BMJ Open

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

Martin Wiegand ,[1,2] Sarah L Cowan ,[3] Claire S Waddington ,[4] David J Halsall,[3] Victoria L Keevil ,[4,5] Brian D M Tom ,[2] Vince Taylor,[6] Effrossyni Gkrania-Klotsas ,[7] Jacobus Preller ,[3] Robert J B Goudie [2]

To cite: Wiegand M, Cowan SL, Waddington CS, *et al*. Development and validation of a dynamic 48-hour in-hospital mortality risk stratification for COVID-19 in a UK teaching hospital: a retrospective cohort study. *BMJ Open* 2022;**12**:e060026. doi:10.1136/bmjopen-2021-060026

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-060026).

Received 28 April 2021
Accepted 13 July 2022

For numbered affiliations see end of article.

Correspondence to
Dr Martin Wiegand;
Martin.Wiegand@mrc-bsu.cam.ac.uk

ABSTRACT

Objectives To develop a disease stratification model for COVID-19 that updates according to changes in a patient's condition while in hospital to facilitate patient management and resource allocation.

Design In this retrospective cohort study, we adopted a landmarking approach to dynamic prediction of all-cause in-hospital mortality over the next 48 hours. We accounted for informative predictor missingness and selected predictors using penalised regression.

Setting All data used in this study were obtained from a single UK teaching hospital.

Participants We developed the model using 473 consecutive patients with COVID-19 presenting to a UK hospital between 1 March 2020 and 12 September 2020; and temporally validated using data on 1119 patients presenting between 13 September 2020 and 17 March 2021.

Primary and secondary outcome measures The primary outcome is all-cause in-hospital mortality within 48 hours of the prediction time. We accounted for the competing risks of discharge from hospital alive and transfer to a tertiary intensive care unit for extracorporeal membrane oxygenation.

Results Our final model includes age, Clinical Frailty Scale score, heart rate, respiratory rate, oxygen saturation/fractional inspired oxygen ratio, white cell count, presence of acidosis (pH <7.35) and interleukin-6. Internal validation achieved an area under the receiver operating characteristic (AUROC) of 0.90 (95% CI 0.87 to 0.93) and temporal validation gave an AUROC of 0.86 (95% CI 0.83 to 0.88).

Conclusions Our model incorporates both static risk factors (eg, age) and evolving clinical and laboratory data, to provide a dynamic risk prediction model that adapts to both sudden and gradual changes in an individual patient's clinical condition. On successful external validation, the model has the potential to be a powerful clinical risk assessment tool.

Trial registration The study is registered as 'researchregistry5464' on the Research Registry (www.researchregistry.com).

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Our dynamic prediction model can incorporate patient data as it accumulates throughout a hospital visit.
⇒ We use the established statistical landmarking approach to dynamic prediction; we account for competing risks for the primary outcome of in-hospital mortality, and the potentially informative availability of clinical and laboratory data.
⇒ The sample size of the first wave of patients admitted with severe COVID-19 was relatively low, due to the lower incidence in Cambridgeshire, but increased significantly during the winter months of 2020/2021, providing the opportunity to temporally validate the model.
⇒ As a single-centre study, the presented model will require external validation to assess its performance in other cohorts; and also if there are significant changes in the characteristics of new variants or the management thereof.
⇒ Our work also highlights the adaptability of the statistical landmarking framework to be used to model individual patient outcomes using frequently collected hospital data.

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. Such point-of-admission models have been proposed for both mortality and composite escalation/mortality outcomes, including new and repurposed severity and early warning scores[1–7] and time-to-event models.[8–13]

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, for example, thromboembolism.[14 15] This may have a significant effect on a patient's prognosis that cannot be foreseen by a point-of-admission model.

A model with the ability to adjust predictions at arbitrary time points by including updated patient information could greatly aid in clinical decision-making. 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 point-of-admission models. Predictive models that incorporate post-admission information are limited in number and scope. Some models for predicting mortality or deterioration have used information after admission, but do not continue beyond the first few days of admission.[16 17] More recent time-varying Cox models (for mortality and escalation)[18 19] and machine learning models (for mortality)[20] have used additional post-admission data. However, time-varying Cox models should not be used for prediction, because they require knowledge of clinical information from the future to calculate the hazard function, which is impossible in practice.[21] Furthermore, 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 for patients thought more likely to benefit from escalation in care, such as transfer to an intensive care unit (ICU). While often ignored, such effects can be strong in electronic health record (EHR) data.[22 23]

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. Our model is based on a principled statistical approach called landmarking[21 24 25] that allows inclusion of any time-varying clinical parameters recorded prior to the time of prediction, while appropriately accounting for censoring and changes in the set of patients at risk. The model 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. Therefore, discharge is a 'competing risk'[26] when viewed from the perspective of in-hospital mortality. We account for competing risks within the landmarking framework using a recently proposed approach[27] that has not previously been used to model individual patient outcomes using frequently collected EHR data.

MATERIALS AND 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 1 March 2020 and 17 March 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, which do not include extracorporeal membrane oxygenation (ECMO).

We report our findings according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis reporting guidelines.[28]

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 (online supplemental appendix 1). Patients with clinically diagnosed COVID-19 (based on symptoms, and the clinical opinion of the treating clinician) were included because diagnostic testing was limited during the early stages of the pandemic.[29]

We include only the first hospital visit for each patient involving (or subsequent to) their first positive test; any readmissions were excluded. Nosocomial infection was defined as a first positive SARS-CoV-2 test or diagnosis more than 10 days after hospital admission. Since we first train our model 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 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 (online supplemental eappendix 2).

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 considered two competing risks: transfer to a tertiary ICU for ECMO; and discharge from the hospital

due to clinical improvement. Patients were followed up until 19 March 2021.

Patient and public involvement

No patients were involved in the design of this study.

Model development

A list of 59 candidate clinical parameters (online supplemental etable 1) were chosen based on existing point-of-admission prediction models for COVID-19; components of scores predicting mortality in critical illness more generally, such as the Sequential Organ Failure Assessment[30] and the Acute Physiology and Chronic Health Evaluation II[31]; and clinical opinion as to other likely predictors. These are divided into five categories: demographics, comorbidities, observations, laboratory tests, and treatments, interventions and level of care.

Basic patient demographics were extracted from the hospital EHR: sex, ethnicity and age at hospital presentation; and deviation from standard ranges of body mass index, using the most recent measurement (up to 1 year old).

Twelve comorbidities that have previously been associated with COVID-19 severity[32] 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). Online supplemental etable 2 provides the ICD-10 codes used to define each comorbidity. In addition to specific comorbidities, frailty among patients over 65 years old was assessed by the Clinical Frailty Scale (CFS) score[33] (online supplemental eappendix 3).

We included the following observations that are regularly recorded in the EHR: heart rate (HR), mean arterial pressure, temperature and respiratory rate (RR). Oxygen saturation (SpO2)/fractional inspired oxygen (FiO2) ratio was calculated (where SpO2 and FiO2 were available at the same time point) to indicate the severity of hypoxia.[34 35] 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 (eg, patients with chronic obstructive pulmonary disease being assigned a lower SpO2 target). Where only oxygen flow rate was available, FiO2 was estimated according to the EPIC II conversion tables.[36] PaO2/FiO2 (P/F) ratio was also included. We summarised the measurements recorded over the previous 24-hour period as follows: mean, minimum and maximum value. We also calculated the 'median-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.

For each of the 31 laboratory tests, we considered (online supplemental eTable 1), 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 seven of the most frequently measured blood tests (C reactive protein (CRP), white cell count (WCC), platelets, haemoglobin, creatinine, sodium, potassium), we included the median trend. 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 17 37] For blood markers where both abnormally low and abnormally high results could potentially be associated with poor prognosis (sodium and pH), we included the maximum deviation below and above the normal range in the previous 24 hours. We adjusted venous pH results by adding 0.03 to approximate arterial pH results.[38]

We included five 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. Steroid administration has been shown to reduce 28-day mortality in patients with COVID-19.[39 40] We therefore include an indicator of whether the patient had received treatment dose steroids (defined as 6 mg dexamethasone daily or an equivalent dose of prednisolone, hydrocortisone or methylprednisolone) during their hospital admission prior to the time the prediction is made.

Models

We used a landmarking approach,[27] which has been proposed for dynamic prediction.[21 24 25] At intervals of 24 hours (referred to as the 'landmark times'), we trained time-to-event models, using clinical parameters recorded before (or at) the landmark time as predictors. This makes it possible to include repeatedly measured clinical parameters into the prediction model, so that predictors reflect any changes in the trajectory of the patient, while appropriately accounting for censoring and changes in the at-risk population. If the primary outcome was recorded within the prediction horizon of a landmark time, we recorded the outcome at the relative time from landmark to event; events after the prediction horizon were censored. Patients who have had any event prior to the landmark time were excluded, since these patients were no longer at risk. The first landmark is 6 hours after presentation to allow time for clinical information to accrue, or at the point of COVID-19 diagnosis for nosocomial patients (diagnosed 10 days or more after presentation). We only used data at each landmark time from patients being actively treated for COVID-19 at that point in time: landmark times after transition to end of life care were omitted, meaning that no predictions were made at these timepoints, although events

occurring within the existing prediction horizon were still included. Patients receiving end of life care were identified from their recorded escalation status in the EHR and by manual review of patient notes containing relevant keywords (online supplemental eappendix 4). We used a supermodel approach in which the time-to-event model is assumed constant across landmark times.[39]

Specifically, we used a competing-risks landmarking approach[27 41] which uses a Fine-Gray competing risk model to predict in-hospital death, and account for the competing risks of hospital discharge and transfer for ECMO.[42 43] A Fine-Gray model uses subdistribution hazards, which are directly related to the cumulative incidence function, by which the probability of an event of interest occurring can be estimated. While the cause-specific hazard function used by Cox models is preferable for inferring biological mechanisms, subdistribution hazard-based models are preferable for prediction.[44]

Missing values

We handled missing data using the missingness indicator approach since 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 condition.[45 46] Conceptually our approach, as well as estimating the 'effect' of a unit increase of a particular clinical parameter (as is standard in all regression approaches), estimates the 'effect' of a variable being 'missing'. A clinical parameter may be 'missing' (also referred to as 'not recorded' or 'not available') due to not being measured or due to not being documented in the EHR. For each potential predictor in the model we also include a missingness indicator, which indicates that no data were recorded during the corresponding time period. This approach allows clinical parameters with an incomplete record to be included in our model and avoids the need to make the missing at random (MAR) assumption that is unlikely to hold in these data.[47] For each parameter, ultimately one of the two expressions is used for prediction for each patient at each timepoint: either a coefficient describing the relationship with the clinical parameter when it is recorded, or a fixed value if the clinical parameter is missing.

Blood tests are considered missing if the most recent measurement was collected more than 48 hours prior to the landmark time. When a blood test is repeated during the previous 48 hours, the most recent result is used. The vital signs we considered did not have missing values at any included landmark.

Predictor selection

To select the most predictive parameters for the model we adopted an approach proposed for Fine-Gray models[48] that uses standard penalised variable selection, specifically smoothly clipped absolute deviations (SCAD), with the tuning parameter chosen to minimise the Bayesian information criterion. We paired parameters together with their corresponding missingness indicator to prevent inclusion of an incompletely recorded parameter without its missingness indicator, using the group SCAD[49]; but also allowed for the missingness indicators to be included by themselves.

The development and validation of the model has been carried out in R V.3.6.[50]

Model assessment

Quantitative assessment of discrimination was performed using the area under the receiver operating characteristic (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 fivefold (split into 80% training, 20% validation data) cross-validation to account for uncertainty and over-optimism due to the complete model building process (including variable selection).[51] We also calculated the precision-recall (PR) curve and 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.[52] We assessed clinical benefit visually via the number needed to evaluate (NNE), defined as 1/positive predictive value (1/PPV), against the sensitivity. We also calculated the net benefit curve.[53]

We assessed calibration visually using a calibration plot of predicted risk against observed mortality rate. We also quantitatively assessed the calibration slope and calibration-in-the-large.[54]

Sensitivity analyses

To assess whether the model is unduly influenced by patients with long hospitalisations we retrained the model using only each patient's first 28 landmark times (spanning 28 days). We also assessed the sensitivity of our model assessment by stratifying by whether COVID-19 was confirmed by a positive SARS-CoV-2 diagnostic test, and according to whether patients had received at least a single COVID-19 vaccination dose prior to admission (Oxford-AstraZeneca Covishield, Pfizer-BionTech Comirnaty or Moderna Spikevax). We also assessed the discrimination of the model if we removed all laboratory tests. Finally, we explored the implications of replacing the 24-hour prediction horizon with a 72-hour prediction horizon.

RESULTS

Development of prediction model

We developed the model using data from wave 1 (1 March 2020 to 12 September 2020), with the end date chosen since only a single patient remained in hospital with COVID-19 on this date. A total of 519 patients presented to hospital with COVID-19 during wave 1, 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 (ie, within 6 hours of presentation). The

Table 1 Cohort demographics and clinical features in wave 1 and wave 2

Characteristic	Wave 1 (training data set)	Wave 2 (validation data set)
Admission dates	1 March 2020–12 September 2020	13 September 2020–17 March 2021
No of patients	473	1119
Female, n (%)	196 (41.4)	579 (48.3)
Age at admission, median (IQR), years	69 (55–81)	65 (49–79)
Admission BMI, median (IQR), kg/m^2 *	25.7 (22.1–29.9)	27.35 (23.5–32.1)
Clinical Frailty Score at admission for over 65 year olds, median (IQR)	5 (3–6)	5 (4–6)
Nosocomial infection, n (%)	36 (7.6)	116 (9.7)
Length of stay, median (IQR), days	10.8 (3.9–19.8)	8.0 (2.95–18.5)†
Ethnicity, n (%)		
White	354 (74.8)	792 (66.1)
Asian	24 (5.1)	61 (5.1)
Black	10 (2.1)	13 (1.1)
Other	11 (2.3)	39 (3.3)
Prefer not to say/not recorded	74 (15.6)	294 (24.5)
Outcomes, n (%)		
Deceased in-hospital‡	99 (20.9)	145 (12.1)
Transferred for ECMO	5 (1.1)	1 (0.0)
Discharged alive	369 (78.0)	967 (80.7)
Remain in hospital on 19 March 2021	0 (0)	86 (7.2)
Support/treatments received during hospital stay, n (%)		
ICU admission	103 (21.8)	238 (19.8)
Invasive mechanical ventilation	82 (17.3)	172 (14.3)
Non-invasive ventilation (CPAP or Bi-PAP)	37 (7.8)	144 (12.9)
Cardiovascular support	86 (18.2)	170 (14.2)
Renal replacement therapy	32 (6.8)	39 (3.3)
Steroid treatment	120 (25.4)	619 (55.3)
Admission observations*, median (IQR)		
Mean arterial pressure, mm Hg	86 (76–96)	88 (78–98)
Heart rate, beats/min	83 (72–93)	81 (72–91)
Temperature, °C	37.2 (36.7, 37.8)	36.9 (36.5, 37.4)
Respiratory rate, breaths/min	19 (17, 22)	18 (17, 20)
SpO2, %	96 (94, 97)	95 (94, 97)
SpO2/FiO2 ratio	448 (337, 457)	452 (400, 462)
Admission blood results*, median (IQR)		
Urea, mmol/L	7.8 (5.0–11.8)	7.1 (5.1–10.5)
Creatinine, µmol/L	75 (61–106)	70 (56–93)
Sodium, mmol/L	137 (134–140)	137 (135–140)
CRP, mg/L	87 (39–178)	55 (26–109)
WCC, 10^9 /L	6.9 (4.9–9.3)	6.8 (5.0–9.8)
Neutrophils, 10^9 /L	5.4 (3.6–7.8)	5.3 (3.5–7.9)
Lymphocytes, 10^9 /L	0.8 (0.5–1.2)	0.8 (0.5–1.2)
D-dimer, ng/mL	334 (177–682)	292 (167–641)
Troponin, ng/L	20.0 (8.3–63.6)	11.0 (4.0–39.5)
pH	7.43 (7.37–7.46)	7.44 (7.39–7.46)
IL-6, pg/mL	15.1 (5.3–40.2)	11.6 (4.0–29.8)

*First after positive SARS-CoV-2 test result for hospital-acquired COVID-19 cases.
†For wave 2, includes stays completed by 19 March 2021 only. Note that because follow-up continued until 19 March 2021, the outcome during the 48-hour prediction horizon was known for all landmark times up to 17 March 2021, and so patients remaining in hospital at the study end-date could be included in the validation.
‡Total in-hospital mortality including those patients who were classified as 'end of life' prior to death.
Bi-PAP, Bi-level positive airway pressure; BMI, body mass index; CPAP, Continuous positive airway pressure; CRP, C reactive protein; ECMO, extracorporeal membrane oxygenation; FiO2, fractional inspired oxygen; ICU, intensive care unit; IL-6, interleukin-6; SpO2, oxygen saturation; WCC, white cell count.

characteristics of the 473 patients included in the development of the model are shown in table 1.

In total, we included 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 hours and the prediction horizon is 48 hours, patient events will usually occur within the prediction horizon of two adjacent landmark times. Online supplemental etable 1 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 (table 2 and online supplemental etable 3) for 48-hour in-hospital mortality includes age, CFS score,[55] HR, RR, SpO2/FiO2 ratio,[34 35] WCC, acidosis (pH <7.35) and interleukin-6 (IL-6). The mortality probability can be calculated using the calculator at http://shiny.mrc-bsu.cam.ac.uk/apps/covid19mortalityrisk/; see online supplemental eappendix 5 for details.

Table 2 Final model coefficients

Predictor	Coefficients when recorded	Coefficients when not recorded*
Age <75 years, at admission	−0.516	–
Age <80 years, at admission	−0.245	–
Clinical Frailty Score, at admission	0.0678	0.0513
Heart rate, beats/min, mean during last 24 hours	0.00282	–
Respiratory rate, breaths/min, minimum during last 24 hours	0.102	–
SpO2/FiO2 ratio, minimum during last 24 hours	−0.0116	–
WCC, 10^9/L, most recent measurement during last 48 hours	0.00651	−0.0015
Acidosis, 7.35 − (lowest pH during last 24 hours), or 0 if all above 7.35	3.18	0.22
IL-6, pg/mL, most recent measurement during last 48 hours	0.000166	−0.0218

*If the predictor value is not recorded (due to not being measured or documented in the EHR within the relevant time window), the fixed value in this column is used, and the coefficient corresponding to the predictor value is ignored.
EHR, electronic health record; SpO2/FiO2, oxygen saturation/fractional inspired oxygen; WCC, white cell count.

Internal performance assessment

Figure 1 shows the internal performance metrics for the model in table 2 (using the training data). The unadjusted internal area under receiver operating characteristic curve (AUROC) was 0.90 (95% CI 0.87 to 0.93) and the median cross-validation AUROC was 0.87, both indicating good discrimination (figure 1A). The PR curve (figure 1B) also showed good discrimination, with an AUPRC of 0.31 in a population with 48 hour in-hospital mortality of 0.017 (1.7%), and we observe that the NNE <10 for sensitivity less than 0.75 (figure 1C). Figure 1D shows the calibration plot. The calibration intercept was −0.02 (95% CI −0.22 to 0.17) indicating that the mean predicted probabilities matched the mean observed mortality, while the calibration slope was 1.16 (95% CI 1.02 to 1.31) suggesting that the observed mortality in high predicted risk patients slightly exceeded the predicted mortality risk, although the CIs are wide. The net benefit curve for risk stratification by the proposed model is clearly higher than for the two non-model alternatives of classifying either everyone or no-one as high risk patients, indicating the clinical utility of the dynamic model (online supplemental efigure 1).

Temporal validation of prediction model

We assessed the performance of the model by applying it to held-out data corresponding to admissions during wave 2 (13 September 2020 and 17 March 2021). A total of 1119 patients presented to the study hospital during this period. In total, we tested the model using 12 981 landmark times, with a median of 6 (2–14) landmark times per patient. In the 48-hour prediction horizon following these landmark times, there were 253 in-hospital death events (1.9% of landmarks), 1615 hospital discharge events (12.4%) and 2 transfers for ECMO (0.015%). Forty-seven landmark times were omitted due to missing vital sign data. Characteristics are summarised in table 1. Of note, compared with wave 1, patients presenting in wave 2 were slightly younger and more likely to be female, evidenced by a more balanced data set.

Figure 2 shows the temporal validation performance metrics, obtained by applying the trained model (table 2) to the wave 2 patients. The receiver operating characteristic (ROC) curve (figure 2A) shows the model continues to discriminate well, with AUROC 0.86 (95% CI 0.83 to 0.88). The PR curve (figure 2B) shows that PPV was consistently well above the 48-hour in-hospital mortality incidence of 0.019 (1.9%) in the wave 2 cohort across all sensitivities. The AUPRC 0.15 is reduced and the NNE is higher than in the training data, although NNE <10 for sensitivities between 0.02 and 0.63 (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 0.35 (95% CI 0.21 to 0.47), suggesting the mean of the predicted probabilities was lower than the mean observed mortality, and calibration slope was 0.90 (95% CI 0.82 to 0.99), indicating that the spread of predicted risk was significantly smaller than the spread of observed mortality. The calibration plot shows a good correspondence between observed mortality rate and predicted mortality risk for the lower-risk patients (<0.4), but due to the low incidence of mortality events among the landmarks corresponds less well for the higher-risk patients. This is evidenced by the considerable CIs. The net benefit curves for the proposed model surpasses both alternatives of classifying everyone and no-one as high-risk patients (online supplemental efigure 2).

Sensitivity analyses

The model trained on only data from the first 28 landmarks for each patient closely resembled the model in table 2, although the Clinical Frailty Score and IL-6 were not selected (online supplemental etable 4).

We did not find evidence that the discrimination of the model was affected by the presence or absence of positive diagnostic SARS-CoV-2 results (rather than solely a clinical diagnosis). In patients with a positive SARS-CoV-2 test the AUROC was 0.90 (95% CI 0.87 to 0.93) in the training dataset and 0.85 (95% CI 0.83 to 0.88) in the validation dataset; whereas for patients with only a clinical diagnosis the AUROC was 0.94 (95% CI 0.88 to 1.00) in the training dataset and 0.88 (95% CI 0.79 to 0.99) in the validation dataset.

A small number of patients (65 patients, for whom 874 landmarks were available) in our validation dataset had received a COVID-19 vaccine at the time of analysis: the AUROC of 0.88 (95% CI 0.71 to 0.99) for these patients suggested good discrimination and is consistent with the

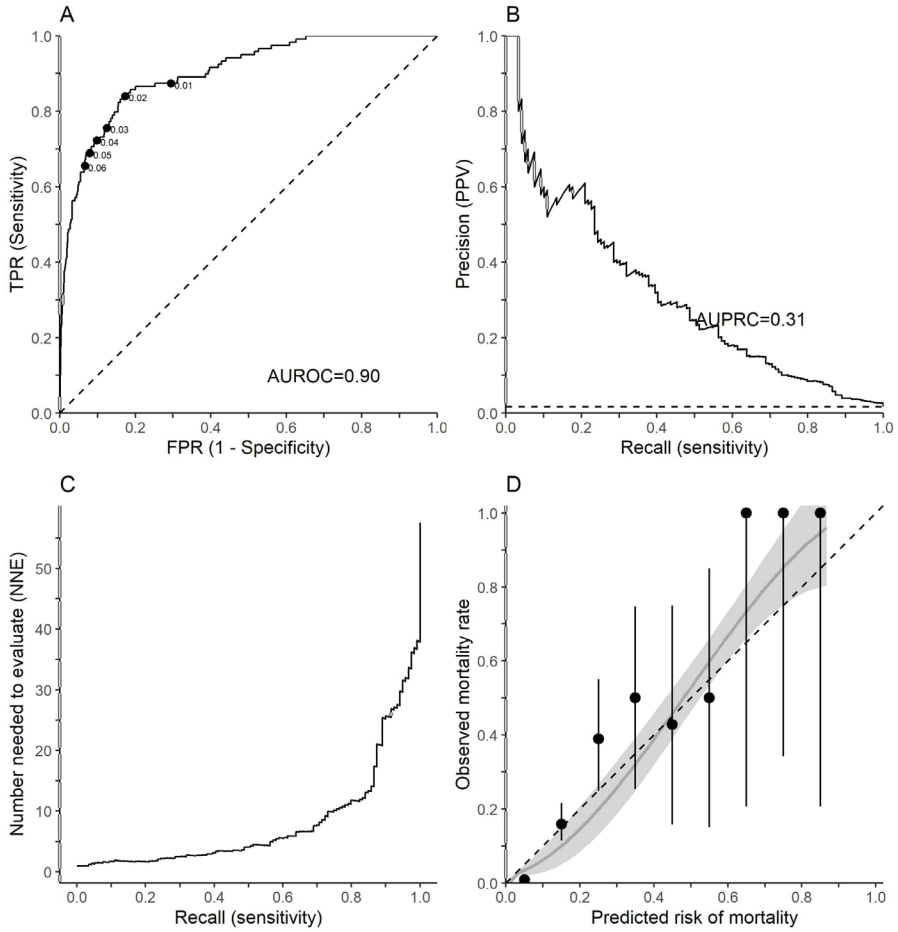


Figure 1 Performance metrics for in-hospital mortality in the training dataset. (A) Receiver operating characteristic plot, with labels indicating the corresponding threshold and the dashed line indicating the line of no discrimination. (B) Precision-recall plot, with the 2.8% observed incidence indicated by the dashed line. (C) NNE against sensitivity. (D) Calibration plot (with 95% CI), by tenths of predicted risk and a LOESS interpolation (grey), with the dashed line indicating perfect calibration. AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristic; FPR, false positive rate; LOESS, locally estimated scatterplot smoothing; NNE, number needed to evaluate; PPV, positive predictive value; TPR, true positive rate.

unvaccinated patients (AUROC 0.85 (95% CI 0.83 to 0.88)).

To assess whether a simpler model could perform similarly we considered removing the laboratory tests from our model (online supplemental etables 5 and 6). The resulting model provided slightly worse discrimination with an AUROC of 0.88 (95% CI 0.85 to 0.90) in the training data and an AUROC of 0.85 (95% CI 0.82 to 0.87) in the validation data.

To assess whether 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 (online supplemental etables 7 and 8). The AUROC was slightly lower on both training (0.89, 95% CI 0.85 to 0.93) and validation (0.84, 95% CI 0.81 to 0.87) data.

To assess the feasibility of our approach with an extended time horizon and the stability of the predictors selected, we refitted the model with a 72-hour prediction horizon. The resulting model matched the previous AUROC in the training data of 0.90 and the AUROC of 0.85 in the validation data set. The full performance metrics are available in online supplemental figures 3 and 4, and the model coefficients in online supplemental etable 9.

DISCUSSION

SARS-CoV-2 causes a wide spectrum of disease that can evolve over time, and may necessitate critical care management and even result in death. In light of the threat of further waves of coronavirus infections there is still a pressing clinical need to be able to anticipate disease severity and the trajectory of illness to facilitate patient management and resource allocation. The model described herein incorporates both static admission 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. The data used in the model were routinely collected demographic and clinical data from during the patient's hospitalisation, automatically extracted from patient EHRs. As such, this

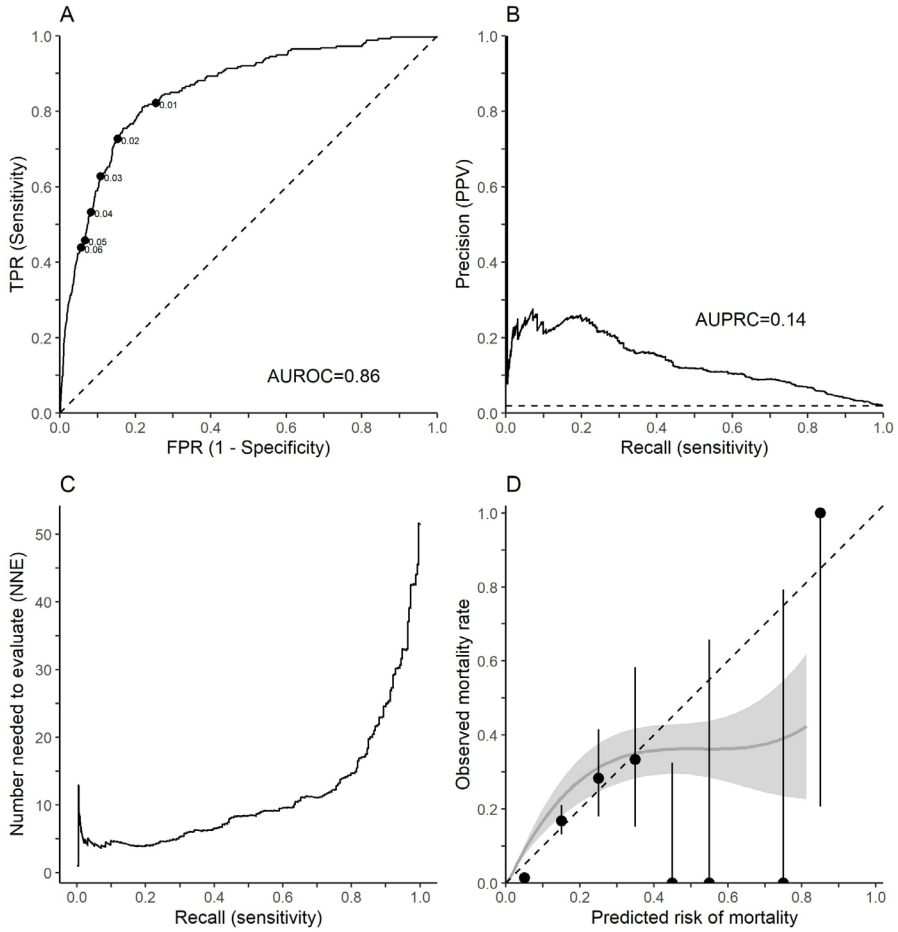


Figure 2 Performance metrics for in-hospital mortality in the validation dataset. (A) Receiver operating characteristic plot, with labels indicating the corresponding threshold and the dashed line indicating the line of no discrimination. (B) Precision-recall plot, with the 3.1% observed incidence indicated by the dashed line. (C) NNE against sensitivity. (D) Calibration plot (with 95% CI), by tenths of predicted risk and a LOESS interpolation (grey), with the dashed line indicating perfect calibration. AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristic; FPR, false positive rate; LOESS, locally estimated scatterplot smoothing; NNE, number needed to evaluate; PPV, positive predictive value; TPR, true positive rate.

model could be readily incorporated into routine clinical care.

An extensive literature of risk stratification models for COVID-19 has developed since the start of the pandemic for a variety of endpoints and time horizons, although many have been described as unreliable in a large systematic review[56] and remain unvalidated. Most of the literature considers point-of-admission scores, including the 4C Mortality score,[57] which is based on demographic information, comorbidities and blood tests taken on admission. This score performed well in both large development and validation cohorts (AUROC 0.77 and 0.79, respectively) and has now been externally validated in several countries.[58–64] Similarly the point-of-admission ISARIC 4C Deterioration score[3] has been externally validated.[58 60 65 66] Promising machine-learning alternatives have also been recently proposed,[67 68] although these have not yet been independently validated in contrast to the 4C scores. Another approach has been to repurpose existing early warning scores, particularly the National Early Warning Score (NEWS) 2 score.[69] The discrimination of NEWS2 for predicting 14-day deterioration (a composite outcome of transfer to ICU or death) has been described as moderate[1 64] (AUROC 0.70), whereas for 24-hour clinical deterioration (initiation of ventilatory support or death) it has been described as reasonable (AUROC 0.78).[64] The addition of age and additional physiological parameters and blood tests led to an improvement in discrimination (AUROC 0.735 and 0.78).[3] NEWS2 is a core component of escalation pathways in most UK hospitals, with a national recommendation that patients with NEWS2 score of 7 or above should be reviewed by the intensive care team. The fact that NEWS2 can, to a degree, predict intensive care admission is therefore unsurprising.

The literature for risk prediction using dynamic, post-admission in-hospital data for patients with COVID-19 is much smaller. Some models have used only data from the first few days of admission.[16 17 70] Others have used additional postadmission data in time-varying Cox models (for mortality and escalation),[18 19] which require knowledge of clinical information from the future and so should

not be used for prediction.[21] Machine-learning models for mortality[20] and respiratory support[71] have also been reported to have good discrimination, although missing data are either discarded or imputed, which ignores the problem of informative missingness and may limit their applicability on an individual patient level. They also do not account for competing risks, for example, hospital discharge. While this paper has been under review, a promising random forest-based dynamic model predicting the need for respiratory support (encompassing high flow nasal oxygen, and both invasive and non-invasive ventilation) and death has been proposed.[72] This model reports high internal discrimination (AUROC 0.89, 95% CI 0.88 to 0.90) although PR is not reported. The most important predictors identified in this model were the SpO2/FiO2 ratio and the RR, both of which also feature in our model. Both these observations play an important role in the decision to escalate the level of respiratory support, and so it is not clear whether for the 1 day horizon this model offers much in prediction that is not already apparent clinically (predicting the need for additional respiratory support after that decision has already been made serves little purpose). In addition, it is again unclear how the model handles missing data. Markov models have also been proposed that seek to model the trajectory of patients rather than focus on prediction as we do in this paper.[73 74]

Our approach has several methodological strengths which we believe sets it apart from previous models. First, we accounted for competing risks, whereby the outcome (risk) of interest (in this case in-hospital mortality) can only happen while 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 is important in predictive modeling.[75] Second, we allowed for the possibility that the availability of observations and investigations may in itself reflect disease severity, as the fact that specific tests have been requested at a given time provides an insight into the treating clinicians' contemporaneous view of the patient's condition. While multiple imputation is often used in clinical prediction models because it gives unbiased estimates under the MAR assumption, it is unlikely that this assumption holds in the routinely collected EHR data that we use.[45] The missing indicator method that we adopted does not rely on the MAR assumption and can improve predictive performance in EHR data.[45–47] Furthermore, we validated our model using data from different waves. As each wave included COVID-19 variants of different infectiousness that are potentially associated with different morbidity and mortality risks, the fact that the model performs reasonably (NNE <10 for sensitivities between 0.02 and 0.63) across waves further attests to the fact that our selected parameters are useful for prognostication in different clinical scenarios. Finally, we did not seek to make prognostic predictions for patients after clinicians have identified them as entering the last few hours or days of life. Since observations and investigations are often discontinued at the end of life, including these time periods would distort the model due to extreme missingness (in our data, no vital sign observations were recorded on 43% of days during end-of-life care, compared with 0% of days during active treatment). In addition, predicting end-of-life after it is clinically apparent would have little clinical utility.

Several predictors of disease severity included in our model have also been identified by point-of-admission severity models, 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 32] Frailty has similarly been shown to be a strong independent predictor of mortality in hospitalised older adults,[76] including those with COVID-19.[55 77]

Respiratory compromise is a common reason for hospital admission and markers of respiratory function, including RR,[3 4 12] SpO2,[3] oxygen requirement[3] and SpO2/FiO2 ratio have been included in previous point-of-admission models. The SpO2/FiO2 ratio, as selected by our model, 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.

Acidosis frequently complicates respiratory, renal and advanced circulatory failure and has previously been noted as a marker of disease severity in COVID-19.[78] The separate inclusion of the severity of acidosis and alkalosis in our set of candidate predictors allowed for pH changes in either direction to be accounted for and avoided, for example, a minor negative effect of alkalosis from masking a more major effect of acidosis.

Our model selected two markers of infection and inflammation: WCC and IL-6. This is consistent with other findings.[11 79 80] 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. The results when refitting the model with CRP in place of IL-6 produced 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.[81]

There are several limitations to our study. We chose to include only laboratory results up to 48 hours and vital signs up to 24 hours before the landmark time; exploiting older data might improve the predictive ability of our model, at the expense of complexity and real-world utility. Our data were gathered from a single centre, and therefore, the generalisability of our findings to other centres and populations are pending external validation. 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 considered clinically appropriate rather than what resources permitted. Finally, while it is encouraging that the model continued to perform well in the wave 2 validation data, changes in clinical care of patients (notably use of steroids and IL-6 inhibitors) and the dominant virus strains (including the Delta and Omicron variants that emerged in the UK while this manuscript was under review) may influence the clinical picture of the disease, its severity and the risk factors for disease. The model will therefore likely need to be updated as the pandemic evolves, but the utilisation of routinely available data in this model makes this straightforward.

Author affiliations
[1]Faculty of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK
[2]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[3]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
[4]Department of Medicine, University of Cambridge, Cambridge, UK
[5]Department of Medicine for the Elderly, Addenbrooke's Hospital, Cambridge, UK
[6]Cancer Research UK, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
[7]Department of Infectious Diseases, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

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

Contributors Conception: JP, RJBG, SLC, EG-K, CSW 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, CW, EG-K, 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, EG-K, SLC and RJBG wrote the initial draft. All authors contributed to substantively revising the article for important intellectual content. All authors approved of the final version of this script to be published and are accountable for the work presented. MW is acting as the guarantor for the project.

Funding MW was funded by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). VLK was funded by the MRC/NIHR Clinical Academic Research Partnership Grant (CARP) (grant code MR/T023902/1). VT was funded by the Cancer Research UK Cambridge Centre. EG-K was supported by the NIHR Clinical Research Network (CRN) Greenshoots Award. BT and RJBG 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).

Disclaimer 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.

Competing interests None Declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval 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.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The deidentified 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 licence for the current study, and so are not publicly available. Data are however available from the authors on reasonable request with permission of Cambridge University Hospitals.

Author note JP and RJBG are last authors.

ORCID iDs
Martin Wiegand http://orcid.org/0000-0003-0276-658X
Sarah L Cowan http://orcid.org/0000-0001-9149-3530
Claire S Waddington http://orcid.org/0000-0002-4120-6027
Victoria L Keevil http://orcid.org/0000-0001-6148-0640
Brian D M Tom http://orcid.org/0000-0002-3335-9322
Effrossyni Gkrania-Klotsas http://orcid.org/0000-0002-0930-8330
Jacobus Preller http://orcid.org/0000-0001-5706-816X
Robert J B Goudie http://orcid.org/0000-0001-9554-1499

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.
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;56:2002113.
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;9:e592.
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;76:442–53.
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.
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;181:1423–33.
7 Shang Y, Liu T, Wei Y, *et al*. Scoring systems for predicting mortality for severe patients with COVID-19. *EClinicalMedicine* 2020;24:100426.
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:e2022058.
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;21:1–7.
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 Geriatr Soc* 2020;68:E19–23.
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.
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;9:1548.

13 Liang W, Yao J, Chen A, *et al*. Early triage of critically ill COVID-19 patients using deep learning. *Nat Commun* 2020;11:1–7.
14 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;191:145–7.
15 Poissy J, Goutay J, Caplan M, *et al*. Pulmonary embolism in patients with COVID-19. *Circulation* 2020;142:184–6.
16 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:e041983.
17 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.
18 Crooks CJ, West J, Fogarty A. 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*2020:20248181.
19 Schwab P, Mehrjou A, Parbhoo S, *et al*. Real-time prediction of COVID-19 related mortality using electronic health records. *Nat Commun* 2021;12:1058.
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 Big Data* 2021;7:38–44.
21 van Houwelingen HC, Putter H. *Dynamic prediction in clinical survival analysis*. Boca Raton: CRC Press, 2012.
22 Goldstein BA, Navar AM, Pencina MJ, *et al*. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 2017;24:198–208.
23 Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ* 2018;361:k1479.
24 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;18:86–92.
25 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;187:1530–8.
26 Wolkewitz M, Cooper BS, Bonten MJM, *et al*. Interpreting and comparing risks in the presence of competing events. *BMJ* 2014;349:g5060.
27 Wu C, Li L, Li R. Dynamic prediction of competing risk events using landmark sub-distribution hazard model with multiple longitudinal biomarkers. *Stat Methods Med Res* 2020;29:3179–91.
28 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55.
29 Iacobucci G. Covid-19: What is the UK's testing strategy? *BMJ* 2020;368:m1222.
30 Vincent J-L, Moreno R, Takala J, *et al*. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/ failure. *Intensive Care Med* 1996;22:707–10.
31 Knaus WA, Draper EA, Wagner DP, *et al*. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818–29.
32 Williamson EJ, Walker AJ, Bhaskaran K, *et al*. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020;584:430–6.
33 Rockwood K, Song X, MacKnight C, *et al*. A global clinical measure of fitness and frailty in elderly people. *CMAJ* 2005;173:489–95.
34 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;132:410–7.
35 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;37:1317–21.
36 Vincent J-L, Rello J, Marshall J, *et al*. International study of the prevalence and outcomes of infection in intensive care units. *JAMA* 2009;302:2323–9.
37 Lin H-A, Lin S-F, Chang H-W, *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:e0241262.
38 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;19:168–75.
39 The RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19 — preliminary report. *N Engl J Med* 2020:NEJMoa2021436.
40 WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JAC, Murthy S, *et al*. Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: a meta-analysis. *JAMA* 2020;324:1330–41.
41 The CRAN repository. Available: https://cran.r-project.org/web/ packages/crrp/index.html
42 van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scand J Stat* 2007;34:70–85.
43 Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999;94:496–509.
44 Ambrogi F, Biganzoli E, Boracchi P. Estimates of clinically useful measures in competing risks survival analysis. *Stat Med* 2008;27:6407–25.
45 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;28:155–66.
46 Sharafoddini A, Dubin JA, Maslove DM, *et al*. A new insight into missing data in intensive care unit patient profiles: observational study. *JMIR Med Inform* 2019;7:e11605.
47 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;125:183–7.
48 Fu Z, Parikh CR, Zhou B. Penalized variable selection in competing risks regression. *Lifetime Data Anal* 2017;23:353–76.
49 Wang L, Chen G, Li H. Group SCAD regression analysis for microarray time course gene expression data. *Bioinformatics* 2007;23:1486–94.
50 R Core Team. R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria, 2021. Available: http://www.r-project.org/
51 Castaldi PJ, Dahabreh IJ, Ioannidis JPA. An empirical assessment of validation practices for molecular classifiers. *Brief Bioinform* 2011;12:189–202.
52 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;68:855–9.
53 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019;3:3.
54 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation and updating*. Cham: Springer, 2019.
55 Marincowitz C, Turner V, Allgar V. Can patient frailty be estimated from inpatient records? A prospective cohort study. *Adv Geriatr Med Res* 2020;2:e200004.
56 Wynants L, Van Calster B, Collins GS, *et al*. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020;369:m1328.
57 Knight SR, Ho A, Pius R, *et al*. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC who clinical characterisation protocol: development and validation of the 4C mortality score. *BMJ* 2020;370:m3339.
58 Ong SWX, Sutjipto S, Lee PH, *et al*. Validation of ISARIC 4C mortality and deterioration scores in a mixed vaccination status cohort of hospitalized COVID-19 patients in Singapore. *Clin Infect Dis* 2022:ciac087.
59 Riley JM, Moeller PJ, Crawford AG. External validation of the COVID-19 4C mortality score in an urban United States cohort. *Am J Med Sci* 2022:30. doi:10.1016/j.amjms.2022.04.030
60 Knight SR, Gupta RK, Ho A, *et al*. Prospective validation of the 4C prognostic models for adults hospitalised with COVID-19 using the ISARIC who clinical characterisation protocol. *Thorax* 2022;77:606–15.
61 Yildiz H, Castanares-Zapatero D, Hannesse C, *et al*. Prospective validation and comparison of COVID-GRAM, NEWS2, 4C mortality score, CURB-65 for the prediction of critical illness in COVID-19 patients. *Infect Dis* 2021;53:640–2.
62 Jones A, Pitre T, Junek M, *et al*. External validation of the 4C mortality score among COVID-19 patients admitted to hospital in Ontario, Canada: a retrospective study. *Sci Rep* 2021;11:18638.
63 Kuroda S, Matsumoto S, Sano T, *et al*. External validation of the 4C mortality score for patients with COVID-19 and pre-existing cardiovascular diseases/risk factors. *BMJ Open* 2021;11:e052708.
64 Gupta RK, Marks M, Samuels THA, *et al*. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. *Eur Respir J* 2020;56:2003498. doi:10.1183/13993003.03498-2020
65 Cowan SL, Wiegand M, Preller J, *et al*. Validation of the 4C deterioration model for COVID-19 in a UK teaching hospital during wave 2. *Am J Med Sci* 2021;362:512–5.

66 Wirth A, Goetschi A, Held U, *et al*. External validation of the modified 4C deterioration model and 4C mortality score for COVID-19 patients in a Swiss tertiary hospital. *Diagnostics* 2022;12:1129.
67 Rasmy L, Nigo M, Kannadath BS, *et al*. Recurrent neural network models (CovRNN) for predicting outcomes of patients with COVID-19 on admission to hospital: model development and validation using electronic health record data. *Lancet Digit Health* 2022;4:e415–25.
68 Garcia-Gutiérrez S, Esteban-Aizpiri C, Lafuente I, *et al*. Machine learning-based model for prediction of clinical deterioration in hospitalized patients by COVID 19. *Sci Rep* 2022;12:7097.
69 Royal College of Physicians. *National early warning score (news) 2: standardising the assessment of acute-illness severity in the NHS*. London: RCP, 2017.
70 Kamran F, Tang S, Otles E, *et al*. Early identification of patients admitted to hospital for covid-19 at risk of clinical deterioration: model development and multisite external validation study. *BMJ* 2022;376:e068576.
71 Youssef A, Kouchaki S, Shamout F, *et al*. Development and validation of early warning score systems for COVID-19 patients. *Healthc Technol Lett* 2021;8:105–17.
72 Wongvibulsin S, Garibaldi BT, Antar AAR, *et al*. Development of severe COVID-19 adaptive risk predictor (SCARP), a calculator to predict severe disease or death in hospitalized patients with COVID-19. *Ann Intern Med* 2021;174:777–85.
73 Momeni-Boroujeni A, Mendoza R, Stopard IJ, *et al*. A dynamic Bayesian model for identifying high-mortality risk in hospitalized COVID-19 patients. *Infect Dis Rep* 2021;13:239–50.
74 Soper BC, Cadena J, Nguyen S, *et al*. Dynamic modeling of hospitalized COVID-19 patients reveals disease state-dependent risk factors. *J Am Med Inform Assoc* 2022;29:864–72.
75 Wolbers M, Koller MT, Witteman JCM, *et al*. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* 2009;20:555–61.
76 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:e444–51.
77 Garibaldi BT, Fiksel J, Muschelli J, *et al*. Patient trajectories among persons hospitalized for COVID-19: a cohort study. *Ann Intern Med* 2021;174:33–41.
78 Chen T, Wu D, Chen H, *et al*. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ* 2020;368:m1091. doi:10.1136/bmj.m1091
79 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;146:799–807.
80 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;180:934–43.
81 Ma L, Zhang H, Yin Y-L, *et al*. Role of interleukin-6 to differentiate sepsis from non-infectious systemic inflammatory response syndrome. *Cytokine* 2016;88:126–35.

Appendix for "Development and validation of a dynamic 48-hour in-hospital mortality risk stratification for COVID-19 in a UK teaching hospital: a retrospective cohort study" by Wiegand et al (2022)

eAppendix 1. 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[e1]. Clinical diagnosis of COVID-19 was identified using International Classification of Diseases 10th Edition (ICD-10) codes in the EHR.

eAppendix 2.

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).

eAppendix 3.

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[e2]. This approach has been shown to have good agreement with CFS scores assigned after face to face assessment (inter-rater reliability kappa 0.84)[e3].

eAppendix 4.

The following keywords were used to identify potentially relevant medical notes: end of life, end-of-life, supportive care, EOL anticipatory med, palliative, comfort care, end of his life, end of her life, terminal wean.

eAppendix 5.

Let $k_i \in \{0,\ldots,n_k\}$ be the event type of patient i, so that $T_{i,k}$ is patient i's corresponding event time.The probability of patient i with covariates $X(s)$ not experiencing any event before landmark s, and incurring event k within time $t \in [0,w]$of this landmark can therefore be calculated by:

$$\mathbb{P}(T_{i,k} < s + t \,|T_{i,k} \geq s, k_i = k, X(s)) = 1 - exp(-\int_{[0,t]} \tilde{\lambda}(u|X(s))du)$$

Here $\lambda(t|X(s))$is the sub-distribution hazard function for event k:

$$\lambda(t|X(s)) = \lambda_{k0}(t)\, exp(\beta^T X(s)),$$

with the subdistribution baseline hazard $\lambda_{k0}(t)$ for event k at time $t \in [0, w]$. We report estimated values for β in Table 2. We report in eTable 3 the estimated cumulative baseline sub-distribution hazard function $\int_{[0,t]} \lambda_{k0}(u)\, du$, and in eTable 6 and eTable 8 for the sensitivity analyses.

To derive the 48-hour mortality probability $(k_i = k_{mort},\ w = 48h)$ from the Fine-Gray model, we therefore compute the probability of the mortality occurring between the landmarking time *s* and the end of the prediction horizon *s*+*w*, given the patient did not have any event prior to the landmark and accounting for the covariates $X(s)$collected at the landmark time s.

$$\mathbb{P}(T_{i,k} < s + w | T_{i,k} \geq s, k_i = k_{mort}, X(s)) = 1 - exp(-\int_{[0,w]} \lambda(t|X(s))dt)$$

To allow for easier access a web-app is available on http://shiny.mrc-bsu.cam.ac.uk/apps/covid19mortalityrisk/ that calculates the 48-hour mortality probability based on user input.

eReferences

[e1] 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.

[e2] 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

[e3] 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

eTable 1. 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 presentation	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%)
Black	yes/no	Black Carribean/ Black African/	155 (2.3%)	– (16.3%)

		Other Black background		
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 (BMI)	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 disease	yes/no	Documented history of	1260 (18.4%)	-

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 Celsius	24h mean	37.0 [36.7, 37.3]	9.3 (0.0%)
		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%)
SpO2/FiO2 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%)

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 measurement	726 [336, 1427]	0.3 (76.6%)

		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 measurement	13.3 [12.5, 14.5]	0.84 (52.6%)

		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%)

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)

[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.

eTable 2. 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

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
	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

eTable 3: Baseline cumulative subdistribution hazards for mortality in the final model, as needed for calculation of the 48 hour survival probabilities (eAppendix 5).

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

eTable 4: Final model coefficients for landmarks less than 28 days from admission (or the first positive SARS-CoV-2 test if infection was nosocomial).

Predictor	**Coefficients when recorded**	**Coefficients if not recorded**[1]
Age <75 years, at admission	-0.269	–
Age <80 years, at admission	-0.135	–
Heart rate, beats/min, mean during last 24h	0.00154	–
Respiratory rate, breaths/min, minimum during last 24h	0.135	–
SpO2/FiO2 ratio, minimum during last 24h	-0.0114	–
WCC, 10^9/L, most recent measurement during last 48h	0.00169	-0.000359
Acidosis, 7.35 - (lowest pH during last 24h), or 0 if all above 7.35	5.97	1.66

[1] If the predictor value is not recorded (due to not being measured or documented in the EHR within the relevant time window), the fixed value in this column is used, and the coefficient corresponding to the predictor value is ignored.

eTable 5. Final model coefficients for model omitting all blood tests.

Predictor	**Coefficients when recorded**	**Coefficients if not recorded**[1]
Age <75 years, at admission	-0.683	–
Age <80 years, at admission	-0.195	–
Clinical Frailty Score, at admission	0.193	-0.170
Heart rate, beats/min, mean during last 24h	0.0175	–
Respiratory rate, breaths/min, minimum during last 24h	0.0570	–
SpO2/FiO2 ratio, minimum during last 24h	-0.0125	–

[1] If the predictor value is not recorded (due to not being measured or documented in the EHR within the relevant time window), the fixed value in this column is used, and the coefficient corresponding to the predictor value is ignored.

eTable 6: Baseline cumulative subdistribution hazards for the alternative model without blood tests, as needed for calculation of the 48 hour survival probabilities (eAppendix 5).

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	0.000687	13	0.006248	25	0.015666	37	0.024447
2	0.00118	14	0.006647	26	0.017095	38	0.02549
3	0.001872	15	0.006911	27	0.018004	39	0.026075
4	0.002229	16	0.007063	28	0.018459	40	0.026245
5	0.002662	17	0.008136	29	0.019544	41	0.027172
6	0.003129	18	0.008366	30	0.020346	42	0.027697
7	0.003357	19	0.008946	31	0.021037	43	0.028356
8	0.003784	20	0.010495	32	0.022007	44	0.030097
9	0.00413	21	0.011433	33	0.022648	45	0.030851
10	0.005232	22	0.012245	34	0.023292	46	0.031778
11	0.005773	23	0.013308	35	0.0239	47	0.032995
12	0.006119	24	0.014341	36	0.024293	48	0.034088

eTable 7. 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**	**Coefficients if not recorded[1]**
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	–
SpO2/FiO2 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

[1] If the predictor value is not recorded (due to not being measured or documented in the EHR within the relevant time window), the fixed value in this column is used, and the coefficient corresponding to the predictor value is ignored.

eTable 8: Baseline cumulative subdistribution hazards for the alternative model with IL-6 replaced by CRP, as needed for calculation of the 48 hour survival probabilities (eAppendix 5).

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

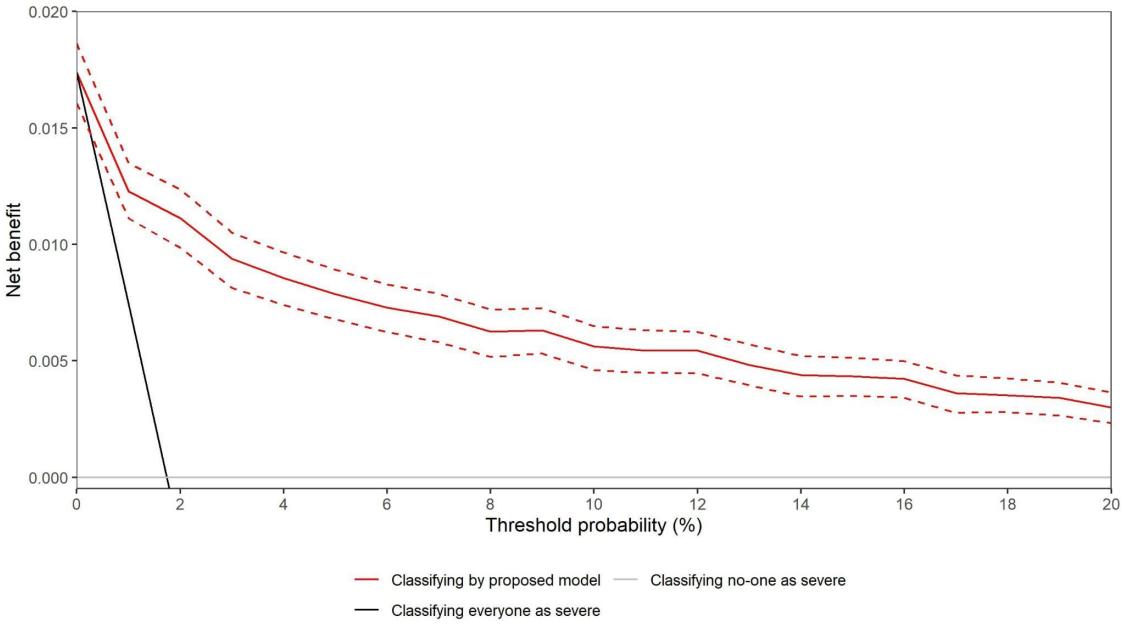
eTable 9. Final model coefficients for 72 hour prediction horizon

Predictor	**Coefficients when recorded**	**Coefficients if not recorded[1]**
Age <75 years, at admission	-0.564	–
Age <80 years, at admission	-0.183	–
History of non-haematological malignancy	0.146	–
Clinical Frailty Score, at admission	0.0721	0.0544
Respiratory rate, breaths/min, minimum during last 24h	0.100	–
SpO2/FiO2 ratio, minimum during last 24h	-0.0111	–
WCC, 10^9/L, most recent measurement during last 48h	0.00806	-0.0198[a]
Acidosis, 7.35 - (lowest pH during last 24h), or 0 if all above 7.35	1.29	0.0.0898
Platelets, 10^9/L, most recent measurement during last 48h	-0.000375	-0.0198[a]

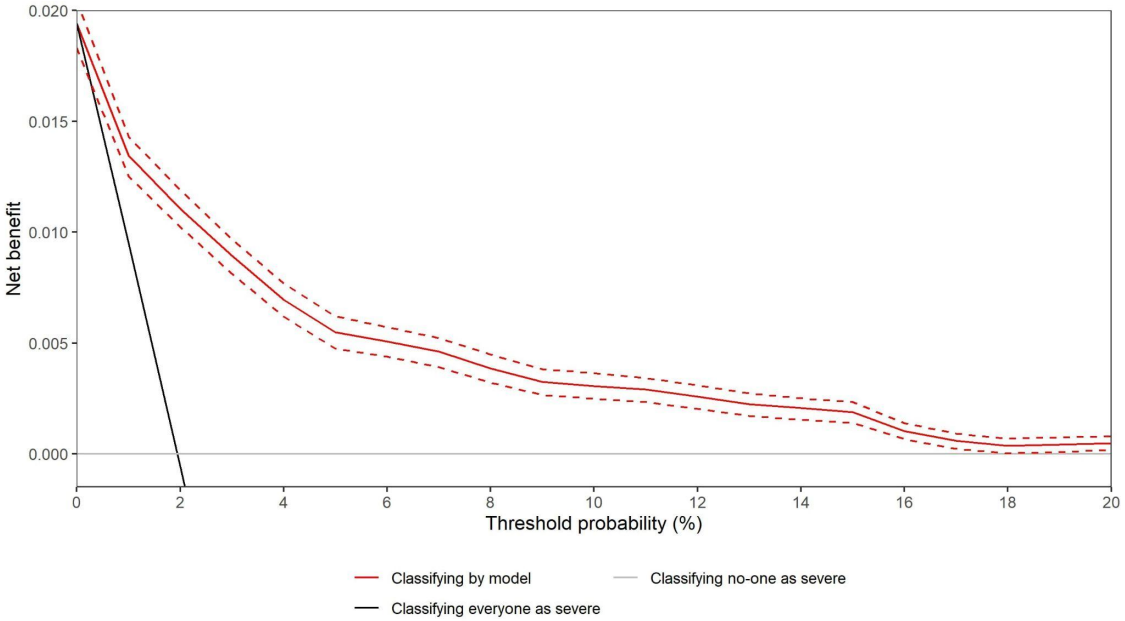
[a] WCC and Platelets are measured on the same sample, therefore the missingness for both clinical parameters is shared.

[1] If the predictor value is not recorded (due to not being measured or documented in the EHR within the relevant time window), the fixed value in this column is used, and the coefficient corresponding to the predictor value is ignored.

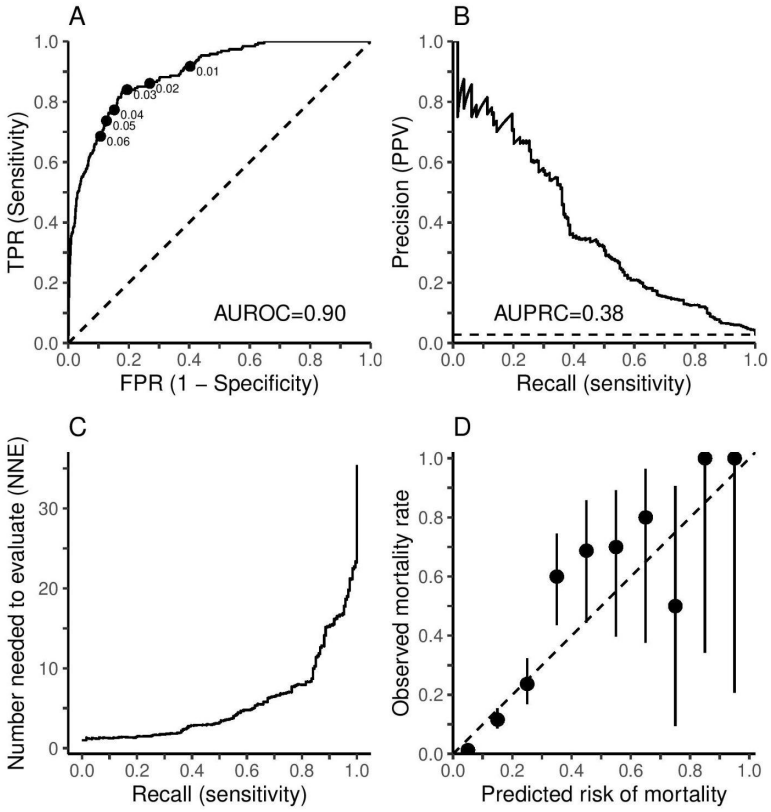
eFigure 1. Net benefit curve for risk stratification in the training dataset using the proposed model and by classifying either everyone or no-one as high risk patients. The dashed line shows 95% CI.



eFigure 2. Net benefit curve for risk stratification in the test dataset using the proposed model and by classifying either everyone or no-one as high risk patients. The dashed line shows 95% CI.



eFigure 3. Performance metrics for in-hospital mortality in the training dataset with 72 hour prediction horizon. (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 2.8% 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.



eFigure 4. Performance metrics for in-hospital mortality in the test dataset with 72 hour prediction horizon. (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 3.1% 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.