SUPPLEMENTARY MATERIAL

Appendix 1. Baseline clinical characteristics of the derivation cohort

All patients (N = 983)	Cohort	Survivors (N=689)	Non-survivors (N=294)	*P* value
Age, years – median (interquartile range)	70 (53 – 83)	61 (50 – 78)	81 (72 – 87)	<0.0001
Age range – no. (%)				
18 – 49 years	176 (17.9)	168 (24.4)	8 (2.7)	<0.001*
50 – 59 years	160 (16.3)	140 (20.3)	20 (6.8)	--
60 – 69 years	151 (15.4)	117 (17.0)	34 (11.6)	--
70 - 79 years	181 (18.4)	106 (15.4)	75 (25.5)	--
≥ 80 years	315 (32.0)	158 (22.9)	157 (53.4)	--
Age – median (IQR) by level of care				
Virtual hospital	53 (43 – 67)	53 (42 – 64)	81 (78 – 86)	0.007
Ward	77 (61 – 86)	70 (55 – 82)	84 (77 – 89)	<0.0001
Ward + received CPAP	71 (61 – 75)	61 (58 – 68)	75 (73 – 81)	<0.0001
Intensive Care Unit (ICU)	60 (52 – 67)	59 (50 – 62)	61 (56 – 71)	0.0349
Male sex – no. (%)	516 (52.5)	351 (50.9)	165 (56.1)	0.137
Ethnic background				
White	760 (77.3)	511 (74.1)	249 (84.7)	0.003*
Asian	162 (16.5)	127(18.4)	35 (11.9)	--
Black	44 (4.5)	37 (5.4)	7 (2.4)	--
Other	17 (1.7)	14 (2.0)	3 (1.0)	--
Smoking history – no. (%)				
Former or current smoker	168 (17.1)	94 (13.7)	74 (25.3)	<0.001
BMI > 30 – no. (%)	243 (24.7)	156 (23.3)	87 (29.8)	0.033
Care Home residency – no. (%)	204 (20.8)	101 (14.7)	103 (35.0)	<0.001
Clinical frailty score – no./total scored (%)				
1 – 4	250/644 (38.8)	192/415 (46.3)	58/229 (25.3)	<0.001*
5 – 6	268/644 (41.6)	152/415 (36.6)	116/229 (50.7)	--
7 – 9	126/644 (19.6)	71/415 (17.1)	55/229 (20.0)	--
Symptoms at presentation– no. (%)				
Fever (temperature >37.3°C)	508 (61.0)	357(62.2)	151 (58.1)	0.259
Breathlessness	482 (57.9)	342 (59.6)	140 (54.1)	0.135
Cough	440 (52.9)	325 (56.7)	115 (44.4)	0.001

Myalgia	181 (21.7)	148 (25.9)	33 (21.8)	<0.001
Headache	62 (7.4)	50 (8.8)	12 (4.6)	0.036
Symptom duration, days – median (IQR)	6 (2 – 11)	7 (3 – 11)	5 (2 – 13)	0.454
Vital baseline observations – no./total (%)				
Respiratory rate >24/min	294/819 (35.9)	155/562 (27.6)	139/257 (54.1)	<0.001
SpO2 ≤92% (on ambient air)	258/822 (31.4)	125/564 (22.2)	133/258 (51.6)	<0.001
Systolic blood pressure <90 mm Hg	23/811 (2.8)	12/556 (2.2)	11/255 (4.3)	0.086
Pulse rate >120/min	81/818 (9.9)	44/560 (7.9)	37/258 (14.3)	0.004
Laboratory findings – no./total (%)				
C-reactive protein >50 mg/L	524/812 (64.5)	306/528 (58.0)	218/284 (76.7)	<0.001
Total white cell count >11 x 10^9/L	175/891 (19.6)	82/600 (13.7)	93/291 (32.0)	<0.001
Lymphocyte count ≤0.7 x 10^9/L	304/890 (34.2)	174/599 (29.1)	130/291 (44.7)	<0.001
Chronic kidney disease – no./total (%)				
Stage 1 (eGFR ≥ 90 ml/min/1.73 m^2)	118/832 (14.2)	94/547 (17.2)	24/285 (8.4)	<0.001*
Stage 2 (eGFR 60-89 ml/min/1.73 m^2)	380/832 (45.7)	286/547 (52.3)	94/285 (33.0)	--
Stage 3 (eGFR 30-59 ml/min/1.73m^2)	237/832 (28.5)	128/547 (23.4)	109/285 (38.3)	--
Stage 4 (eGFR 15-29 ml/min/1.73m^2)	65/832 (7.8)	29/547 (5.3)	36/285 (12.6)	--
Stage 5 (eGFR <15 ml/min/1.73m^2)	32/832 (3.9)	10/547 (1.8)	22/285 (7.7)	--
≥4 abnormal CXR zones – no./total (%)	338/895 (37.8)	203/615 (33.0)	135/280 (48.2)	<0.001
Co-morbid conditions – no. (%)				
Hypertension	475 (48.4)	288 (41.8)	187 (63.8)	<0.001
Ischaemic heart disease	194 (19.7)	103 (15.0)	91 (31.0)	<0.001
Cardiac failure	33 (3.4)	21 (3.1)	12 (4.1)	0.41
Cardiac arrhythmias	34 (3.5)	20 (2.9)	14 (4.8)	0.144
Diabetes mellitus	232 (23.6)	150(21.8)	82 (28.0)	0.036
Respiratory disease	293 (30.0)	199 (29.1)	94 (32.1)	0.35
Chronic kidney disease	196 (20.0)	102 (14.8)	94 (32.1)	<0.001
Cerebrovascular disease	107 (10.9)	47 (6.8)	60 (20.4)	<0.001
Mental health/behavioural disorders				
Dementia	157 (16.0)	73 (10.6)	84 (28.6)	<0.001
Anxiety, depression or both	156 (15.9)	113 (16.4)	43 (14.6)	0.486
Prescribed medications ≥5 – no. (%)	550 (56.1)	340 (49.4)	210 (71.7)	<0.001
Median length of stay – days (IQR)	7 (3.0 – 13.5)	7 (3.0 – 15)	7 (3.0 – 12)	0.564

Abbreviations: BMI – body mass index; CXR – chest radiograph; ED –Emergency Department; GFR – glomerular filtration rate; IQR – inter-quartile range. *P* values denote comparisons between survivors and non-survivors. *χ2 test comparing all subcategories.

Appendix 2. In-patient mortality rate by age bracket in the derivation and validation cohorts

Age group	Derivation cohort	Validation cohorts	
	(N=983)	**ISARIC (N=14,231)**	**Aintree (N=303)**
< 50	4.6%	5.6%	9.7%
50 - 59	12.5%	13.8%	7.9%
60 – 69	22.5%	25.4%	27.0%
70 – 79	41.4%	37.5%	44.1%
≥ 80	49.8%	46.2%	64.9%
Overall mortality	29.9%	30.9%	31.7%

Appendix 3. In-patient mortality rate by level of care in the derivation cohort (N=983)

		LEVEL OF MAXIMAL CARE			
	Whole cohort	**VH**	**Ward**	**Ward CPAP**	**ICU**
N	983	228	627	41	87
Median age (IQR)	70 (53 – 83)	53 (43 – 67)	77 (61 – 86)	71 (61 – 75)	60 (52 – 67)
Deaths	294/983 (29.9%)	4/228 (1.8%)	216/627 (34.4%)	20/41 (48.8%)	54/87 (62.1%)
% of non-survivors by ethnicity at each level of care		W 3 (75.0%) A 1 (25.0%) B 0 OTH 0	W 194 (90.0%) A 17 (7.9%) B 4 OTH 1	W 16 (80.0%) A 3 (15.0%) B 0 OTH 1	W 37 (68.5%) A 13 (24.1%) B 3 OTH 1

The difference between the median age of patients by level of care, expressed as the Chi-square value with ties, is 150.542 (with 80 degrees of freedom) (*P* <0.001).

Abbreviations: A – Asian; B – Black; CPAP – continuous positive airway pressure; ICU – Intensive Care Unit; IQR – interquartile range; OTH – other ethnicity; VH – virtual hospital; W – White.

Appendix 4. 11-predictor score (1 – 18 points) for predicting in-hospital COVID-19 death

PREDICTOR	POINTS
SpO_2	
> 92% on air	0
≤ 92% on air	1
Obesity (BMI >30)	
absent	0
present	1
Age	
< 50	0
50 – 59	1
60 – 69	2
70 – 79	3
≥ 80	4
Respiratory rate	
≤ 24/min	0
> 24/min	1
Stroke/CVA	
absent	0
present	1
Ever smoked	
no	0
yes	1
Dementia	
no	0
yes	1
CKD stage	
1	1
2	2
3	3
4	4
5	5
White cell count >11 $x10^9$	
no	0
yes	1
Lymphocytes ≤ 0.7 $x10^9$	
no	0
yes	1
CXR (≥ 4 zones affected)	
no	0
yes	1

Appendix 5. Comparison of key parameters between the derivation and validation cohorts

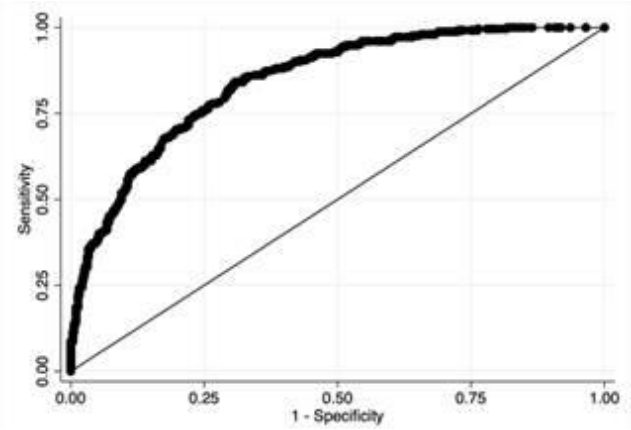
	Derivation cohort N = 983	**Validation (ISARIC) N = 14,231**	***P* value**
Median age (IQR)	70 (53 – 83)	73 (59 – 83)	NS
Age range (years)			
70 – 79	18.4%	21.9%	<0.05
≥ 80	32.5%	34.8%	NS
Mortality	29.9%	30.9%	NS
Ethnicity (% of cohort)			
White	77.3%	69.0%	<0.01
BAME	22.7%	31.0%	
Male sex (% of cohort)	52.5%	55.7%	<0.05

	Derivation cohort N = 983	**Validation (Aintree) N = 303**	***P* value**
Median age (IQR)	70 (53 – 83)	67 (57 – 77)	NS
Age range (years)			
70 – 79	18.4%	22.4%	NS
≥ 80	32.5%	18.8%	<0.001
Mortality	29.9%	31.7%	NS
Ethnicity (% of cohort)			
White	77.3%	95.7%	<0.001
BAME	22.7%	4.3%	
Male sex (% of cohort)	52.5%	61.1%	<0.01

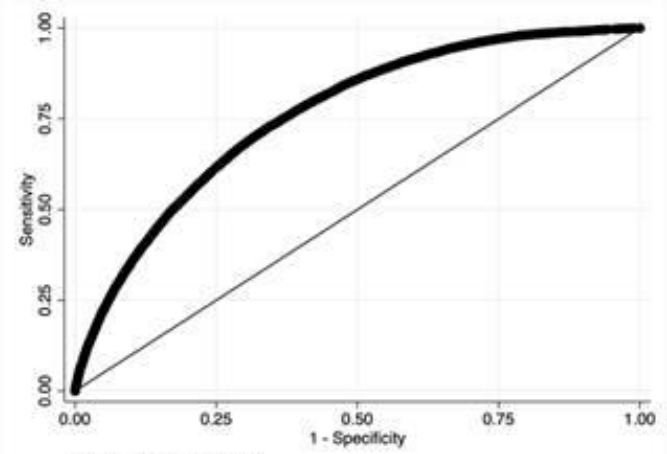
Abbreviations: BAME – Black Asian or other Minor Ethnicity Black; IQR – interquartile range

Appendix 6. Receiver operating curves (ROCs) of the 11-predictor and 5-predictor (SOARS) scores on the derivation and validation cohorts

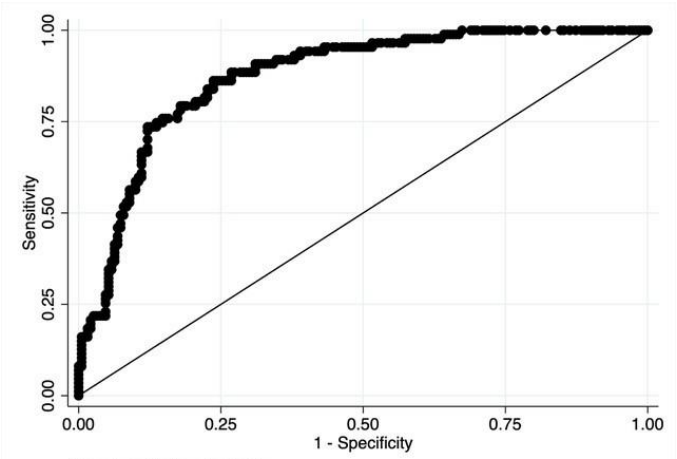
Area under the ROC (AUROC) for the 11-predictor score on the derivation cohort (0.84)



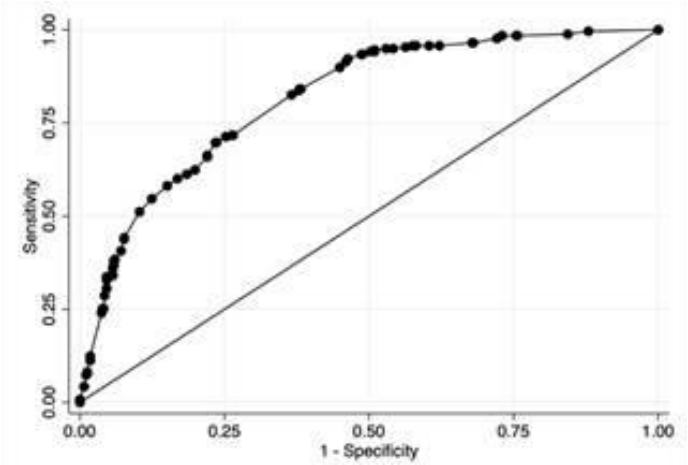
Area under the ROC (AUROC) for the 11-predictor score on the ISARIC validation cohort (0.77)



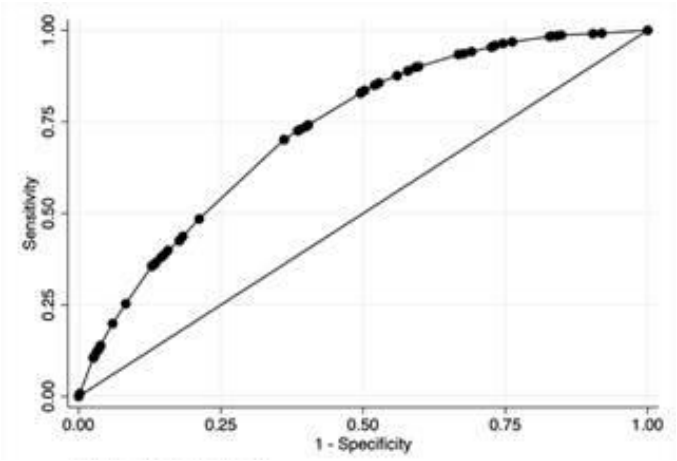
Area under the ROC (AUROC) for the 11-predictor score on the Aintree validation cohort (0.87)



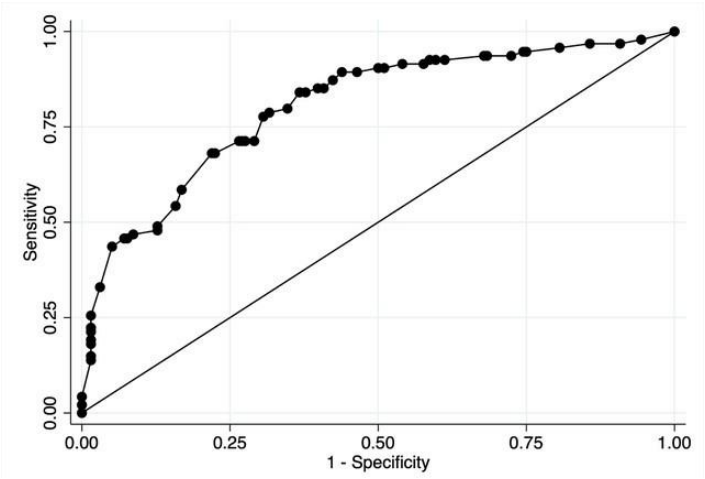
Area under the ROC (AUROC) for the 5-predictor score on derivation cohort (0.82)



Area under the ROC (AUROC) for the 5-predictor score on ISARIC validation cohort (0.74)



Area under the ROC (AUROC) for the 5-predictor score on Aintree validation cohort (0.80)



Appendix. Pre-specified study protocol for clinical data collection

PREDICT COVID Clinical Data Collection protocol
Date: 1st March 2020

Aims

1. To prospectively collect data on all adults (> 18 years) with laboratory-confirmed SARS-CoV-2 infection (COVID-19) presenting to Watford Hospital, West Hertfordshire NHS Trust, during the first wave of the SARS-CoV-2 pandemic, including an outcome of in-hospital death or hospital discharge,
2. To develop a prognostic (risk prediction) score using the above derivation data,
3. To construct a practical clinical scoring system for predicting mortality (and risk of morbidity) after validation of score against external, i.e. independent cohorts of COVID-19 patients from other UK sites,
4. To include and assess outcomes of patients referred to the COVID-19 Virtual Hospital (out-of-hospital monitoring) in the same NHS Trust,
5. To monitor and characterise surviving patients for up to 12 months from the time of confirmation of SARS-CoV-2 infection, including the domains of psychological, physiological and radiological impairment and recovery.

Primary outcome
In-hospital mortality with minimum 30-day follow-up data.

Secondary outcomes
Longer term mortality and morbidity: radiological, psychological and cardiorespiratory. This will also be assessed against ongoing health care needs.

Patient inclusion
All adult patients (aged >18 years) with SARS-CoV-2 real-time reverse transcriptase polymerase chain reaction (rRT-PCR) confirmation.
Completed admission and outcomes at 3 months and 12 months.

Inclusion criteria:

- Readily available patient or clinical characteristic to attending clinicians upon presentation to hospital (Accident & Emergency department, Acute Medical Receiving Unit)
- Blood markers should be commonly measured and results available for review within the first 24 hours of admission
- All parameters relating to oxygen supplementation and advanced respiratory support including one or more of: continuous positive airway pressure (CPAP), bilevel non-invasive ventilation (NIV), high-flow nasal cannula (HFNC) oxygenation and invasive mechanical ventilation (IMV) needs

Exclusion criteria
All SARS-CoV-2 rRT-PCR negative patients irrespective of clinical suspicion of COVID-19.
All individuals aged <18y.

Selection of candidate variables for initial data collection
Candidate variables were chosen based on knowledge of their potential association/s with SARS-CoV-2 infection and clinical disease (COVID-19). A systematic literature search was undertaken to identify these variables with respect to their predictive association for mortality and other adverse outcomes including COVID-19 severity and disease-related complications such as requirement for critical care and the development of COVID-19-associated acute respiratory distress syndrome (ARDS).

Systematic literature for English language articles in the following search databases: PubMed, EMBASE, WHO Medicus, Web of Science and Google Scholar (particularly for pre-print publications on medRxiv). Search terms included SARS-CoV-2; COVID-19; coronavirus; ARDS; pneumonia; sepsis; influenza; risk prediction; risk score; prognosis; validation. No date restrictions were imposed.

Statistical analysis for derivation and validation modelling

In analysing the data collected from the derivation (West Herts) cohort, categorical variables will expressed as frequency (%), with significance determined by the Chi-squared test. Continuous variables will be analysed for median (interquartile range) or mean (standard deviation) outcomes and analysed by the t-test, Kruskal-Wallis or Mann-Whitney U test, as appropriate. Missing data in the derivation cohort will be expected even with prospective data collection; missingness of data will be assumed to be at random and handled by multiple imputation by chained equations (MICE) with at least ten imputations, provided the proportion of missingness for the defined parameter constitutes no more than 20% of the cohort. Collated data will be subjected to univariate and multivariate logistic regression in order to determine odds ratios (OR) for in-hospital mortality. The latter will also be internally validated by bootstrapping using a minimum of 1000 re-samples. Predictor interactions will be analysed by appropriate methodology such as the likelihood ratio (LR) test comparing broad and narrow (constrained) models.

All performance metrics against external validation cohorts will be analysed using the prediction score and not with the multivariate regression model. The performance of the derivation model will be assessed for discriminatory ability (area under the receiver operating characteristic, AUROC) and calibration (graphical representation of Hosmer-Lemeshow analysis).

All statistical analyses including risk modelling calculations will be performed using STATA, version 16 (Stata Corp., Texas, USA).

R Vancheeswaran, F Chua, A Draper, T Vaghela and A Barlow (study design and responsible persons for data analysis and interpretation)

MARCH 2020