no	Documented history of	220 (3.2%)	-
Observations				
Heart rate (HR)	Beats/min	24h mean	83 [73, 93]	17.3 (0.0%)
		24h min	72 [63, 82]	
		24h max	94 [83, 106]	
		Trend	0 [-4.5, 4]	– (6.3%)
Mean arterial pressure	mmHg	24h mean	86 [79, 94]	17.1 (0.0%)
		24h min	74 [66, 83]	
		24h max	100 [91, 109]	
		Trend	0.0 [-0.4, 0.4]	– (6.3%)
Temperature	Degrees	24h mean	37.0 [36.7, 37.3]	9.3 (0.0%)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

	Celsius	24h min	36.4 [36.1, 36.7]	
		24h max	37.5 [37.1, 38.1]	
		Trend	0 [-0.3, 0.2]	– (6.3%)
Respiratory Rate (RR)	Breaths/min	24h mean	18.5 [17, 21]	18.8 (0.0%)
		24h min	16 [15, 18]	
		24h max	20 [19, 26]	
		Trend	0 [-1,1]	– (6.3%)
S/F ratio		24h mean	431 [325, 456]	12.8 (0.0%)
		24h min	392 [250, 448]	
		24h max	457 [443, 467]	
		Trend	0 [-7.0, 7.4]	– (6.3%)
P/F ratio	mmHg	24h mean	184 [136, 250]	1.5 (77.9%)
		24h min	140 [98, 201]	
		24h max	229 [171, 310]	
		Trend	2 [-18, 24]	– (80.3%)
Glasgow coma scale (GCS)	Lowest GCS in the last 24h	<9	33.8%	–
		<12	47.0%	–
Laboratory tests				
Urea	mmol/L	Most recent measurement during last 48h	8.8 [5.6, 14.1]	0.9 (38.0%)
Creatinine	μmol/L	Most recent measurement during last 48h	70 [52, 106]	0.9 (25.8%)
		Trend	-1 [-7, 4]	– (43.4%)
Sodium	mmol/L	Not included	138.6 [135.5, 142]	2.7 (24.8%
		Trend	0.0 [-1.0, 1.5]	– (41.9%)
Hyponatraemia	Na < 135 mmol/L	135 - (lowest sodium during last 24h), or 0 if all above 135	82% = 0 After excluding zero: 2 [1, 4]	2.7 (24.8%)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Hypernatraemia	Na > 145 mmol/L	(highest sodium during last 24h) - 145, or 0 if all below 145	85.2% = 0 After excluding zero: 3.8 [2, 6.2]	
Potassium	mmol/L	Most recent measurement during last 24h	4.1 [3.7, 4.4]	2.7 (25.0%)
		Trend	0 [-0.2, 0.2]	– (42.1%)
Albumin	g/L	Most recent measurement during last 48h	24 [20, 28]	1.2 (33.6%)
Alanine Transaminase (ALT)	U/L	Most recent measurement during last 48h	36 [22, 61]	0.7 (37.2%)
Alkaline phosphatase (ALP)	U/L	Most recent measurement during last 48h	100 [73, 149]	2.3 (19.8%)
Bilirubin	μmol/L	Most recent measurement during last 48h	8 [5, 13]	0.7 (37.4%)
Lactate dehydrogenase (LDH)	U/L	Most recent measurement during last 48h	335 [261, 436]	0.2 (80.5%)
C-reactive protein (CRP)	mg/L	Most recent measurement during last 48h	56 [22, 131]	1.6 (12.8%)
		Trend	0 [-0.2, 0.3]	– (45.9%)
Procalcitonin (PCT)	ng/ml	Most recent measurement during last 48h	0.24 [0.08, 0.83]	0.2 (84.4%)
Ferritin	μg/L	Most recent	726 [336, 1427]	0.3 (76.6%)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

		measurement during last 48h		
Haemoglobin	g/L	Most recent measurement during last 48h	104 [89, 122]	5.4 (12.4%)
		Trend	-1 [-5, 3]	– (43.9%)
White cell count (WCC)	10^9/L	Most recent measurement during last 48h	7.9 [5.7, 10.6]	1.6 (12.7%)
		Trend	0 [-1, 1]	– (46.1%)
Neutrophils	10^9/L	Most recent measurement during last 48h	5.7 [3.9, 8.2]	1.6 (13.4%)
Lymphocytes	10^9/L	Most recent measurement during last 48h	1.1 [0.7, 1.5]	1.6 (13.4%)
Neutrophil-Lymphocytes ratio	Ratio	Most recent measurement during last 48h	5.4 [3.2, 9.5]	1.6 (13.5%)
Eosinophils	10^9/L	Most recent measurement during last 48h	0.1 [0.02, 0.28]	1.6 (14.1%)
Monocytes	10^9/L	Most recent measurement during last 48h	0.45 [0.3, 0.64]	1.6 (13.5%)
Platelets	10^9/L	Most recent measurement during last 48h	280 [193, 387]	1.6 (12.8%)
		Trend	3 [-17, 26]	– (46.3%)
Red cell distribution width (RDW)	%	Most recent measurement during last 48h	15.1 [14, 16.4]	1.6 (13.3%)
Prothrombin Time	sec	Most recent	13.3 [12.5, 14.5]	0.84 (52.6%)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

		measurement during last 48h		
Activated partial thromboplastin time (APTT)	sec	Most recent measurement during last 48h	32.3 [29.6, 35.3]	0.84 (53.8%)
D-Dimer	ng/ml	Most recent measurement during last 48h	552 [284, 1677]	0.4 (71.9%)
Troponin	ng/L	Most recent measurement during last 48h	17 [5.6, 48.7]	0.3 (80.5%)
Interferon Gamma (IG)	pg/ml	Most recent measurement during last 48h	0.9 [0.9, 2.5]	0.2 (84.2%)
TNF-Alpha (TNFA)	pg/ml	Most recent measurement during last 48h	12.3 [8.3, 18.2]	0.2 (84.2%)
Interleukin-1 beta (IL-1)	pg/ml	Most recent measurement during last 48h	0.5 [0.3, 0.9]	0.2 (84.2%)
Interleukin-6 (IL-6)	pg/ml	Most recent measurement during last 48h	13.6 [4.7, 31.9]	0.3 (84.0%)
Interleukin-10 (IL-10)	pg/ml	Most recent measurement during last 48h	1.88 [0.7, 4.4]	0.2 (84.2%)
Interleukin ratio (IL-ratio, IL6/ IL10)	Ratio	Most recent measurement during last 48h	7.6 [2.9, 20.1]	0.2 (84.3%)
Lactate	mmol/L	Most recent measurement during last 48h	1.3 [1.0, 1.7]	3.7 (60.4%)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

pH – arterial or (venous + 0.03)		Not included	7.41 [7.36, 7.44]	1.9 (66.4%)
Acidosis	pH-value < 7.35	7.35 - (lowest pH during last 24h), or 0 if all above 7.35	67.1% = 0 After excluding zero: 0.06 [0.028,0.107]	1.9 (66.4%)
Alkalosis	pH-value > 7.45	(Highest pH during last 24h) - 7.45, or 0 if all below 7.45	66.8%= 0 After excluding zero: 0.024 [0.011, 0.039]	
Treatments, interventions and level of care				
Visited ICU	yes/no	During last 24h	1789 (26.1%)	–
Mechanically ventilated	yes/no	During last 24h	1420 (20.7%)	–
Cardiovascular support	yes/no	During last 24h	705 (10.3%)	–
Renal replacement therapy	yes/no	During last 24h	348 (5.1%)	–
Steroids (oral or intravenous; dexamethasone, hydrocortisone, prednisolone, methylprednisolone)	yes/no	Ever during this hospital visit up to now	2770 (40.5%)	–

[a] For yes/no items shown as number (%) across landmark times; for quantitative items shown as median [IQR] across landmark times

[b] Shown as mean number of measurements per landmark (% landmarks with no measurement)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

[c] For 45 patients for whom no CFS score had been recorded by the treating team, a consultant or specialist registrar in Geriatric Medicine reviewed the clinical records and assigned a CFS score using only information recorded at the time of admission.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 S2. ICD-10 codes used to identify comorbidities.

Diagnosis	**ICD-10 codes**	**Description**
Hypertension	I10	Essential hypertension
	I11	Hypertensive heart disease
	I12	Hypertensive renal disease
	I13	Hypertensive heart and renal disease
	I15	Secondary hypertension
Diabetes	E10	Type 1 diabetes mellitus
	E11	Type 2 diabetes mellitus
	E12	Malnutrition-related diabetes mellitus
	E13	Other specified diabetes mellitus
	E14	Other unspecified diabetes mellitus
Chronic liver disease	K70	Alcoholic liver disease
	K71	Toxic liver disease
	K72	Hepatic failure, not elsewhere classified
	K73	Chronic hepatitis, not elsewhere classified
	K74	Fibrosis and cirrhosis of the liver
	K75	Other inflammatory diseases of the liver
	K76	Other diseases of the liver
	K77	Liver disorders in disease classified elsewhere
Asthma	J45	Asthma
Non-haematological malignancy	C0	Malignant neoplasm of lip
	C1	Malignant neoplasm of base of tongue
	C2	Malignant neoplasm of other unspecified parts of tongue
	C3	Malignant neoplasm of gum
	C4	Malignant neoplasm of floor of mouth
	C5	Malignant neoplasm of palate
	C6	Malignant neoplasm of other and unspecified parts of mouth
	C7	Malignant neoplasm of parotid gland

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

Haematological malignancy	C8	Malignant neoplasm of other and unspecified major salivary glands
	C9	Malignant neoplasm of tonsil
Stroke	I63	Cerebral infarction
	I65	Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
	I66	Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
Chronic kidney disease	N18.1-N18.5	Chronic kidney disease stage 1-5
	N18.9	Chronic kidney disease, unspecified
	I13	Hypertensive and renal disease
Chronic heart disease	I20	Angina pectoris
	I21	Acute myocardial infarction
	I22	Subsequent myocardial infarction
	I23	Certain current complications following acute myocardial infarction
	I24	Other acute ischaemic heart diseases
	I25	Chronic ischaemic heart disease
	I34	Nonrheumatic mitral valve disorders
	I35	Nonrheumatic aortic valve disorders
	I36	Nonrheumatic tricuspid valve disorders
	I37	Pulmonary valve disorders
	I42	Cardiomyopathy
	I43	Cardiomyopathy in diseases classified elsewhere
	I44	Atrioventricular and left bundle-branch block
	I50	Heart failure
Immunocompromised	D80	Immunodeficiency with predominantly antibody defects
	D81	Combined immunodeficiencies
	D82	Immunodeficiency associated with other major defects
	D83	Common variable immunodeficiency
	D84	Other immunodeficiencies
Dementia	F01	Vascular dementia
	F02	Dementia in other diseases classified elsewhere

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 .

	F03	Unspecified dementia
	G30, G31	Alzheimer disease & Other degenerative diseases of nervous system, not elsewhere classified
	F10.27	Alcohol dependence, with alcohol-induced persisting dementia
	F10.97	Alcohol use, unspecified with alcohol-induced persisting dementia
	F19.97	Other psychoactive substance use, unspecified with psychoactive substance-induced persisting dementia
Respiratory disease	I27	Other pulmonary heart diseases
	J6*-J7*	Lung diseases due to external agents
	J41	Simple and mucopurulent chronic bronchitis
	J42	Unspecified chronic bronchitis
	J43	Emphysema
	J44	Other chronic obstructive pulmonary disease
	J47	Bronchiectasis

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 S3: Baseline cumulative subdistribution hazards for mortality in the final model

Time after landmark (hours)	Cumulative subdistribution hazard	Time after landmark (hours)	Cumulative subdistribution hazard	Time after landmark (hours)	Cumulative subdistribution hazard	Time after landmark (hours)	Cumulative subdistribution hazard
1	0.00094	13	0.00852	25	0.02143	37	0.03339
2	0.00161	14	0.00906	26	0.02339	38	0.03480
3	0.00255	15	0.00942	27	0.02463	39	0.03559
4	0.00304	16	0.00963	28	0.02525	40	0.03582
5	0.00363	17	0.01109	29	0.02673	41	0.03707
6	0.00427	18	0.01140	30	0.02783	42	0.03778
7	0.00458	19	0.01220	31	0.02876	43	0.03867
8	0.00516	20	0.01434	32	0.03008	44	0.04101
9	0.00563	21	0.01563	33	0.03095	45	0.04200
10	0.00713	22	0.01674	34	0.03182	46	0.04322
11	0.00787	23	0.01821	35	0.03265	47	0.04481
12	0.00834	24	0.01962	36	0.03318	48	0.04625

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 S4. Model coefficients for the alternative model with IL-6 replaced by CRP and re-calculating the model coefficients through the same penalised likelihood function used in the SCAD algorithm.

Predictor	**Coefficients**	
	When recorded	**If unrecorded**
Age <75 years, at admission	-0.115	–
Age <80 years, at admission	-0.0582	–
Clinical Frailty Score, at admission	0.0672	0.150
Heart rate, beats/min, mean during last 24h	0.0128	–
Respiratory rate, breaths/min, minimum during last 24h	0.0515	–
S/F ratio, minimum during last 24h	-0.00346	–
WCC, 10^9/L, most recent measurement during last 48h	0.00239	-0.116
Acidosis, 7.35 - (lowest pH during last 24h), or 0 if all above 7.35	2.73	0.474
C-reactive protein, pg/ml, most recent measurement during last 48h	-0.0000350	0.220

medRxiv preprint doi: https://doi.org/10.1101/2021.02.15.21251150; this version posted February 18, 2021. 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 S5: Baseline cumulative subdistribution hazards for the alternative model with IL-6 replaced by CRP

Time after landmark (hours)	Cumulative subdistribution hazard	Time after landmark (hours)	Cumulative subdistribution hazard	Time after landmark (hours)	Cumulative subdistribution hazard	Time after landmark (hours)	Cumulative subdistribution hazard
1	0.00010	13	0.00091	25	0.00224	37	0.00342
2	0.00017	14	0.00097	26	0.00243	38	0.00355
3	0.00028	15	0.00101	27	0.00255	39	0.00363
4	0.00033	16	0.00103	28	0.00262	40	0.00365
5	0.00039	17	0.00118	29	0.00276	41	0.00378
6	0.00046	18	0.00122	30	0.00287	42	0.00384
7	0.00049	19	0.00130	31	0.00296	43	0.00393
8	0.00056	20	0.00152	32	0.00309	44	0.00416
9	0.00061	21	0.00165	33	0.00318	45	0.00425
10	0.00077	22	0.00176	34	0.00326	46	0.00437
11	0.00084	23	0.00191	35	0.00334	47	0.00452
12	0.00089	24	0.00205	36	0.00340	48	0.00466