Intensive Care Med (2021) 47:549–565
https://doi.org/10.1007/s00134-021-06389-z

ORIGINAL

Natural history, trajectory, and management of mechanically ventilated COVID-19 patients in the United Kingdom

Brijesh V. Patel[1,2*], Shlomi Haar[3,4,5,6], Rhodri Handslip[1,2], Chaiyawan Auepanwiriyakul[3,4], Teresa Mei-Ling Lee[1,2], Sunil Patel[1,2], J. Alex Harston[3,4], Feargus Hosking-Jervis[7], Donna Kelly[8], Barnaby Sanderson[9], Barbara Borgatta[10], Kate Tatham[1,11], Ingeborg Welters[12], Luigi Camporota[9], Anthony C. Gordon[1,13], Matthieu Komorowski[1,13], David Antcliffe[1,13], John R. Prowle[14], Zudin Puthucheary[14] and Aldo A. Faisal[3,4,15,16*]on behalf of the United Kingdom COVID-ICU National Service Evaluation

Abstract

Purpose: The trajectory of mechanically ventilated patients with coronavirus disease 2019 (COVID-19) is essential for clinical decisions, yet the focus so far has been on admission characteristics without consideration of the dynamic course of the disease in the context of applied therapeutic interventions.

Methods: We included adult patients undergoing invasive mechanical ventilation (IMV) within 48 h of intensive care unit (ICU) admission with complete clinical data until ICU death or discharge. We examined the importance of factors associated with disease progression over the first week, implementation and responsiveness to interventions used in acute respiratory distress syndrome (ARDS), and ICU outcome. We used machine learning (ML) and Explainable Artificial Intelligence (XAI) methods to characterise the evolution of clinical parameters and our ICU data visualisation tool is available as a web-based widget (https://www.CovidUK.ICU).

Results: Data for 633 adults with COVID-19 who underwent IMV between 01 March 2020 and 31 August 2020 were analysed. Overall mortality was 43.3% and highest with non-resolution of hypoxaemia [60.4% vs17.6%; $P<0.001$; median PaO_2/FiO_2 on the day of death was 12.3(8.9–18.4) kPa] and non-response to proning (69.5% vs.31.1%; $P<0.001$). Two ML models using weeklong data demonstrated an increased predictive accuracy for mortality compared to admission data (74.5% and 76.3% vs 60%, respectively). XAI models highlighted the increasing importance, over the first week, of PaO_2/FiO_2 in predicting mortality. Prone positioning improved oxygenation only in 45% of

*Correspondence: brijesh.patel@imperial.ac.uk; aldo.faisal@imperial.ac.uk
[1] Division of Anaesthetics, Pain Medicine & Intensive Care, Department of Surgery & Cancer, Faculty of Medicine, Imperial College London, London, UK
[3] Brain & Behaviour Lab, Dept. Of Computing, Imperial College London, London, UK
Full author information is available at the end of the article

Brijesh V. Patel, Shlomi Haar, Rhodri Handslip are first authors.
John R. Prowle, Zudin Puthucheary, Aldo A Faisal are joint senior authors.

The members of the United Kingdom COVID-ICU National Service Evaluation are mentioned in Acknowledgements section.

Springer

patients. A higher peak pressure (OR 1.42[1.06–1.91]; $P < 0.05$), raised respiratory component (OR 1.71[1.17–2.5]; $P < 0.01$) and cardiovascular component (OR 1.36 [1.04–1.75]; $P < 0.05$) of the sequential organ failure assessment (SOFA) score and raised lactate (OR 1.33 [0.99–1.79]; $P = 0.057$) immediately prior to application of prone positioning were associated with lack of oxygenation response. Prone positioning was not applied to 76% of patients with moderate hypoxemia and 45% of those with severe hypoxemia and patients who died without receiving proning interventions had more missed opportunities for prone intervention [7 (3–15.5) versus 2 (0–6); $P < 0.001$]. Despite the severity of gas exchange deficit, most patients received lung-protective ventilation with tidal volumes less than 8 mL/kg and plateau pressures less than 30cmH_2O. This was despite systematic errors in measurement of height and derived ideal body weight.

Conclusions: Refractory hypoxaemia remains a major association with mortality, yet evidence based ARDS interventions, in particular prone positioning, were not implemented and had delayed application with an associated reduced responsiveness. Real-time service evaluation techniques offer opportunities to assess the delivery of care and improve protocolised implementation of evidence-based ARDS interventions, which might be associated with improvements in survival.

Keywords: COVID-19, ARDS, Mechanical ventilation, Prone position, Mortality risk, Artificial intelligence

Introduction

Coronavirus disease 2019 (COVID-19) caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) was declared a global pandemic on March 11, 2020 by the World Health Organisation. COVID-19 related severe acute hypoxemic respiratory failure invariably leads to intensive care unit (ICU) admission. These patients fulfil Acute Respiratory Distress Syndrome (ARDS) Berlin definition criteria [1–4]. However, uncertainties around the extent of pathological and physiological differences between COVID-19 related ARDS and other causes of ARDS and the pulmonary angiopathy of COVID-19 further fuel the uncertainties regarding the full disease progression and management of these acutely unwell patients with a high mortality rate [5–9]. This ambiguity leads to an ongoing debate on the application of existing evidence-based ARDS management to COVID-19 patients [10].

Pre-COVID evidence-based guidelines for ARDS management include lung-protective ventilation, prone positioning, conservative fluid strategies with the option of open lung strategy and neuromuscular blockade (NMBA) alongside patients with severe hypoxaemia refractory to these interventions having timely access to extracorporeal membrane oxygenation (ECMO) support [9, 11, 12]. Moreover, reports suggest that real-world compliance with evidence-based ARDS management strategies is difficult at a system level [13]. Furthermore, these interventions are implemented at various stages of ARDS progression and are time-sensitive over the natural history of illness [14, 15]. Monitoring of dynamic responsiveness to interventions is fundamental to clinical practice in critical care and is increasingly facilitated by artificial intelligence analytics [16]. Whilst there have been reports of the epidemiological and admission characteristics of hospitalised patients with COVID-19 admitted to intensive care from around the world, none have a focus on a complete clinical trajectory in combination with clinical application and response to ARDS management strategies [1–4].

Take-home message

In a cohort of COVID-19 patients treated in the United Kingdom, progressive respiratory failure was increasingly associated with mortality. Evidence-based triggers for ARDS interventions, in particular prone position, were not implemented, had delayed application, or showed poor responsiveness in a sizeable proportion of patients with progressive hypoxaemia. How this implementation gap and lack of response to conventional ARDS interventions may have contributed to excess mortality across the pandemic deserves further interrogation.

Accordingly, we undertook a cohort study across several intensive care units in the United Kingdom, to report the natural history and management of mechanically ventilated COVID-19 patients. Our specific aims were to define, from routine clinical measurements, crucial factors associated with disease progression and mortality; and to ascertain use, compliance, duration and effect of established evidence-based ARDS management strategies.

Methods

Study design

We performed a multicentre, observational cohort study in patients with SARS-CoV-2 infection who required mechanical ventilation for severe COVID-19 infection in the United Kingdom.

Eligibility criteria

Adult patients (aged ≥ 18 years) with laboratory-confirmed SARS-CoV-2 infection who required invasive mechanical ventilation (IMV) in the United Kingdom between March 1st and August 31st, 2020. Only patients transferred to study sites within 48 h of intubation were included and due to the nature of ECMO provision in the UK those patients were excluded (see Supp. Methods: ECMO).

Ethical approval

The UK Health Research Authority exempted this study from review by an NHS Research Ethics Committee due to its urgent need. Each site registered the study protocol as a service evaluation. The "Strengthening the Reporting of Observational Studies in Epidemiology" statement guidelines were applied (see Supp. Appendix p.4–5) [17].

Data collection and procedures

We setup standardised data processing pipelines to manage the considerable daily data flow. Only routine, pseudonymised data were collected with no change to clinical care. In brief, the case report form captured admission demographics, twice daily (8 am and 8 pm) respiratory physiology and blood gas results, daily ARDS interventions, daily COVID-19 interventions, daily blood results and outcome status. Table S1 lists the participating sites. Data were extracted from either electronic healthcare records (EHRs) or paper-based records into the COVID-ICU secure REDCap database (REDCap v10.0.10; Vanderbilt University, US).

Missing data

We made the heuristic decision of setting the threshold of data completeness (i.e. missingness) to balance off the number of patients, against the number of variables. We defined this by examination of the available variables in the first 48 h of admission or the last 36 h before prone or the first 36 h after prone. If in these 3 or 4 12-h measurement points, all were missing, then we counted this patient as 'missing' data. The missingness is thus the percentage of patients where there is no measurement in this 36/48-h window for a modality. Percentage of missing data per modality are shown in Table S2, and details of missing data are shown in Table S3 and S4.

Data presentation and group definitions

Descriptive variables are expressed as percentage, or median and interquartile range (IQR), as appropriate. Continuous variables were analysed with Mann–Whitney-*U* or Kruskal–Wallis tests, as appropriate. Categorical variables were compared using Fisher's exact test or the Chi-square test for an equal proportion, as appropriate. All statistical tests were 2-sided and $P \leq 0.05$ was considered statistically significant. Mortality was defined as ICU mortality. The incidence and duration of interventions, as well as ventilation settings, were analysed and reported to current strategies e.g., low tidal volume ventilation and ARDSNet Positive End Expiratory Pressure (PEEP) tables. We defined an intervention period as a daily application of the intervention with a day of no intervention defining the end of the current period. For group-wise analysis, the outcome of the therapies was measured as categorical variables of "Mild, Moderate, or Severe", "Survival or Death", "resolver or non-resolver", and "prone responder or non-responder". The *severity of hypoxaemia* was categorised as per Berlin Definition criteria [18]. To evaluate features associated with the progression of hypoxaemia, we analysed the evolution of hypoxaemia over the first 7 days of invasive mechanical ventilation and categorised them into two groups, "resolvers" and "non-resolvers". Patients whose hypoxemia categorisation improved or got discharged from ICU were considered "resolvers" while those whose hypoxaemia categorisation worsened, or died, were considered "non-resolvers". We further considered the longer-term effect on PaO_2/FiO_2 after prone positioning and defined prone responsiveness as maintenance of a mean $PaO_2/FiO_2 > 20$ kPa over 7 days after the first prone episode. We defined a *prone opportunity* as per inclusion criteria for the PROSEVA study to assess opportunities to apply a prone intervention [19]: a $PaO_2/FiO_2 < 20$ kPa, with an $FiO_2 \geq 0.6$, a PEEP $\geq 5 cmH_2O$. Prone opportunities were measured at 8 am and 8 pm with the ventilator and arterial blood gas (ABG) evaluation.

Multivariate logistic regression using backward method was applied for variable selection (with screening univariate, $P < 0.1$) to each outcome variable to test associations with independent variables. The full list of variables tested for inclusion (and missingness) in these models is shown in Supplementary Table S2. For details see Supp. Methods "Logistic regression in statistical analysis (details)".

Statistical analysis of longitudinal measures

The association between the change over time of each independent variable and the outcome measures was tested in repeated measures (rm) ANOVA. For the survival and first week resolver outcome, rmANOVA was applied on the physiology variables over the first week of mechanical ventilation, while for the prone responder outcome, it was applied on the physiology variables over

a week from the day before the first PP episode. We accounted for multiple comparisons in the interaction statistic by controlling the false discovery rate (FDR). For details see Supp. Methods "Statistical analysis of longitudinal measures (details)".

Machine learning models for daily and week-long-mortality prediction

We evaluated both logistic regression and more potent machine learning models at predicting mortality. We used these models to predict (a) mortality based on a single day's data, for each of the first seven days of admission of a patient, and (b) used data from the whole week together to predict mortality. We designed our model to be compatible with both daily prediction data (all used clinical parameters for each day in the first week) and weekly prediction (all used clinical parameters over all 7 days). We used a deep learning framework as a foundation to implement both a logistic regression (LR) predictor and a deep neural network (three-layer multilayer perceptron model—3MLP) predictor. We provide the complete details in the Supp. Materials "Machine Learning models for of daily and weeklong- mortality prediction (details)".

Grouping of dynamic clinical parameter importance through Explainable AI (XAI)

We analysed our deep learning model and the logistic regression model with an XAI approach. We used SHAP (SHapley Additive exPlanations) to explain how the prediction models weighted the importance of individual input features for its output, i.e. here, the clinical parameters [20]. The SHAP value effectively is the difference in how well a system performs when knowing all parameters minus the performance of the system when knowing all but one parameter, namely the one for which the SHAP contribution is computed. It thus measures, the explanatory cost of leaving the parameter out, and so high SHAP values mean that they are very important for the prediction. SHAP values should not be confounded with conventional regression weights, as a high SHAP value of a variable may imply that a larger, or a lower value, or vicinity to a specific value may increase mortality. It is, therefore, that we used SHAP values, as we could directly compare deep learning and logistic regression models using a common 'currency' of explainability. For complete details please see Supp. Methods "Grouping of dynamic clinical parameter importance through Explainable AI (details)".

All statistical analyses were carried out using MATLAB (MathWorks Inc., Natick, MA). Detailed data science methods are described in the supplementary appendix.

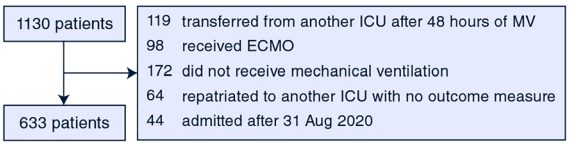

Fig. 1 Study population flowchart

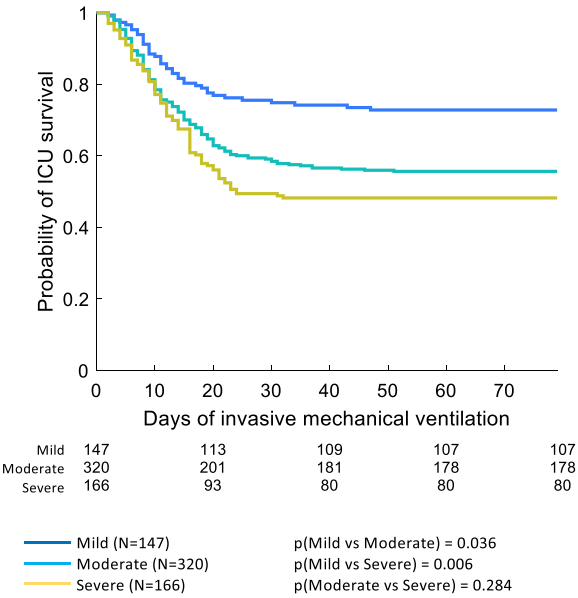

Fig. 2 ICU survival curves based on admission severities of hypoxaemia as defined by the Berlin definition of ARDS

Results

Clinical progression of critically ill COVID-19 patients

A total of 633 mechanically ventilated patients admitted to 13 UK National Health Service (NHS) Trusts with 18 ICU sites between 01 March 2020 and 31 August 2020 had complete daily data up to ICU death or discharge (Fig. 1, Table S1). Baseline demographics (Fig. S1 and Table S5) were similar to the Intensive Care National Audit and Research Centre cohort [21] (Table S6). On initiation of mechanical ventilation, the severity of mild, moderate and severe hypoxaemia was 23.2%; 50.6%, and 26.2%, respectively, with mortality increasing with severity (Fig. 2, Table S7). On admission, increased severity was associated with higher settings for mechanical ventilation, higher severity of organ failure (including dynamic respiratory system compliance, oxygenation index (OI), and ventilatory ratio (VR)) (Table 1), and greater application of interventions (Table 2).

Determinants of mortality

Survival to ICU discharge was 57.7%. There was a difference in mortality between quartiles of patients admitted

Table 1 Clinical and physiological characteristics and outcomes according to severity of hypoxaemia on admission

Label	All			PaO_2/FiO_2 > 26.7		PaO_2/FiO_2 13.3–26.6		PaO_2/FiO_2 < 13.3		
	Units	Total *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group N	median [IQR] / *N* (%)	*p*-value (KruskalWallis/ Chi^2)
Clinical characteristics										
Male		633	481 (76%)	147	107 (72.8%)	320	250 (78.1%)	166	124 (74.7%)	0.411
White		542	250 (46.1%)	126	50 (39.7%)	274	135 (49.3%)	142	65 (45.8%)	0.243
Age	years	633	59 [51-66]	147	59 [48-66]	320	60 [52.5-67]	166	59 [52-65]	0.272
BMI	kg/m^2	524	28.1 [24.9-32.8]	123	28 [24.8-32.4]	278	28 [24.8-32.1]	123	29.3 [25.4-34.6]	0.088
Time since onset of symptoms	days	408	8 [6-12]	90	9 [6-13]	226	8 [6-12]	92	7.5 [6-12]	0.851
ICU length of stay	days	633	14 [8-23]	147	13 [7-24.8]	320	14 [8-23]	166	14 [9-22]	0.355
Length of mechanical ventilation	days	633	13 [7-22]	147	11 [6-23.8]	320	13 [8-22]	166	14 [8-21]	0.295
ICU Mortality	%	633	268 (42.3%)	147	40 (27.2%)	320	142 (44.4%)	166	86 (51.8%)	< 0.001
Ventilation										
FiO_2	%	628	60 [45-80]	143	40 [30-50]	319	60 [50-70]	166	80 [70-90]	< 0.001
PaO_2 to FiO_2 ratio	kPa	626	18.3 [13-25]	142	35.2 [29.9-42.4]	319	18.7 [16-22]	165	10.9 [9.4-12.1]	< 0.001
Tidal Volume	ml/Kg(IBW)	472	6.8 [6-7.8]	112	6.9 [6.2-8]	254	6.7 [6-7.6]	106	7 [6-7.8]	0.264
Respiratory rate	bpm	627	18.8 [16-22]	142	18 [16-22]	319	19 [16-22]	166	20 [16-22.7]	0.075
Minute ventilation	L/minute	606	8.5 [6.9-10.4]	138	8.2 [6.5-9.7]	310	8.6 [6.9-10.5]	158	8.6 [6.9-10.4]	0.078
Peak pressure	ml/Kg(RBW)	599	26 [23-30]	138	24 [21-26]	306	26 [23-29]	155	29 [25-31]	< 0.001
Plateau pressure	ml/Kg(IBW)	80	26 [22.5-28.5]	10	24 [22-26]	48	25 [22-28]	22	28 [25-29]	0.032
PEEP	cmH_2O	603	10 [8-12]	140	10 [7-10]	302	10 [8-12]	161	10 [10-12]	< 0.001
Mean airway pressure	cmH_2O	387	16 [13.2-19]	96	15 [12-17]	194	16 [13-19]	97	17 [15-20]	< 0.001
Pressure support	cmH_2O	371	10 [5.3-14]	83	10 [5-12]	202	10 [6-14]	86	10 [5-14]	0.610
Dynamic Compliance	mls/cmH_20	564	31.5 [24.3-40.2]	135	32.4 [26-40.8]	284	32.9 [25-42]	145	27.1 [22.3-35.8]	< 0.001

Table 1 (continued)

Label	All			PaO_2/FiO_2 > 26.7		PaO_2/FiO_2 13.3–26.6		PaO_2/FiO_2 < 13.3		
	Units	Total *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group N	median [IQR] / *N* (%)	*p*-value (KruskalWallis/ Chi^2)
Oxygenation Index		387	8.1 [5.1-12.5]	96	4 [3.2-5.2]	194	8.2 [6.1-9.9]	97	15.2 [11.8-18.2]	< 0.001
Ventilatory Ratio		470	1.5 [1.2-2.1]	110	1.4 [1-1.9]	254	1.5 [1.3-2]	106	1.7 [1.3-2.4]	<0.001
Arterial Blood Gas (ABG)										
Oxygen saturation	%	518	95 [93-98]	120	97 [95-99]	281	95 [93-97]	117	93 [90-96]	< 0.001
pH		630	7.4 [7.3-7.4]	145	7.4 [7.3-7.4]	320	7.4 [7.3-7.4]	165	7.3 [7.3-7.4]	0.001
PaO_2	kPa	630	10.7 [9.2-13.1]	145	12.8 [11-17.5]	320	10.7 [9.4-12.9]	165	9.3 [7.8-10.7]	< 0.001
$PaCO_2$	kPa	630	6 [5.2-7.2]	145	5.4 [4.7-6.2]	320	6 [5.3-7]	165	6.6 [5.5-8]	< 0.001
Base excess		630	– 0.3 [– 2.6-2.2]	145	– 0.9 [– 3.4-1.3]	320	– 0.3 [– 2.5-1.9]	165	0.7 [– 2.3-4]	< 0.001
HCO_3^-	mmol/L	629	24.5 [22.5-26.7]	145	23.8 [21.2-25.6]	319	24.3 [22.6-26.6]	165	25.3 [22.8-28.1]	< 0.001
Lactate	mmol/L	604	1.2 [1-1.6]	137	1.1 [0.9-1.4]	304	1.2 [1-1.6]	163	1.3 [1-1.8]	0.001
Sequential Organ Failure Assessment (SOFA) score										
SOFA score		428	9 [7-11]	96	7 [5-9]	210	8.5 [7-10]	122	11 [8-12]	< 0.001
SOFA Respiratory		626	3 [3-4]	142	2 [1-2]	319	3 [3-3]	165	4 [4-4]	< 0.001
SOFA Nervous		508	3 [0-4]	119	3 [0-4]	247	0 [0-4]	142	4 [0-4]	0.008
SOFA Cardiovascular		579	3 [3-4]	128	3 [3-4]	297	3 [3-4]	154	3 [3-4]	0.096
SOFA Liver		588	0 [0-0]	138	0 [0-0]	293	0 [0-0]	157	0 [0-0]	0.808
SOFA Coagulation		619	0 [0-0]	143	0 [0-0]	313	0 [0-0]	163	0 [0-0]	0.647
SOFA Kidneys		629	0 [0-1]	145	0 [0-1]	318	0 [0-1]	166	0 [0-1]	0.857
Full Blood Count (FBC)										

Table 1 (continued)

Label	All			$PaO_2/FiO_2 > 26.7$		PaO_2/FiO_2 13.3–26.6		$PaO_2/FiO_2 < 13.3$		
	Units	Total *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group N	median [IQR] / *N* (%)	*p*-value (KruskalWallis/ Chi^2)
Haemoglobin	g/dL	619	114 [92-128]	143	113 [89.3-128]	313	113 [91.8-129]	163	117 [95.5-128]	0.558
Haematocrit		312	0.4 [0.3-0.4]	63	0.3 [0.3-0.4]	172	0.4 [0.3-0.4]	77	0.4 [0.3-0.4]	0.011
White blood cell count	$\times 10^9$/L	619	9.6 [7-13.1]	143	9.3 [6.8-12.6]	313	9.5 [6.9-13]	163	10.1 [7.5-13.7]	0.350
Neutrophils	$\times 10^9$/L	618	8.1 [5.7-11.3]	143	7.5 [5.4-10.5]	313	8 [5.5-11.1]	162	8.6 [6.6-12.3]	0.060
Monocytes	$\times 10^9$/L	614	0.4 [0.3-0.7]	143	0.6 [0.3-0.8]	310	0.4 [0.3-0.7]	161	0.4 [0.3-0.6]	0.005
Lymphocytes	$\times 10^9$/L	615	0.8 [0.5-1.2]	143	0.8 [0.6-1.4]	310	0.8 [0.5-1.2]	162	0.8 [0.5-1.1]	0.359
Basophils	$\times 10^9$/L	498	0 [0-0.1]	110	0 [0-0.1]	248	0 [0-0.1]	140	0 [0-0]	0.135
Eosinophils	$\times 10^9$/L	493	0 [0-0.1]	113	0 [0-0.1]	244	0 [0-0.1]	136	0 [0-0.1]	0.925
Coagulation										
Platelet Count	$\times 10^9$/L	619	246 [185.3-320.8]	143	244 [190.3-332]	313	250 [183-322.8]	163	237 [185-307.5]	0.811
APTT	U/L	398	32.1 [28.3-37.4]	81	31 [28.6-36.4]	195	31.8 [28-36.5]	122	33.9 [28.8-38]	0.188
PT	U/L	396	13.9 [12.4-15.2]	81	13.9 [12.7-15]	193	13.7 [12.1-15]	122	14.3 [12.9-15.6]	0.057
INR	U/L	212	1.1 [1.1-1.2]	43	1.2 [1.1-1.3]	114	1.1 [1-1.2]	55	1.1 [1.1-1.2]	0.060
Fibrinogen	U/L	410	6.8 [5.6-8.1]	100	6.2 [5.4-7.6]	218	7.1 [5.9-8.4]	92	6.5 [5-7.6]	0.001
D– dimer	IU/L	391	2642 [990.5-7701.3]	85	2310 [980.8-5809.5]	183	2050 [930.3-6278.8]	123	3390 [1366.3-9851.8]	0.018
Electrolytes										
Blood Urea Nitrogen (BUN)	mmol/L	498	7.4 [4.9-11.8]	111	6.3 [4.4-10.6]	249	7.7 [5.3-12.2]	138	7.8 [4.8-11.8]	0.045
Creatinine	µmol/L	621	88 [66-140]	144	85 [65.5-148.5]	314	89 [71-141]	163	85 [63-131]	0.322
Sodium	mmol/L	621	139 [136-142]	144	139 [136.5-142]	314	138 [135-141]	163	139 [136-142]	0.049

Table 1 (continued)

Label	All			PaO_2/FiO_2 > 26.7		PaO_2/FiO_2 13.3–26.6		PaO_2/FiO_2 < 13.3		
	Units	Total *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group N	median [IQR] / *N* (%)	*p*-value (KruskalWallis/ Chi^2)
Potassium	mmol/L	620	4.4 [4-4.8]	143	4.5 [4.1-4.8]	314	4.4 [4.1-4.8]	163	4.4 [4-4.8]	0.512
Liver Enzymes										
Bilirubin	µmol/L	588	10 [7-15]	138	10 [7-14]	293	10 [7-15]	157	10 [7-15]	0.344
Alkaline Phosphatase	U/L	600	77 [58.5-113]	138	76 [55-109]	300	73.5 [53-107.5]	162	87.5 [65-133]	0.002
AST	U/L	99	59 [39.3-85]	22	62 [46-108]	52	59.5 [37-84.5]	25	56 [42.3-74.5]	0.615
ALT	U/L	592	37 [24-59]	138	32 [22-60]	294	37 [25-56]	160	42.5 [27-68.5]	0.072
Inflammation measurements										
LDH (Lactate dehydrogenase)	IU/L	194	649 [452-921]	36	553.5 [394-832]	107	622 [467.5-792.8]	51	728 [512-1228]	0.048
Ferritin	ng/mL	290	1218.5 [696-2320]	61	1070 [651.8-1871.8]	157	1296 [701.5-2342.5]	72	1356 [717.5-2743]	0.363
CRP	mg/L	592	215.7 [135-311]	144	165.5 [86-260.5]	292	223.8 [145.7-316.4]	156	242.7 [167.8-321.1]	< 0.001
Procalcitonin	ug/L	125	0.7 [0.3-2.2]	34	0.3 [0.2-1]	57	0.8 [0.3-2.5]	34	1.4 [0.5-2.6]	0.002
Cardiac Enzymes										
Creatinine Kinase	U/L	243	217 [83.5-637.3]	50	285.5 [68-1078]	123	222 [98-571]	70	171.5 [93-484]	0.701
High sensitivity Troponin	ng/L	253	21.8 [11-61]	52	22.6 [11-110.9]	130	20.3 [10-55.1]	71	22 [12.2-75.5]	0.360
NT Pro BNP	pg/ml	58	537.5 [165-1478]	8	255.5 [103-1201]	33	433 [158.5-1545]	17	692 [416.3-1763]	0.457
Fluid Balance										
Cumulative Fluid balance	L	589	343 [– 212.3-1058.3]	135	205 [– 301.9-821.3]	298	365 [– 185-1074]	156	452.5 [– 264.5-1597]	0.044

(peak: 31st March; median: 1st April 2020) during the first surge ($P=0.053$). This showed the first quartile [1st–26th March 2020] of admitted patients during the surge had a mortality of 37.3%; the second quartile [27th March–2nd April 2020], 53%; the third quartile [3rd–9th April 2020], 43.4%; and the last quartile [10th April–31st August], 35.9% (see Fig. S1, Table S8). Admission respiratory SOFA increased across the pandemic quartile ($P=0.036$). In those that died, active withdrawal of support occurred in 65% of patients (85/130), in the 13 sites which reported, and unanticipated cardiac arrest occurred in 11% of patients (13/122). There was an increased rate of reported withdrawal of life support in patients admitted during the second and third quartiles of the surge (first quartile, 55.9%; the second quartile, 73.8%; the third quartile, 71%; and the last quartile, 56.5%; $P=0.018$). Patients who had life support withdrawn had a median age of 64(57–70) years, a length of mechanical ventilation of 11 (6–18) days; a last PaO_2/FiO_2 of 12.8 (10–19.5) kPa and had a higher application of prone intervention (72%). Median PaO_2/FiO_2 in non-survivors on the day of death was 12.3(8.9–18.4)kPa.

Our multivariate model showed clinical variables on ICU admission independently associated with mortality were higher age (HR 1.95 per decade, 95% CI 1.58–2.4), male gender (HR 2.05, 95% CI 1.17–3.61), higher lactate (HR 1.52 per quartile (0.6 mmol/L), 95% CI 1.21–1.92), and higher SOFA coagulation score (HR 1.95, 95% CI 1.17–3.26) (Fig. S2; Table S9). Over the first week, statistically significant interaction differences were noted in the group-wise ANOVA between survivors and non-survivors within several respiratory, inflammatory and coagulation parameters (Fig. S2; Table S10). Machine learning models using admission data predicted mortality with 60% accuracy. Predictive capacity increased to 74.5% and 76.3% accuracy, respectively, when longitudinal data from the first week were added to LR and 3MLP models (Fig. 3). Critically, using Explainability AI methods, we were able to identify key clinical parameters which started at relatively low importance at admission but then greatly increased and exceeded others in importance over the first week (Fig. 3): these were lower PaO_2/FiO_2, higher peak pressure, higher ventilatory ratio (VR), lower pH, higher lactate, lower platelet count, higher C-Reactive Protein (CRP), lower oxygen saturations, and higher $PaCO_2$ (see Fig. S3).

Determinants of oxygenation

Movement across hypoxaemia severity groups (mild, moderate and severe PaO_2/FiO_2 group) showed deterioration in 31.4% of cases, stasis in 45.1%, and resolution in only 23.5% of patients over the first 7 days (Fig. 4 and Table S11). Overall, progression to a worse PaO_2/FiO_2 group occurred in twice the number of patients as compared to pre-COVID studies of ARDS (Table S11). ICU mortality in those who did not resolve hypoxaemia within the first week was significantly higher than those that did (60.4% versus 17.6%; $P<0.001$; Fig. S4). Admission and time-course differences between resolvers and non-resolvers in demographic, ventilatory, physiological, and laboratory parameters are shown in Fig. S4 and Tables S12 and S13. Resolvers were younger [57 (47–64) vs 60 (54–67) years; $P<0.001$] and showed a longer duration of symptoms prior to ICU admission 9.0 (7–14) vs 7 (6–11) days ($P=0.004$). Multivariate regression showed that increased age and worse cardiovascular SOFA were associated with deteriorating hypoxaemia within the first week of IMV (Fig. S4; Table S14).

ICU management

The application, median start date and duration of the first episode of each intervention and for each site is shown in Figs. S5, S6 and Table S15. The reported ideal body weight overestimated our calculated ideal body weight derived from reported height (http://ardsnet.org) in 92.6% of patients (Fig. S7). Hence, median tidal volume per kg on actual ideal body weight was 7.0 [IQR 6.0–8.4] mL/kg across all breaths and 5.6 [IQR 4.7–6.6] mL/kg on reported ideal body weight. Survivors and non-survivors showed the same distribution of tidal volume variation. Over 65% of reported PEEP values were set outside $\pm 1 cmH_2O$ and 53% set outside $\pm 2 cmH_2O$ of the ARDSNet PEEP-FiO_2 tables (Fig. S7). Patients with BMI < 40 had a higher set PEEP than recommended by the PEEP-FiO_2 table. In contrast, patients with BMI > 40 had a lower set PEEP than recommended by the PEEP-FiO_2 table. Inhaled nitric oxide and prostacyclin were commenced on day 6 (3–9) and 7 (3–15) and were continued for 4 (2–7) days and 3 (1–7) days, respectively. Tracheostomy was performed in 29% at a median 14(9–18) days in patients mainly likely to survive (40% versus 10.9%; $P<0.001$). Application of high PEEP, NMBA, and prone position was significantly higher during the second and third quartiles (Table S8). Corticosteroid usage increased across the surge whereas use of diuresis reduced (Table S8).

Responsiveness to open lung and prone interventions

Changes in PEEP were widespread over the first 7 days of IMV with both increases and decreases leading to unpredictable changes in PaO_2/FiO_2 (Fig. S7). We analysed the immediate change in PaO_2/FiO_2 over 36 h around the first prone intervention. Indeed, there were both positive and negative changes in PaO_2/FiO_2 in response to prone intervention over the first 36 h (Fig. S8). Improvements in oxygenation in response to prone position was found

Table 2 Interventions according to severity of hypoxaemia on admission

Adjunctive interventions	Number of completed records *N*	Total number in which intervention applied median [IQR] / *N* (%)	$PaO_2/FiO_2 > 26.7$ Group N	$PaO_2/FiO_2 > 26.7$ median [IQR] / *N* (%)	PaO_2/FiO_2 13.3–26.6 Group *N*	PaO_2/FiO_2 13.3–26.6 median [IQR] / *N* (%)	$PaO_2/FiO_2 < 13.3$ Group *N*	$PaO_2/FiO_2 < 13.3$ median [IQR] / *N* (%)	*p*-value (KruskalWallis/ Chi2)
Label									
Was patient transferred in from another hospital?	633	159 (25.1%)	147	31 (21.1%)	320	86 (26.9%)	166	42 (25.3%)	0.407
PEEP > 10	622	459 (73.8%)	143	73 (51%)	314	242 (77.1%)	165	144 (87.3%)	< 0.001
Neuro-muscular blockade	617	434 (70.3%)	139	74 (53.2%)	314	222 (70.7%)	164	138 (84.1%)	< 0.001
Inhaled nitric oxide	521	73 (14%)	121	10 (8.26%)	283	36 (12.7%)	117	27 (23.1%)	0.032
Inhaled prostacyclin	521	55 (10.6%)	121	10 (8.26%)	283	34 (12%)	117	11 (9.4%)	0.216
Bronchoscopy	521	51 (9.79%)	121	7 (5.79%)	283	29 (10.2%)	117	15 (12.8%)	0.246
Renal replacement therapy	555	211 (38%)	126	44 (34.9%)	297	110 (37%)	132	57 (43.2%)	0.607
Diuretics	598	443 (74.1%)	134	95 (70.9%)	311	235 (75.6%)	153	113 (73.9%)	0.128
Corticosteroids	582	304 (52.2%)	130	62 (47.7%)	307	153 (49.8%)	145	89 (61.4%)	0.129
Therapeutic heparin	438	55 (12.6%)	99	11 (11.1%)	242	34 (14%)	97	10 (10.3%)	0.195
Antibiotics	313	219 (70%)	63	39 (61.9%)	173	126 (72.8%)	77	54 (70.1%)	0.021
Proning									
Prone Positioning	551	273 (49.5%)	127	35 (27.6%)	291	152 (52.2%)	133	86 (64.7%)	< 0.001
No of prone periods	273	1 [1-2]	35	1 [1-2]	152	1 [1-2]	86	1 [1-2]	0.177
Mechanical ventilation prior first prone event (days)	273	3 [1.8-6]	35	4 [2-7.8]	152	4 [2-7]	86	2 [1-4]	0.005
Duration of first prone period (days)	273	2 [1-4]	35	2 [1-4]	152	2 [1-3.5]	86	2 [1-4]	0.913
Oxygentaion responder to prone position	270	119 (44.1%)	35	23 (65.7%)	150	71 (47.3%)	85	25 (29.4%)	0.087
Missed opportunities prior to prone	273	3 [1-6]	35	0 [0-5]	152	3 [1-7]	86	2 [1-5]	0.014
Missed opportunities in unproned	360	3 [1-10]	112	1 [0-3]	168	4 [1-10]	80	7.5 [3-16]	< 0.001
Tracheostomy									
Tracheostomy	516	145 (28.1%)	117	38 (32.5%)	282	80 (28.4%)	117	27 (23.1%)	0.059

Table 2 (continued)

Adjunctive interventions	Number of completed records	Total number in which intervention applied	$PaO_2/FiO_2 > 26.7$		PaO_2/FiO_2 13.3–26.6		$PaO_2/FiO_2 < 13.3$		
	N	median [IQR] / *N* (%)	Group N	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	Group *N*	median [IQR] / *N* (%)	*p*-value (KruskalWallis/ Chi2)
Mechanical ventilation prior to tracheostomy (days)	145	13 [9-18]	38	14 [10-19]	80	14 [8.5-19]	27	11 [6.3-15]	0.050
Duration of tracheostomy (days)	145	14 [6-21.3]	38	14 [11-22]	80	15 [7-21.5]	27	10 [3.3-19.5]	0.122

to decrease the later the prone episode was initiated after intubation (Fig. S8; Spearman r = − 0.16, $P = 0.012$). Patients that resolved hypoxaemia in the first week had prone position applied significantly earlier (2 [1–5] vs 4 [2–7] days; $P = 0.007$) than those that did not resolve. Importantly, in those that received no prone position, there were a higher number of missed opportunities to prone in non-resolvers compared to resolvers (6 [3–13] versus 1 [0–4] opportunities per patient; $P < 0.001$; Table S12).

Only 44.4% of patients maintained a mean $PaO_2/FiO_2 > 20$ kPa over 7 days after the initiation of prone position. Mortality was significantly higher in prone non-responders than in responders (69.5% versus 31.1%, $P < 0.001$ as seen in Fig. 5 and Table S16). Time series analysis showed that non-responders showed worse mean airway pressure, worse oxygenation index (OI), higher platelet count and higher alkaline transaminase (ALT) over the first week of prone position (Fig. S8 and Table S17). Multivariate analysis showed non-responders to be older with a higher pre-pronation peak pressure (OR 1.42[1.06–1.91]; $P < 0.05$), higher respiratory component (OR 1.71[1.17–2.5]; $P < 0.01$) and higher cardiovascular component (OR 1.36[1.04–1.75]; $P < 0.05$) of the sequential organ failure assessment (SOFA) score and raised lactate (OR 1.33[0.99–1.79]; $P = 0.057$) (Fig. S8 and Tables S18 and S19). Whilst there were no significant differences in the duration of IMV prior to the first prone period, the duration of the first period, or the number of future prone periods between responders and non-responders; non-responders had a higher number of missed prone opportunities (prior to first prone position event) than responders (3 [1–7] versus 2 [1–5] opportunities per patient; $P < 0.05$; Table S16).

Clinical implementation gap in proning interventions

The application of prone position occurred in 49.5% of patients and was applied on day 2 (1–5) and lasted 2 (1–4) days. Prone position was applied earlier in patients with greater severity on admission [mild: 4 (2–8) days; moderate 4 (2–7) days; severe: 2(1–4) days after onset of IMV; $P < 0.001$]. While patients that did not undergo prone position may overall have had a milder disease, we found that 76% of patients who had moderate hypoxaemia and 46% who had severe, at any stage of admission, did not undergo prone position at all. We measured the opportunity to apply prone position when there was a $PaO_2/FiO_2 < 20$ kPa, with an $FiO_2 \geq 0.6$, and a PEEP $\geq 5 cmH_2O$, as per the PROSEVA study [20]. In patients who received no prone positioning, there was 1 (IQR 0–2) prone opportunity per patient ignored during the first 48 h and 3 (IQR 1–10) during the whole patient journey. In patients who received prone interventions,

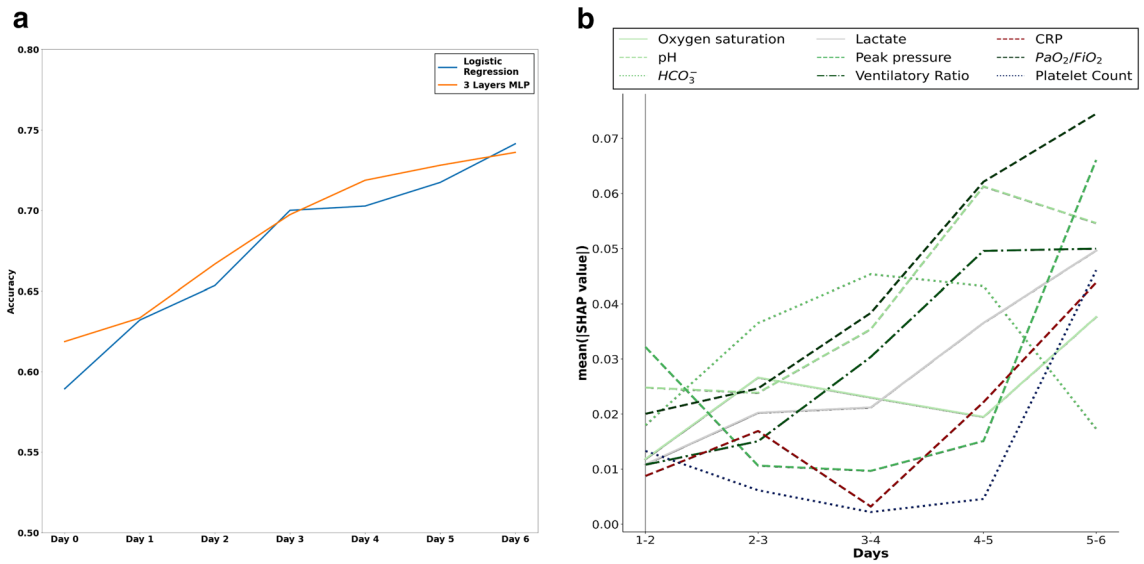

Fig. 3 **a** Week-long day-by-day performances of our logistic regression (LR) and three-layer multilayer perceptron (3-MLP) model. As shown in the figures, our models' performances rise daily from admission to the end of the first week. Daily accuracy for mortality prediction are: [Log-Reg, 3MLP]: Day 0—[58.9%, 61.9%], Day 1—[63.2%, 63.3%], Day 2—[65.3%, 66.7%], Day 3—[70%, 69.7%], Day 4—[70.3%, 71.9%], Day 5—[71.7%, 72.8%], Day 6—[74.1%, 73.6%]. **b** The importance of clinical parameters for the week-long mortality prediction model changes day-by-day. We are showing here the evolution of the SHAP (SHapley Additive exPlanations) importance of the clinical variables from admission (Day 1–2) to end of the first week, particularly PaO_2/FiO_2, higher peak pressure, higher ventilatory ratio, lower pH, higher lactate, lower platelet count, higher CRP, lower oxygen saturations, higher PEEP and higher $PaCO_2$. Note, how some of the variables are systematically high importance or low importance while others are systematically increasing from day 1 and vice versa, see Fig. S4 for details

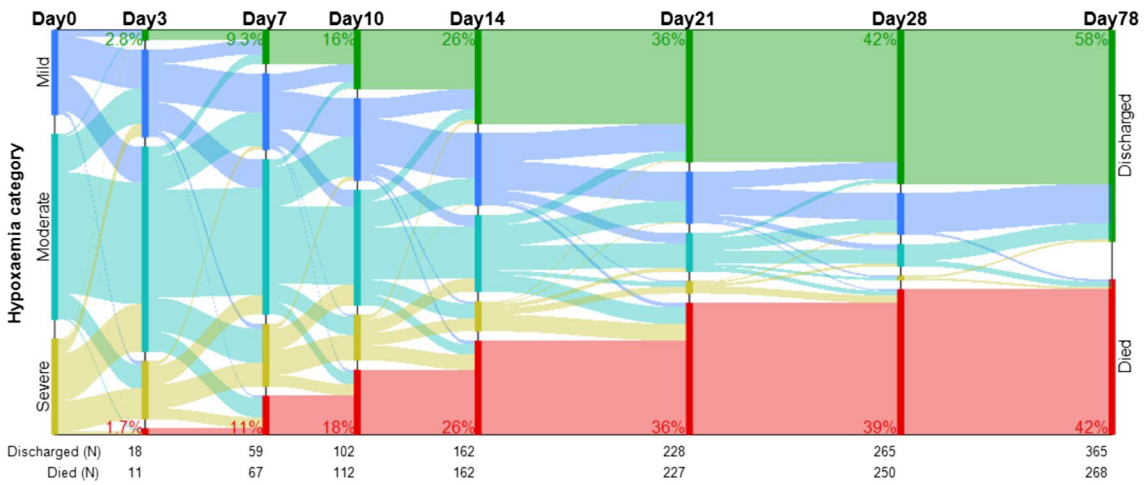

Fig. 4 Alluvial diagram of patient movements between ARDS severity groups: Mild hypoxaemia ($PaO_2/FiO_2 > 26.7$), Moderate hypoxaemia (PaO_2/FiO_2: 26.6–13.3), and severe hypoxaemia ($PaO_2/FiO_2 < 13.3$) and patient outcome (discharged—green, deceased—red). Each solid bar represents an ARDS severity group at a given number of days since initiation of IMV. Shaded coloured streams between bars represent transitions of patients between the severity groups from one-time point to the next, which is either their new severity or their outcome. The height of the bars represents the proportion of patients at that time point (i.e., they stack up to 100%) and the height of a stream field represents the size of the components contained in both bars connected by the stream

there were on average 3 (IQR 1–6) prone opportunities per patient before prone initiation that were missed. There was no difference in the number of prone sessions between survivors and non-survivors, however, patients who died without receiving prone position had a greater number of missed prone opportunities [7 (3–15) versus

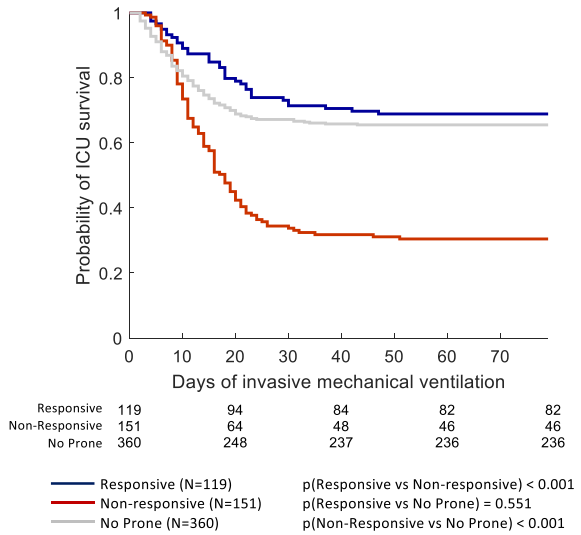

Fig. 5 ICU survival curves for the prone responder (blue) versus non-responder (red) versus patients who received no prone position (grey). Responsiveness to prone position with responders defined as maintenance of a mean $PaO_2/FiO_2 > 20$ kPa over 7 days after the first prone episode

2(0–6); $P < 0.001$; Table S7]. Patients admitted before the peak of the surge had a lower application of prone position, a greater duration of IMV prior to application of first prone position and a tendency towards having more missed prone opportunities.

Discussion

We describe the natural longitudinal history of critically ill COVID-19 patients undergoing invasive mechanical ventilation (IMV). Mortality was 43.3%, consistent with described mortality from IMV from the UK intensive care national audit and research centre (ICNARC; 47.7%) [21]. The median PaO_2/FiO_2 in non-survivors on the day of death was 12.3 (8.9–18.4)kPa suggesting many patients died with refractory hypoxaemia. This was associated with variable application of, and non-responsiveness to, ARDS interventions such as high PEEP and prone position. Evidence-based ARDS measures were imperfectly implemented, with inaccurate tidal volume calculations and missed prone positioning opportunities noted. Non-pulmonary clinical factors were associated with a lack of response to prone positioning, suggesting a role for a wider diagnostic assessment. Our machine learning models highlight the importance of including longitudinal week-long data to more accurately assess mortality prognostication. We then used Explainable AI to look under the hood of the machine learning models by computing SHapley Additive exPlanations for all clinical parameters for each day of ICU stay. This approach suggests that the focus of attention should shift over the course of the first week after admission to specific clinical parameters (such as PaO_2/FiO_2) which increase in predictive importance.

Progression of COVID-19 respiratory failure in ICU

Trajectory in terms of severity of ARDS and oxygenation is not only dependent on cardiopulmonary factors (e.g. pulmonary consolidation, thrombosis, fibrosis and right ventricular compromise) but also responsiveness to interventions (many of which aim to reduce ventilator-induced lung injury). The longitudinal natural history shows key modalities associated with pulmonary dysfunction, i.e. PaO_2/FiO_2, VR and peak pressure, had the highest importance in predicting mortality across the entire first week. Patients with progressive hypoxaemia over the first week suffered a mortality of 59.4% versus 16.3% in those that resolved hypoxaemia. Over 75% of patients remained in either static or worse hypoxaemic categories, despite an increased application of adjunctive ARDS interventions suggesting that many patients were refractory to traditional ARDS interventions, ultimately dying with refractory hypoxaemia. We hypothesise that this progressive gas exchange failure (hypoxaemia and hypercarbia) observed in COVID-19 may be due to the immunothrombotic nature of the disease pathophysiology, with increasing clot burden to the lung and subsequent right heart dysfunction in patients that progress and show reduced responses. We recently showed in COVID-19, that right ventricular fractional area change (FAC) and ventricular-pulmonary artery coupling (as measured by FAC:Right ventricular systolic pressure (RVSP) ratio) correlated significantly not only with troponin, BNP and pulmonary vascular resistance but also with measures of ventilation (namely PEEP and PaO_2/FiO_2) and a liver marker of congestion (ALT) [22]. A further determinant of hypoxaemia trajectory (in addition to underlying disease processes) could be the evolving interplay between timing, application and responsiveness of ARDS interventions which protect the lung from ventilator-induced lung injury (VILI) [23].

Responsiveness to ARDS interventions in COVID-19

We identified four key points that challenged our assumptions and inform management about COVID-ARDS during the pandemic: (1) changes in PEEP did not equate to improvements in PaO_2/FiO_2, suggesting other approaches to PEEP titration are needed, e.g. electrical impedence tomography [24] or recruitment/inflation index [25]; (2) over half of the patients who underwent prone positioning did not maintain a sustained response in PaO_2/FiO_2 over the following week. Multivariate analysis showed that patients with a higher peri-pronation

lactate, $PaCO_2$, peak pressure and worse cardiovascular and respiratory SOFA had a worse oxygenation response to prone position; (3) responsiveness to prone positioning decayed with a longer duration of IMV prior to the first prone position intervention. Solutions could include the earlier implementation of prone position either, immediately after or even prior to intubation for less severe states.; (4) half of patients with severe ARDS did not have prone interventions applied at all and those which did had a significant number of missed opportunities when it could have been applied earlier. Those that resolved hypoxaemia in the first week underwent prone position on average 2 days earlier than those that did not and, prone non-responders had many opportunities to receive prone positioning earlier in their course.

Opportunities for rapid improvements in mortality

This evaluation aims to recognise this gap in the implementation of the current ARDS evidence base and enable real-time feedback during a pandemic. The mortality for patients receiving IMV in our cohort is considerably higher than reports from other countries [2–4] and these outcomes may reflect existing clinician prognostication biases rather than prognostic characteristics in the "natural course" of COVID-19 ARDS. We show increased mortality in patients admitted during the peak of the surge with many patients dying with (and possibly as a result of) severe hypoxaemia refractory to many interventions. Other reports with lower mortality also show a greater application of ARDS interventions e.g. prone position (70% [2] and 76% [3] versus 50% in our study), and earlier in the disease process. Whilst prone positioning improves oxygenation, there are conflicting reports as to whether this physiological response equates to improved mortality [26, 27]. It is important to consider if the excess mortality in our cohort is secondary to worse VILI (as a result of non-application of therapies in overwhelmed heath systems during the pandemic) which has been shown to have a causal association to mortality [28]. Patients in our cohort that showed improvements in oxygenation with prone position also showed improvements in $PaCO_2$, OI, VR, and lower peak pressures. Oxygenation should not be used as a standalone measure of response to proning as survival benefit is likely a non-linear interaction between improved ventilation/perfusion matching, more homogenous distribution of lung stress and lung strain with lower VILI, and reduced loading and strain of the right ventricle [29]. We chose PaO_2/FiO_2 as it is the main criterion for starting, terminating, and assessing response to ARDS interventions [30]. Additionally, it also shows a strong correlation to OI (Fig. S9) which has been suggested over the first 7 days to predict the failure of interventions in clinical trials [31]. Most patients received lung-protective ventilation with tidal volumes less than 8 mL/kg and plateau pressures less than 30 cmH_2O. This was despite systematic errors in measurement of height and derived ideal body weight. However, PEEP was set higher than the low PEEP ARDSNet table and conversely lower than advised by the high PEEP ARDSNet table.

Mortality assessment across the pandemic

While many reasons may exist for differences in mortality between countries (e.g. illness severity, healthcare dilution from large numbers etc.), our data show ICU mortality being highest during the busiest period for admissions around the peak of the surge (second and third quartiles between 27th March and 9th April 2020). Hence, a poor implementation of proven evidence-based ARDS interventions during pandemic surge may have contributed to this higher mortality (e.g. systems-related or lack of clinical awareness, judgement that hypoxaemia is not severe enough or cardiovascular instability). In a pandemic, there may have been other workload pressures including inadequate staffing and training where the healthcare system is overwhelmed. Prior to COVID-19, the most common explanation for why prone position was not applied was oxygenation not being sufficiently impaired with application reserved for rescue therapy for severe hypoxemia [32]. We examined national data for the period analysed in this study and found that approximately 20% of IMV patients with COVID-19 (1596 of 7874 patients) were referred for advice to the NHS England severe acute respiratory failure ECMO service with approximately only 4% (306 patients) fulfilling new acceptance criteria [33] for ECMO in the United Kingdom (data from NHS England commissioned service).

Strengths and limitations

A key strength of this study was to take a longitudinal view and evaluate trajectories based on granular clinical ICU data. These insights enabled by our use of ML and XAI methods complementing standard techniques, helped us separate human bias (from pre-COVID ARDS) from objective, data-derived analysis on actual COVID-ARDS. There are limitations of this service evaluation, not least its observational, retrospective nature. While predictive models traditionally use a derivation/validation model, this is not applicable in the setting of an evolving pandemic and hence, we used within cohort separation (70/30) as discussed in the methods. Furthermore, the impact of overwhelmed healthcare systems during surge on variables and progression remains uncertain. We chose to focus on patients undergoing invasive mechanical ventilation as this remains a key defining criteria for admission to ICU as well as active treatment [34, 35]. We

opted for a twice-daily collection of data in contrast to a worst daily value, to appreciate the overall progression of disease and impact of complex interventions, but also achieve a pragmatic balance with ease of data collection for sites. In view of this, an important caveat to our analysis of prone position is that we were unable to accurately collect the duration of each prone position event, which also determines responsiveness [19]. We took a pragmatic approach to define responsiveness to prone position using PaO_2/FiO_2 as a surrogate. This has limitations but is clinically relevant for implementation and termination of prone position [30]. However, other factors may influence PaO_2/FiO_2 further along the ICU stay, such as superinfection (e.g. antibiotic usage in non-responders was 79.3% versus 52.9% in responders, although this was non-significant). With respect to missing data, this study evaluated routine clinical care, and hence, "missing" data is predominantly through variations in care, for instance, a site not performing the test in the first place. Finally, the collection of certain parameters may not be physiologically "pure" measurements e.g. dynamic compliance is that shown on the ventilator and not calculated [VT/(End-inspiratory plateau pressure—PEEP)].

Implications for clinical service and future research

We show in a cohort of mechanically ventilated patients with COVID-19, that a trajectory of worsening respiratory failure as a result of disease factors and a lack of responsiveness, inappropriate timing and non-implementation of ARDS interventions, is associated with worse outcome. Our XAI analyses of longitudinal disease trajectory emphasises the importance of gas exchange, respiratory mechanics, inflammation, thrombosis, haemodynamic/cardiac dysfunction (particularly reflected by the cardiovascular and coagulation SOFA components) in predicting disease progression. In those that had prone positioning applied, less than half maintained a PaO_2/FiO_2 above 20 kPa after application, and crucially, its effectiveness decayed the later it was applied. This potential lack or decay in responsiveness contrasts traditional ARDS interventions and prompt the development of studies to develop a COVID-19 specific evidence-base. While this evidence-base evolves, management may benefit from serial re-evaluation of actual disease trajectory and prognostic models due to (1) the application and impact of novel interventions (e.g. dexamethasone and tocilizumab) and (2) variations in clinical practice that may influence the implementation gap, for instance, better-prepared processes to prevent healthcare services being overwhelmed. Our data-driven approach demonstrates how a form of "standing" multicentre service evaluation could help monitor and directly inform better clinical practice and future research during the pandemic.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s00134-021-06389-z.

Author details

[1] Division of Anaesthetics, Pain Medicine & Intensive Care, Department of Surgery & Cancer, Faculty of Medicine, Imperial College London, London, UK. [2] Department of Adult Intensive Care, The Royal Brompton and Harefield NHS Foundation Trust, Sydney Street, London, UK. [3] Brain & Behaviour Lab, Dept. Of Computing, Imperial College London, London, UK. [4] Brain & Behaviour Lab, Dept. Of Bioengineering, Imperial College London, London, UK. [5] Dept. of Brain Sciences, Imperial College London, London, UK. [6] UK Dementia Research Institute Care Research and Technology Centre, Imperial College London, London, UK. [7] Department of Surgery & Cancer, Faculty of Medicine, Imperial College London, London, UK. [8] Department of Critical Care, Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [9] Department of Critical Care, Guy's and St Thomas' NHS Foundation Trust, St Thomas' Hospital, London, UK. [10] Department of Critical Care, Aintree University Hospital Foundation Trust, Liverpool, UK. [11] Department of Anaesthetics and Critical Care, The Royal Marsden NHS Foundation Trust, London, UK. [12] Department of Critical Care, Liverpool University Hospitals NHS Foundation Trust and University of Liverpool, Liverpool, UK. [13] Department of Critical Care, Imperial College Healthcare NHS Trust, London, UK. [14] Critical Care and Peri-Operative Medicine Research Group, William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [15] UKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, London, UK. [16] MRC London Institute for Medical Sciences, London, UK.

Acknowledgements

The authors are grateful to the patients and families of those affected by COVID-19 for whom the NHS had the privilege of delivering care during the pandemic. We are grateful to the staff of the UK National Health Service (NHS) and the National Institute for Health Research (NIHR). The data extracted for this evaluation would not have been possible without the expertise and goodwill of multiple doctors, nurses, other allied health professionals, and data coordinators across the NHS and university organizations that contributed. We thank the NHS commissioned service for severe acute respiratory failure and extracorporeal membrane oxygenation. We finally thank the Intensive Care Society for endorsing this national service evaluation. All collaborators are listed in the supplementary appendix page 3. Study Committee: Brijesh V Patel, Shlomi Haar, Rhodri Handslip, David Antcliffe, Ingeborg Welters, Luigi Camporota, John Prowle, Zudin Puthucheary, and Aldo Faisal. Writing Committee: Brijesh V Patel, Shlomi Haar, David Antcliffe, Ingeborg Welters, Barnaby Sanderson, Luigi Camporota, John Prowle, Zudin Puthucheary, and Aldo Faisal. Data Science Team: Brijesh Patel, Rhodri Handslip, Feargus Hosking-Jervis, Chaiyawan Auepanwiriyakul, Matthieu Komorowski; Shlomi Haar, Aldo Faisal. Site Contributions (*Site leads): Aintree University Hospitals NHS Foundation Trust: Emma Addie; Barbara Borgatta*; Roz Chisholm; Alexandra Crocokft; Anna Gilfedder; Peter Harding; Gladys Madzamba; Namratha Mathai; Maya Patel; Amie Reddy; Leanne Smith; Ian Turner-Bone; Lilian Wajero; Raegan Wignall; Laura Wilding. Northern health and social care Trust: (Antrim Area hospital, Causeway hospital) Paul Johnston*; Ciaran Mackle; Nicholas McNaughton; Scott Rafferty. Guys & St Thomas' NHS Foundation Trust: Gill Arbane; Luigi Camporota*; Alison Dixon; Barnaby Sanderson*. Imperial College Heathcare NHS Trust: David Antcliffe*; Mariana Ferrari Barbosa; Carlos Gomez; Abdalla Kaware, Fady Tarek Khalil, Matthieu Komorowski*, Maie Templeton, Raja Bilal Zafar. University Hospitals of Leicester NHS Trust: Michelle Craner; Sarah Edwards: Neil Flint; Graziella Isgro; Hakeem Yusuff*. Liverpool University Hospitals NHS Foundation Trust: Jaime Fernandez Roman; Oliver Hamilton; Brian Johnson; Emily Johnson; Maria Lopez Martinez; Suleman Mulla; David Shaw; Alicia Waite; Victoria Waugh; Ingeborg Welters*; Karen Williams. Newcastle upon Tyne Hospitals NHS Foundation Trust: Benjamin Evans; Lewis Gray; Sam Horne; Natalie Jones; Donna Kelly*; Rebecca Lawrance; Jennifer Partridge; Antonia Snell; Lucy Venyo; Thomas Walker. Royal Brompton and Harefield

NHS Foundation Trust: Frank Barrett; Sebastian Cabero Alonso; Masoomeh Farazdaghi; Ben Garfield; Clara Hernandez Caballero; Stephane Ledot; Teresa Lee; James Leslie; Helen McGuire; Lorraine O'Connor; Maurizio Passariello; Brijesh Patel*; Sunil Patel; Alex Rosenberg; Jennifer Thomson; Dhulamini Wijeratne. Royal Cornwall Hospitals NHS Trust: Sarah Bean; Karen Burt; Rosanna Jury; Oliver Quick; Michael Spivey*; Josh Thorns. Barts Health NHS Trust: Sana Ali; Thomas Beazer; Matilda Boa; Rosemarie Burd; Zoe Burdon; J Burgess; Natalie Cameron; A Daou; Charles Fadipe; FAJ Freer; Gayle Fung; Lindsey Iles; Aaron Jones; Agne Jovaisaite; Gabriella Knight; Veda Kudva; J Lampard; Ruby Lightfoot; Nicholas Low; Caroline Olabisi; Osama Omrani; Rebecca Pease; John Prowle*; Zudin Puthucheary*; Johnathan Mak; Aamir Saiyed; Maria Salamanca; Gauri Saxena; W Stanley; Nida Tahir; Yize Wan; Anton Weatherhead. Royal Papworth Hospital NHS Foundation Trust: Linzi Dunham; Melissa Earwaker*; Amy Gladwell. Royal Marsden Hospital NHS Foundation Trust: Thomas Bemand; Ethel Black; Samantha Coulson; Arnold Dela Rosa; Shamen Jhanji; Ravishankar Rao Baikady; Kate Tatham; Zoszka Webb*. South Tyneside and Sunderland NHS Foundation Trust: Sarah Cornell; Anthony Rostron*; Fiona Wakinshaw; Lindsey Woods.

Author contributions

Study concepts & design: BVP, SH, DA, JP, ZP, AAF. Literature search: BVP, SH, AAF. Database development: BVP, FHJ, RH, SH, TL. Data collection: All authors. Data science: BVP, SH, CA, RH, TL, KT, DA, MK, LC, JP, ZP, AAF. Statistical analysis: SH, BVP, JP, ZP, AAF. Machine learning analysis: CA, BVP, SH, AAF. Manuscript preparation: BVP, SH, RH, TL, KT, DA, MK, LC, JP, ZP, AAF. Figures: SH, BVP, CA, AAF. Online tool: JAH, AAF. Final manuscript review: All authors.

Funding

This study was supported through an internal research grant from the Imperial College London COVID-19 Research Fund to AAF and BVP; a UKRI Turing AI Fellowship to AAF, and an award from the Royal Brompton & Harefield Hospitals charity to BVP and general support from the NIHR Imperial Biomedical Research Centre. ACG is funded by an NIHR Research Professorship (2015–06-019). Some clinical sites have had indirect support from UK NIHR clinical research network. The funders had no role in the study design or writing. All authors had full access to all the data in the study. The writing committee had the sole responsibility for the decision to submit for publication.

Declarations

Conflicts of interest

BVP reports consulting fees from GSK and Faraday Health; grants from Mermaid Care A/C, grants from ESICM, grants from Royal Brompton & Harefield Charity, grants from European Commission, grants from Academy of Medical Sciences, grants from Imperial College London Covid Fund. ACG reports grants from NIHR Imperial BRC, grants from NIHR Research Professorship, during the conduct of the study; other from Bristol-Meyers Squibb, other from GSK. ZP reports honoraria for consultancy from GlaxoSmithKline, Lyric Pharmaceuticals, Faraday Pharmaceuticals and Fresenius-Kabi, and speaker fees from Orion, Baxter, Nutricia and Nestle and educational grants from Baxter and VitaFlo. AAF reports grants from Imperial College London and the UKRI, during the conduct of the study. All other authors report no conflicts of interest.

Publisher's Note

Received: 28 February 2021 Accepted: 18 March 2021
Published online: 11 May 2021

References

1. Grasselli G, Tonetti T, Protti A et al (2020) Pathophysiology of COVID-19-associated acute respiratory distress syndrome: a multicentre prospective observational study. Lancet Respir Med. https://doi.org/10.1016/s2213-2600(20)30370-2
2. Ferrando C, Suarez-Sipmann F, Mellado-Artigas R et al (2020) Clinical features, ventilatory management, and outcome of ARDS caused by COVID-19 are similar to other causes of ARDS. Intens Care Med. https://doi.org/10.1007/s00134-020-06192-2
3. Schmidt M, Hajage D, Demoule A et al (2020) Clinical characteristics and day-90 outcomes of 4244 critically ill adults with COVID-19: a prospective cohort study. Intens Care Med. https://doi.org/10.1007/s00134-020-06294-x
4. Botta M, Tsonas AM, Pillay J et al (2020) Ventilation management and clinical outcomes in invasively ventilated patients with COVID-19 (PRoVENT-COVID): a national, multicentre, observational cohort study. Lancet Respir Med. https://doi.org/10.1016/s2213-2600(20)30459-8
5. Gattinoni L, Coppola S, Cressoni M et al (2020) Covid-19 does not lead to a "Typical" acute respiratory distress syndrome. Am J Resp Crit Care. https://doi.org/10.1164/rccm.202003-0817le
6. Patel BV, Arachchillage DJ, Ridge CA et al (2020) Pulmonary angiopathy in severe COVID-19: physiologic, imaging and hematologic observations. Am J Resp Crit Care. https://doi.org/10.1164/rccm.202004-1412oc
7. Ackermann M, Verleden SE, Kuehnel M et al (2020) Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. N Engl J Med. https://doi.org/10.1056/nejmoa2015432
8. Sinha P, Calfee CS, Cherian S et al (2020) Prevalence of phenotypes of acute respiratory distress syndrome in critically ill patients with COVID-19: a prospective observational study. Lancet Respir Med. https://doi.org/10.1016/s2213-2600(20)30366-0
9. Fan E, Beitler JR, Brochard L et al (2020) COVID-19-associated acute respiratory distress syndrome: is a different approach to management warranted? Lancet Respir Med. https://doi.org/10.1016/s2213-2600(20)30304-0
10. Bos LDJ, Sinha P, Dickson RP (2020) The perils of premature phenotyping in COVID: a call for caution. Eur Respir J. https://doi.org/10.1183/13993003.01768-2020
11. Schmidt M, Hajage D, Lebreton G et al (2020) Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome associated with COVID-19: a retrospective cohort study. Lancet Respir Med. https://doi.org/10.1016/s2213-2600(20)30328-3
12. Barbaro RP, MacLaren G, Boonstra PS et al (2020) Extracorporeal membrane oxygenation support in COVID-19: an international cohort study of the Extracorporeal Life Support Organization registry. Lancet 396:1071–1078. https://doi.org/10.1016/s0140-6736(20)32008-0
13. Needham DM, Colantuoni E, Mendez-Tellez PA et al (2012) Lung protective mechanical ventilation and two year survival in patients with acute lung injury: prospective cohort study. Bmj Br Medical J 344:e2124. https://doi.org/10.1136/bmj.e2124
14. Bellani G, Laffey JG, Pham T et al (2016) Epidemiology, Patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. JAMA 315:788–800. https://doi.org/10.1001/jama.2016.0291
15. Duggal A, Rezoagli E, Pham T et al (2020) Patterns of use of adjunctive therapies in patients with early moderate to severe ARDS. Chest 157:1497–1505. https://doi.org/10.1016/j.chest.2020.01.041
16. Komorowski M, Celi LA, Badawi O et al (2018) The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 24:1716–1720. https://doi.org/10.1038/s41591-018-0213-5
17. von Elm E, Altman DG, Egger M et al (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies*. B World Health Organ 85:867–872. https://doi.org/10.2471/blt.07.045120
18. The ARDS Definition Task Force (2012) Acute respiratory distress syndrome: the berlin definition. Jama 307(23):2526–2533. https://doi.org/10.1001/jama.2012.5669

19. Guérin C, Reignier J, Richard J-C et al (2013) Prone positioning in severe acute respiratory distress syndrome. N Engl J Med 368:2159–2168. https://doi.org/10.1056/nejmoa1214103
20. Lundberg S, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions. Arxiv
21. Richards-Belle A, Orzechowska I, Gould DW et al (2020) COVID-19 in critical care: epidemiology of the first epidemic wave across England, Wales and Northern Ireland. Intens Care Med. https://doi.org/10.1007/s00134-020-06267-0
22. Bleakley C, Singh S, Garfield B et al (2020) Right ventricular dysfunction in critically ill COVID-19 ARDS. Int J Cardiol. https://doi.org/10.1016/j.ijcard.2020.11.043
23. Marini JJ, Gattinoni L (2020) Time course of evolving ventilator-induced lung injury: The "Shrinking Baby Lung." Crit Care Med 48:1203–1209. https://doi.org/10.1097/ccm.0000000000004416
24. van der Zee P, Somhorst P, Endeman H, Gommers D (2020) Electrical impedance tomography for positive end-expiratory pressure titration in COVID-19–related acute respiratory distress syndrome. Am J Resp Crit Care 202:280–284. https://doi.org/10.1164/rccm.202003-0816le
25. Pan C, Chen L, Lu C et al (2020) Lung recruitability in COVID-19–associated acute respiratory distress syndrome: a Single-Center Observational Study. Am J Resp Crit Care 201:1294–1297. https://doi.org/10.1164/rccm.202003-0527le
26. Lemasson S, Ayzac L, Girard R et al (2006) Does gas exchange response to prone position predict mortality in hypoxemic acute respiratory failure? Intens Care Med 32:1987–1993. https://doi.org/10.1007/s00134-006-0390-4
27. Albert RK, Keniston A, Baboi L et al (2014) Prone position–induced improvement in gas exchange does not predict improved survival in the acute respiratory distress syndrome. Am J Resp Crit Care 189:494–496. https://doi.org/10.1164/rccm.201311-2056le
28. The Acute Respiratory Distress Syndrome Network (2000) Ventilation with Lower Tidal Volumes as Compared with Traditional Tidal Volumes for Acute Lung Injury and the Acute Respiratory Distress Syndrome (2000). N Engl J Med 342:1301–1308. https://doi.org/10.1056/nejm200005043421801
29. Vieillard-Baron A, Charron C, Caille V et al (2007) Prone positioning unloads the right ventricle in severe ARDS. Chest 132:1440–1446. https://doi.org/10.1378/chest.07-1013
30. Guérin C, Albert RK, Beitler J et al (2020) Prone position in ARDS patients: why, when, how and for whom. Intens Care Med. https://doi.org/10.1007/s00134-020-06306-w
31. Go L, Budinger GRS, Kwasny MJ et al (2016) Failure to Improve the Oxygenation Index Is a Useful Predictor of Therapy Failure in Acute Respiratory Distress Syndrome Clinical Trials. Crit Care Med 44:e40–e44. https://doi.org/10.1097/ccm.0000000000001295
32. Guérin C, Beuret P, Constantin JM et al (2018) A prospective international observational prevalence study on prone positioning of ARDS patients: the APRONET (ARDS Prone Position Network) study. Intens Care Med 44:22–37. https://doi.org/10.1007/s00134-017-4996-5
33. Camporota L, Meadows C, Ledot S et al (2021) Consensus on the referral and admission of patients with severe respiratory failure to the NHS ECMO service. Lancet Respir Med. https://doi.org/10.1016/s2213-2600(20)30581-6
34. Docherty AB, Harrison EM, Green CA et al (2020) Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 369:m1985. https://doi.org/10.1136/bmj.m1985
35. Knight SR, Ho A, Pius R et al (2020) 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 370:m3339. https://doi.org/10.1136/bmj.m3339