Supplementary methods

Sample size

We estimated we required at least 21 in-hospital deaths per parameter to provide consistent and reliable estimates from the prediction model (eTable 2).(1) To allow for flexible modelling of potential non-linear associations between physiological and laboratory variables and mortality, we assumed that the model would require expanding continuous variables, so that each continuous variable was represented by up to four parameters. Therefore, a model including up to 116 parameters (allowing for maximum degrees of freedom for non-linear predictors and all pre-specified interactions) would require a minimum of 2436 in-hospital deaths within 28 days to minimise model overfitting and provide sufficiently precise model predictions.

Handling of missing data

Ten imputed datasets were produced using fully conditional specification(2) under the assumption that data were missing at random conditional on the observed data. When required, logarithmic transformation for non-normality was used (eTable 3). All the potential predictors that were considered in the analysis (with or without missing values) were entered into the imputation model plus the outcome variable(3) and auxiliary variables of mechanical ventilation and sedation during the first 24 hours of critical care. Height and weight were imputed separately, and body mass index was determined from the actual or imputed data. Models were fitted to each of the ten imputed datasets and the results were combined using Rubin's rules.(4)

Model development

The predictor-outcome relationship for continuous predictors was explored by expanding each variable into multiple terms using restricted cubic splines with 3, 4 or 5 knots, plotting the relationships with a running line smoother as reference, and testing nonlinearity. Knot positions were selected following the recommendations of Harrell.(5) When the hypothesis of linearity was rejected, we chose the model that gave the best plausible fit.

The model building process consisted of the following stages:

Development of a parsimonious physiology model

After appropriate functional forms were selected in the univariable setting, a multivariable full physiology model was fitted. Tests were applied to test the global significance of the predictors and their linearity following a two-step procedure: testing the model on the pooled analysis (t-test) and independently to each of the 10 imputed data sets (Wald test). Considering a p-value < 0.1 as the selection criterion, a variable was dropped from the model if it was non-significant in the multiple imputation analysis and in all 10 individual analyses. The model was refitted, and the functional form was tested. The process continued until all the factors in the model were significant.

Development of a full main terms model

Starting from this simplified physiology model, a full multivariable model adding the rest of the potential predictors was fitted and the variable selection was done as described above. The model was refined by dropping predictors with non-significant global effects and/or data reduction (e.g. dropping nonlinear terms, combining categories) where appropriate. The process continued until all the predictors in the model were significant.

Modelling of severe conditions in the past medical history

Due to small sample sizes with individual conditions, two alternative approaches to modelling severe conditions in the past medical history were compared head-to-head: a single binary covariate for any severe condition in the past medical history; and separate binary covariates for immunocompromise and any other condition.

Model simplification

To get a more parsimonious model, each predictor was removed from the model one by one in turn and the c index(6) (equivalent to the area under the receiver operating characteristic curve) and Brier model accuracy score(7) (mean squared error between outcome and prediction) without that predictor were compared. Only the predictors that made an important contribution were retained in the model.

Interactions and re-evaluation of non-selected variables

Pre-specified interaction terms between ethnicity and BMI, and between ethnicity and glucose, were added, assessed, and retained if they made an important contribution. Predictors that were previously considered but found not to be significant in previous steps were also added back in to the model and reassessed.

A final process of deleting, refitting and verifying was performed to ensure that all included predictors were significant and were making an important contribution to the model.

Handling of calendar time

After the above steps, calendar time (days from 1 March 2020) was added to the model. The contribution of all predictors to the model was reassessed, as well as the predictors that had been previously considered but found not to be important, following the same approach described above.

Model validation

The prediction models were validated for their discrimination, calibration and overall fit. For the temporal validation of the model including calendar time, rather than extrapolating the linear time trend beyond the range of the development data, all predictions were based on a date of admission to critical care of 30 April 2020. The discrimination of the models was estimated by the c index and accuracy was assessed by Brier score and its associated R-squared measure. To assess optimistic performance within the development data, the percentage of overfitting was estimated by refitting each model in 100 bootstrap samples from the development dataset and evaluating each of these in the original development dataset to produce optimism-corrected statistics. We assessed calibration graphically with predicted outcome on the X-axis and observed outcome on the Y-axis in 10 equal-sized groups by predicted outcome, and by smoothing the individual patient data using locally weighted scatterplot smoothing (LOWESS) applied to the observed and predicted log odds of mortality. In the temporal validation data, calibration was additionally assessed by the ratio of observed to predicted mortality (calibration in the large) and by Cox's calibration regression (linear recalibration of the predicted log odds).(8)

eTable 1. Potential predictors

Predictor	Definition	Parameters
Calendar time	Number of days from 1 March 2020.	1
Age	The age of the patient in whole years at admission to critical care.	1-4
Sex	The genotypical sex of the patient.	1
Ethnicity	The ethnicity of the patient based on self-report, as recorded in the medical records or as seen. Recorded according to NHS ethnic category codes and grouped as White, Asian, Black, Other (including Mixed) or not stated.	4
Quintile of deprivation	Population level quintiles within country based on small area multiple deprivation measures combining indicators of income, employment, education, health, crime, barriers to housing and services, and living environment. The latest published version of the measure for each country was used – the English Index of Multiple Deprivation 2019 for postcodes in England, the Welsh Index of Multiple Deprivation 2019 for postcodes in Wales or the Northern Ireland Multiple Deprivation Measure 2017 for postcodes in Northern Ireland.	4
Body mass index	Calculated from the weight (either measured or estimated) and height (either measured or estimated) of the patient as weight in kilograms divided by height in metres squared.	1-4
Any dependency prior to hospital admission	Dependency prior to admission to acute hospital, assessed as the best description for the dependency of the patient in the two weeks prior to admission to acute hospital and prior to the onset of the acute illness based on the level of assistance required with daily activities. Daily activities include bathing, dressing, going to the toilet, moving in/out of bed/chair, continence and eating.	1
Severe conditions in the past medical history	Defined according to APACHE II.(9) Must have been evident in the six months prior to admission to the critical care unit and documented prior to or at admission to the unit. • Respiratory disease: Permanent shortness of breath with light activity due to pulmonary disease or a requirement for home ventilation. • Cardiovascular disease: Fatigue, claudication, dyspnoea or angina at rest (New York Heart Association Functional Class IV). • Liver disease: Biopsy-proven cirrhosis, portal hypertension or hepatic encephalopathy. • Renal disease: Current requirement for chronic renal replacement therapy for irreversible renal disease. • Immunocompromise: AIDS (HIV positive and AIDS-defining illness), congenital immunohumoral or cellular immune deficiency state, metastatic disease (distant metastases), haematological malignancy (acute or chronic), chemotherapy, radiotherapy or daily high dose steroid treatment (≥ 0.3 mg kg^{-1} prednisolone or equivalent).	1-5
Highest temperature	Highest central temperature during the first 24 hours following admission to critical care. If no central temperatures recorded, highest non-central temperature +0.5°C is substituted.	1-4
Lowest systolic blood pressure	Lowest systolic blood pressure during the first 24 hours following admission to critical care.	1-4
Highest heart rate	Highest heart rate during the first 24 hours following admission to critical care.	1-4
Highest respiratory rate	Highest respiratory rate (either ventilated or non-ventilated) during the first 24 hours following admission to critical care.	1-4

Urine output	Total urine output during the first 24 hours following admission to critical care.	1-4
PaO_2/FiO_2 ratio	Ratio of arterial oxygen tension (PaO_2) to fractional inspired oxygen concentration (FiO_2) from the arterial blood gas with the lowest PaO_2 from blood sampled during the first 24 hours following admission to critical care.	1-4
Highest blood lactate concentration	Highest blood lactate concentration during the first 24 hours following admission to critical care.	1-4
Highest serum sodium concentration	Highest serum/plasma sodium concentration during the first 24 hours following admission to critical care.	1-4
Highest blood glucose concentration	Highest blood glucose concentration during the first 24 hours following admission to critical care.	1-4
Highest serum creatinine concentration	Highest serum/plasma creatinine concentration during the first 24 hours following admission to critical care.	1-4
Highest serum urea concentration	Highest serum/plasma urea concentration during the first 24 hours following admission to critical care.	1-4
Lowest haemoglobin concentration	Lowest haemoglobin concentration during the first 24 hours following admission to critical care.	1-4
Lowest white blood cell count	Lowest white blood cell count during the first 24 hours following admission to critical care.	1-4
Neutrophil count	Absolute neutrophil count from the lowest white blood cell count during the first 24 hours following admission to critical care.	1-4
Lowest platelet count	Lowest platelet count during the first 24 hours following admission to critical care.	1-4

eTable 2. Sample size criteria based on Riley et al 2020

Sample size criteria	Sample size requirement
B1. Estimate the overall outcome proportion with sufficient precision Assuming an overall outcome of 42% at 28-days, to estimate this with a precision of ±2.5% requires at least 1475 patients.	1475 patients
B2. Target a small mean absolute prediction error Assuming a mean absolute error in predicted probabilities of ≤0.05 and 30 candidate predictor parameters.	1462 patients
B3. Target a shrinkage factor of 0.9 $R^2_{Cox\text{-}Snell}$ of 0.23 was estimated based on fitting models of similar complexity to random samples of historic patients with viral pneumonia and evaluating these models on independent validation data	19.66 events per parameter
B4. Target small optimism of 0.05 in the apparent $R^2_{Nagelkerke}$ Assuming an outcome of 42% at corresponds to max($R^2_{Cox\text{-}Snell}$)= 0.743 and a shrinkage factor of 0.861	20.96 events per parameter

eTable 3. Missing values and approach to imputation

Predictor	Missing values, n (%)	Imputation model
Calendar time	0 (0.0)	N/A
Age	0 (0.0)	N/A
Sex	4 (0.0)	Logistic regression
Ethnicity	0 (0.0)	N/A
Quintile of deprivation	130 (1.7)	Multinomial logistic regression
Weight	216 (2.9)	Linear regression
Height	365 (4.9)	Linear regression
Any dependency prior to hospital admission	96 (1.3)	Logistic regression
Respiratory disease	79 (1.1)	Logistic regression
Cardiovascular disease	79 (1.1)	Logistic regression
Liver disease	79 (1.1)	Logistic regression
Renal disease	79 (1.1)	Logistic regression
Immunocompromise	79 (1.1)	Logistic regression
Highest temperature	110 (1.5)	Linear regression
Lowest systolic blood pressure	114 (1.4)	Linear regression
Highest heart rate	113 (1.5)	Linear regression
Highest respiratory rate	119 (1.6)	Linear regression
PaO_2/FiO_2	382 (5.1)	Linear regression (log-transformed)
Highest blood lactate concentration	391 (5.2)	Linear regression (log-transformed)
Highest serum creatinine	294 (3.9)	Linear regression (log-transformed)
Highest serum urea	340 (4.5)	Linear regression (log-transformed)
Lowest haemoglobin concentration	231 (3.1)	Linear regression
Lowest platelet count	258 (3.4)	Linear regression (log-transformed)
Highest blood glucose concentration	482 (6.4)	Linear regression
Lowest white blood cell count	255 (3.4)	Linear regression
Neutrophil count	283 (3.8)	Linear regression
Urine output	417(5.6)	Linear regression (log-transformed)

eTable 4. Characteristics of the development and validation samples

Characteristics	All patients (N=10,401)	Development (N=8,666)	Temporal validation (N=1,735)
Physiological and laboratory variables from first 24h of critical care, median (IQR)			
Highest temperature (°C)	38.5 (37.7, 39.3)	38.5 (37.8, 39.3)	38.1 (37.4, 38.9)
Lowest systolic blood pressure (mmHg)	96 (86, 110)	95 (85, 108)	100 (90, 115)
Highest heart rate (beats min^{-1})	105 (92, 119)	105 (92, 119)	105 (92, 118)
Highest respiratory rate (breaths min^{-1})	29 (22, 38)	28 (22, 37)	33 (26, 40)
Urine output (ml)	1395 (950, 1995)	1360 (925, 1947)	1560 (1060, 2290)
Glasgow Coma Scale *, n (%)			
15	4,706 (83.1)	3,619 (83.9)	1,087 (80.4)
9-14	849 (15.0)	611 (14.2)	238 (17.6)
4-8	58 (1.0)	41 (1.0)	17 (1.3)
3	53 (0.9)	43 (1.0)	10 (0.7)
PaO_2/FiO_2 (kPa)	15.7 (11.3, 21.8)	15.6 (11.3, 21.5)	16.0 (11.4, 24.0)
Highest blood lactate concentration (mmol l^{-1})	1.4 (1.1, 1.9)	1.4 (1.1, 1.8)	1.5 (1.2, 2.2)
Highest serum sodium concentration (mmol l^{-1})	139 (136, 141)	139 (136, 141)	139 (135, 141)
Highest blood glucose concentration (mmol l^{-1})	9.2 (7.3, 12.2)	9.1 (7.3, 12.1)	9.4 (7.5, 12.6)
Highest serum creatinine concentration (µmol l^{-1})	82 (63, 119)	84 (64, 121)	74 (57, 107)
Highest serum urea concentration (mmol l^{-1})	6.8 (4.7, 10.4)	7.0 (4.8, 10.6)	6.3 (4.4, 9.8)
Lowest haemoglobin concentration (g l^{-1})	12.0 (10.7, 13.2)	12.1 (10.8, 13.2)	11.6 (9.8, 13.0)
Lowest white blood cell count ($\times 10^9$ l^{-1})	8.6 (6.3, 11.6)	8.6 (6.4, 11.5)	8.8 (6.1, 12.2)
Neutrophil count ($\times 10^9$ l^{-1})	7.2 (5.0, 9.9)	7.2 (5.1, 9.8)	7.2 (4.9, 10.4)
Lowest platelet count ($\times 10^9$ l^{-1})	232 (175, 301)	232 (177, 301)	229 (167, 302)

IQR, interquartile range

* For patients free of the effects of sedative agents at any time during the first 24h of critical care

eTable 5. Summary of model development

Step	Result
Development of a parsimonious physiology model	From the full physiology model, the following potential predictors were dropped: highest sodium; lowest haemoglobin; lowest systolic blood pressure; urine output; highest temperature; lowest white blood cell count.
Development of a full main terms model	All non-physiological predictors were added and sex was dropped. The functional form of BMI change to 3 knots.
Modelling of severe conditions in the past medical history	The model with a single binary covariate for any severe condition in the past medical history was chosen.
Model simplification	Highest glucose was dropped.
Interactions and re-evaluation of non-selected variables	None of the interactions or previously non-selected variables were significant.
Adjustment for calendar time	Time trend was added; there were no changes to other predictors.

eTable 6. Multivariable logistic regression models for death within 28 days of start of critical care (N=8,666)

Predictor	Model without calendar time		Model with calendar time	
	Coefficient (95% CI)	P value	Coefficient (95% CI)	P value
Calendar time (days from 1 March 2020)	N/A	N/A	−0.0147 (-0.0194, -0.0100)	<0.0001
Age (years) – RCS (42,60,74)		<0.0001		<0.0001
Spline base variable 1	0.0402 (0.0303, 0.0501)		0.0410 (0.0311, 0.0510)	
Spline base variable 2	0.0273 (0.0167, 0.0379)		0.0255 (0.0149, 0.0361)	
Quintile of deprivation (vs 1, least deprived)		0.0054		0.0012
2	0.0758 (-0.1015, 0.2530)		0.0944 (-0.0839, 0.2727)	
3	0.0275 (-0.1435, 0.1984)		0.0440 (-0.1271, 0.2152)	
4	0.1566 (-0.0084, 0.3217)		0.1859 (0.0202, 0.3516)	
5 (most deprived)	0.2646 (0.0990, 0.4303)		0.3022 (0.1361, 0.4682)	
Ethnicity (vs white)		<0.0001		<0.0001
Asian	0.4201 (0.2743, 0.5660)		0.4233 (0.2775, 0.5691)	
Black	0.0846 (-0.0906, 0.2599)		0.0556 (-0.1190, 0.2302)	
Mixed/other	-0.0161 (-0.1999, 0.1677)		-0.0272 (-0.2111, 0.1567)	
Not stated	0.2833 (0.0079, 0.5588)		0.2583 (-0.0180, 0.5345)	
Body mass index (kg m^{-2}) – RCS (23.1,28.3,38.4)		0.0231		0.0252
Spline base variable 1	-0.0093 (-0.0305, 0.0119)		-0.0117 (-0.0328, 0.0094)	
Spline base variable 2	0.0245 (-0.0034, 0.0524)		0.0269 (-0.0007, 0.0546)	
Any dependency prior to hospital admission	0.3854 (0.2066, 0.5642)	<0.0001	0.3950 (0.2155, 0.5745)	<0.0001
Any severe condition medical history	0.3703 (0.1705, 0.5702)	<0.0001	0.3993 (0.1975, 0.6010)	<0.0001
Neutrophil count ($\times 10^9$ l^{-1})– RCS (3,6,8.6,16.4)		0.0059		0.0029
Spline base variable 1	-0.0275 (-0.0953, 0.0402)		-0.0259 (-0.0897, 0.0380)	
Spline base variable 2	0.3533 (-0.0172, 0.7239)		0.3516 (-0.0014, 0.7046)	
Spline base variable 3	-0.9217 (-1.8269, -0.0166)		-0.9186 (-1.7856, -0.0517)	
PaO_2/FiO_2 (kPa)– RCS (8.5,15.7,29.5)		<0.0001		<0.0001
Spline base variable 1	-0.0532 (-0.0889, -0.0175)		-0.0566 (-0.0916, -0.0216)	
Spline base variable 2	-0.1280 (-0.3585, 0.1026)		-0.1304 (-0.3564, 0.0956)	
Spline base variable 3	0.3841 (-0.1173, 0.8854)		0.4000 (-0.0918, 0.8919)	
Lowest platelet count ($\times 10^9$ l^{-1}) – RCS (133.1,235,385)		<0.0001		<0.0001
Spline base variable 1	−0.0040 (−0.0052, −0.0025)		-0.0032 (-0.0045, -0.0019)	
Spline base variable 2	0.0030 (0.0011, 0.0044)		0.0020 (0.0004, 0.0036)	
Highest respiratory rate (min^{-1}) – RCS (18,28,45)		0.0049		0.0024
Spline base variable 1	-0.0248 (-0.0384, -0.0111)		-0.0230 (-0.0366, -0.0094)	
Spline base variable 2	0.0339 (0.0149, 0.0528)		0.0335 (0.0146, 0.0523)	
Highest urea (μmol l^{-1}) – RCS (3.5,7,16.6)		0.0043		0.0082
Spline base variable 1	0.0588 (0.0200, 0.0976)		0.0611 (0.0222, 0.0999)	
Spline base variable 2	-0.0787 (-0.1314, -0.0260)		-0.0809 (-0.1336, -0.0282)	
Highest heart rate (min^{-1})	0.0085 (0.0060, 0.0110)	<0.0001	0.0087 (0.0062, 0.0112)	<0.0001
Highest blood lactate (mmol l^{-1}) – RCS (0.9,1.4,2.6)		<0.0001		<0.0001
Spline base variable 1	0.2298 (-0.2581, 0.7176)		0.2355 (-0.2401, 0.7111)	
Spline base variable 2	1.7982 (-2.5421, 6.1386)		1.8097 (-2.4483, 6.0677)	
Spline base variable 3	-4.8098 (-14.0357, 4.4160)		-4.8384 (-13.8966, 4.2198)	
Highest creatinine (μmol l^{-1}) – RCS (45,66,83,115,323)		<0.0001		<0.0001
Spline base variable 1	-0.0161 (-0.0273, -0.0048)		-0.0172 (-0.0284, -0.0060)	
Spline base variable 2	1.0491 (0.4330, 1.6652)		1.0407 (0.4282, 1.6532)	
Spline base variable 3	-2.7076 (-4.4125, -1.0027)		-2.6488 (-4.3430, -0.9546)	
Spline base variable 4	1.8349 (0.5484, 3.1214)		1.7634 (0.4846, 3.0421)	
Constant	-1.9557 (-3.1921, -0.7193)		-1.4247 (-2.6729, -0.1766)	

RCS (k_1,…,k_j) indicates restricted cubic spline with knots at positions k_1 to k_j, corresponding to the following base variables for predictor x:

Spline base variable 1 = x

Spline base variable i+1 = $[\max((x - k_i)^3, 0) - (k_j - k_i) \times \max((x - k_{j-1})^3, 0)/(k_j - k_{j-1}) + (k_{j-1} - k_i) \times \max((x - k_j)^3, 0)/(k_j - k_{j-1})]/(k_j - k_1)^2$; i =1, … , j−2

References

1. Riley RD, Ensor J, Snell KIE, Harrell FE, Jr., et al: Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; 368:m441
2. van Buuren S: Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007; 16(3):219-242
3. Moons KG, Donders RA, Stijnen T, Harrell FE, Jr.: Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 2006; 59(10):1092-1101
4. Rubin DB: Multiple Imputation for Nonresponse in Surveys. New York, J Wiley & Sons, 1987
5. Harrell FE, Jr.: Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, Springer, 2001
6. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, et al: Evaluating the yield of medical tests. *JAMA* 1982; 247(18):2543-2546
7. Brier GW: Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 1950; 75:1-3
8. Cox DR: Two further applications of a model for binary regression. *Biometrika* 1958; 45:562-565
9. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13(10):818-829