



BMJ Open Biomarker identification using dynamic time warping analysis: a longitudinal cohort study of patients with COVID-19 in a UK tertiary hospital

Hannah Burke,¹ Anna Freeman,¹ Paul O'Regan,² Oskar Wysocki,² Andre Freitas ,² Ahilanandan Dushianthan,¹ Michael Celinski,¹ James Batchelor,¹ Hang Phan,^{3,4} Florina Borca,⁵ Natasha Sheard,⁶ Sarah Williams,⁶ Alastair Watson,¹ Paul Fitzpatrick,⁷ Dónal Landers ,⁸ Tom Wilkinson,¹ On behalf of the REACT COVID group

To cite: Burke H, Freeman A, O'Regan P, *et al.* Biomarker identification using dynamic time warping analysis: a longitudinal cohort study of patients with COVID-19 in a UK tertiary hospital. *BMJ Open* 2022;**12**:e050331. doi:10.1136/bmjopen-2021-050331

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-050331>).

HB, AF, P O'R and OW are joint first authors.

Received 16 February 2021
Accepted 27 January 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Anna Freeman;
a.freeman@soton.ac.uk

ABSTRACT

Objectives COVID-19 is a heterogeneous disease, and many reports have described variations in demographic, biochemical and clinical features at presentation influencing overall hospital mortality. However, there is little information regarding longitudinal changes in laboratory prognostic variables in relation to disease progression in hospitalised patients with COVID-19.

Design and setting This retrospective observational report describes disease progression from symptom onset, to admission to hospital, clinical response and discharge/death among patients with COVID-19 at a tertiary centre in South East England.

Participants Six hundred and fifty-one patients treated for SARS-CoV-2 between March and September 2020 were included in this analysis. Ethical approval was obtained from the HRA Specific Review Board (REC 20/HRA/2986) for waiver of informed consent.

Results The majority of patients presented within 1 week of symptom onset. The lowest risk patients had low mortality (1/45, 2%), and most were discharged within 1 week after admission (30/45, 67%). The highest risk patients, as determined by the 4C mortality score predictor, had high mortality (27/29, 93%), with most dying within 1 week after admission (22/29, 76%). Consistent with previous reports, most patients presented with high levels of C reactive protein (CRP) (67% of patients >50 mg/L), D-dimer (98% > upper limit of normal (ULN)), ferritin (65% > ULN), lactate dehydrogenase (90% > ULN) and low lymphocyte counts (81% < lower limit of normal (LLN)). Increases in platelet counts and decreases in CRP, neutrophil:lymphocyte ratio ($p < 0.001$), lactate dehydrogenase, neutrophil counts, urea and white cell counts (all $p < 0.01$) were each associated with discharge.

Conclusions Serial measurement of routine blood tests may be a useful prognostic tool for monitoring treatment response in hospitalised patients with COVID-19. Changes in other biochemical parameters often included in a 'COVID-19 bundle' did not show significant association with outcome, suggesting there may be limited clinical benefit of serial sampling. This may have direct clinical

Strengths and limitations of this study

- Close alignment of research and clinical practice in a near real-time manner.
- Longitudinal data collection and sampling opportunities.
- A single-centre study with data collection reflective of clinical need rather than a strict protocolised time frame.
- Use of novel artificial intelligence techniques for data analysis.
- Analysis of a 'first wave' cohort prior to approval for use of COVID-19 specific treatments, including dexamethasone.

utility in the context of escalating healthcare costs of the pandemic.

INTRODUCTION

COVID-19 is a heterogeneous disease with variable clinical outcomes, ranging from asymptomatic carriage to severe pneumonia and multi-organ failure.^{1 2} Understanding this heterogeneity and its implications for prognosis and therapeutic response is key to improving outcomes in COVID-19. Despite the rapid development of effective vaccinations,^{3 4} the clinical repercussions of SARS-CoV-2 infection are likely to continue to impact on health services, and an increased understanding of the disease is still necessary.

Large observational studies have characterised the clinical features of hospitalised patients with COVID-19² and highlighted specific risk factors associated with mortality such as older age, male sex and chronic comorbidity.^{5 6} Furthermore, robust models that predict prognosis of COVID-19 have been developed for use both in the general

population⁷ and on admission to hospital.⁸ These models have incorporated known phenotypic risk factors but also included simple clinical variables such as blood urea level and C-reactive protein (CRP). Much has been documented about the common laboratory parameters in COVID-19 and whether these can be used as diagnosis aids or to predict outcome.⁹ However, the focus has been solely on admission blood parameters, and little is known about how these parameters change during admission and whether these changes could help prognosticate.

In addition, little is known about the relationship between timing of symptom onset and presentation to hospital and subsequent disease trajectory. The WHO International Severe Acute Respiratory and emerging Infection Consortium (ISARIC) cohort reports the median duration of symptoms before admission as 4 days; however, there is a wide range of 1–8 days.⁵ The timing of presentation to hospital may have important consequences as it may relate to the underlying pathology of the disease.¹⁰ For example, initial evidence suggested that acute respiratory distress syndrome (ARDS)-associated with severe COVID-19 presents in the second week of the illness.¹¹ The emerging literature suggest that there may be two distinct but overlapping pathological subsets: the first triggered by the virus itself and the second by the host response.^{11–13} Therefore, the point at which patients deteriorate and require hospitalisation may be key to understanding the dominant pathological mechanism at play and may have important implications for treatment.

We hypothesise that timing of presentation to hospital with COVID-19 may have a bearing on the outcome of disease and may relate to an individual's COVID-19 risk profile, as defined by a weighted risk score based on the clinical and biochemical features described previously. Furthermore, identifying specific changes in common laboratory blood markers may provide additional prognostic information and aid clinicians working in the field.

Here we report the clinical characteristics, laboratory measures and outcomes of all patients presenting with SARS-CoV-2 positive swabs to a tertiary academic medical centre in Southampton (UK) from 7 March to 4 September 2020. We examine timing of presentation to hospital against individual patients' COVID-19 risk, and determine temporal changes of laboratory blood variables that predict outcome during hospital admission. The granularity of our data and longitudinal nature of the analysis adds novelty and depth to the more cross-sectional analyses in the literature to date. While dynamic time warping has been used in modelling and forecasting the number COVID-19 cases,^{14 15} we are the first, to our knowledge, to use dynamic time warping to look much more closely at the patients, aiming to discover patterns in biomarkers trajectories.

Objectives

- ▶ Compare disease progression from symptom onset to final outcome between risk groups.

- ▶ Identify features at admission to hospital that are associated with outcome.
- ▶ Identify changes in biochemical parameters from serial sampling that are associated with outcome.

METHODS

Study design and setting

As part of the Research Evaluation Alongside Clinical Treatment in COVID-19 (REACT) observational and biobanking study of ¹⁶COVID-19,¹⁷ data were collected for COVID-19 positive patients who were admitted to University Hospital Southampton between 7 March and 4 September 2020.

Participants

Patients were included in the study if they were admitted to the hospital and were confirmed positive for SARS-CoV-2 on real-time reverse transcription PCR from a nasopharyngeal swab or bronchoalveolar lavage. Patients without a definitive outcome (eg, death or discharge), either due to ongoing treatment or missing outcome data at the time of analysis, were excluded. Possible second COVID-19 infections or readmissions with no subsequent evidence of death or discharge were also excluded.

Variables

Patients' characteristics included demographics (age, sex and body mass index) and comorbidities (including asthma, chronic obstructive pulmonary disease (COPD), cardiac disease and others). The following data were collected at admission and throughout hospitalisation as part of routine clinical care. Laboratory tests included full blood count, renal profile, liver profile, CRP, ferritin, D-dimer and lactate dehydrogenase (LDH), and vital signs included blood pressure, heart rate, respiration rate and peripheral oxygen saturations. Timing, dose and duration of all treatments including corticosteroids, anticoagulants, antibiotics, antivirals and antifungals were collected.

Outcomes

The primary outcome was in-hospital mortality. Analysis of associations between biochemical parameters at admission and outcome was restricted to patients who were hospitalised for 2 or more days and had a final outcome within 28 days of admission. For analysis of changes in parameters, additional restrictions were that patients' last specimens must have been taken at least 2 days after admission and no more than 4 days prior to final outcome.

Data sources/measurement

Clinical data were captured longitudinally, with change over time treated as explicit. All data collected from the cohort in the study are kept in a highly secure contemporary encrypted data platform BCInsight (within the Clinical Informatics Research Unit, University of Southampton) that was set up in a Microsoft Data Centre in South UK. A detailed outline of study protocol and methodology is published elsewhere.¹⁷

In order to adjust analysis of mortality based on known risk factors for COVID-19, weighted risk scores were calculated for patients at admission (first available value up to

and including the day after admission) using available variables and equivalent weightings, as described previously for the 4C mortality score.⁸ Briefly, the following weightings were applied: age (50–60 years score +2, 60–70 years score +4, 70–80 years score +6, >80 years score +7); gender (male score +1); number of relevant comorbidities (1 score +1, >1 score +2); respiration rate (20–30 score +1, >30 score +2); peripheral oxygen saturation (<92% score +2); urea (7–14mmol/L score +1, >14mmol/L score +3); CRP (50–100mmol/L score +1, >100mmol/L score +2). Glasgow Coma Scale values were not included in risk score calculation, as approximately 90% of patients did not have values available. Participants were classified based on these scores using the same thresholds described previously: low risk (weighted risk score 0–3), intermediate (score 4–8), high (score 9–14) or very high (score >14) risk.⁸

Values for individual laboratory tests were classed as low/normal/high based on the thresholds defined in online supplemental table 1.

Statistical methods

Continuous data are summarised as median (IQR), and categorical data are summarised as frequency (percentage). Differences between cohorts were tested using the Kruskal-Wallis rank sum test for quantitative variables and using Pearson's χ^2 test for count data. Biochemical parameters measured at admission were compared with the last available records using Wilcoxon signed-rank test within groups stratified by outcome. Associations between parameters and outcome were investigated using logistic regression adjusted for age, number of comorbidities and gender. Changes in biochemical parameters were also tested using logistic regression adjusted for age, number of comorbidities and gender and including both parameter value as well as the difference between first and last available values. P values were adjusted for multiple testing using Holm-Bonferroni method. Trajectories of biochemical parameters were clustered using the k-means clustering algorithm together with dynamic time warping^{18 19} used as a distance metric, setting the number of clusters $k=4$, which was found to be the optimal setting for the majority of biochemical parameters on analysis using the silhouette and elbow method.^{20 21}

Missing data

Given the real-world nature of the study, there were a number of missing data points, and as this paper is mainly descriptive, we have not performed any imputation for these missing data but describe the data as they stand. For each model the number of patients may vary due to missing values.

Bias

The analysis population includes only patients hospitalised with COVID-19, and as such, is likely to be biased towards more severe COVID-19 infection and/or people at higher risk of death due to the presence of known risk factors. Treatment and intervention pathways may have evolved over the course of the study period. The analysis does not adjust for any differences in interventions between patients over the

course of the study period. The study period largely predates the finding that dexamethasone treatment improves survival; while no adjustment is made for differences in interventions, dexamethasone use in the population was low and is not considered likely to affect the results or their interpretation.

Patient and public involvement (PPI)

Patient and public involvement was sought in the design and oversight of the broader Southampton Research Biorepository, within which the sampling arm of the REACT COVID study is nested, and patient representatives were involved in the design and management of the WATCH study,²² on which the REACT COVID Database is based. For further detail on PPI involvement in the REACT COVID study, please see ref¹⁷.

RESULTS

Participants

Six hundred and fifty-one patients who had a confirmed case of COVID-19 infection were included in this analysis. Five hundred patients had a final outcome of either died or discharged. Of these, date of symptom onset was recorded for 455 patients: 96 had date of symptom onset after admission (classed as nosocomial cases) and were excluded from further analysis. The analysis population consists of 359 patients admitted after symptom onset, who had a final outcome (died or discharged) recorded.

In order to support comparisons between cohorts, weighted risk scores at admission were calculated based on weightings described previously for 4C mortality score predictor.⁸ As seen in figure 1, while weighted risk scores predict outcome with reasonable accuracy, mortality was underestimated in most cases.

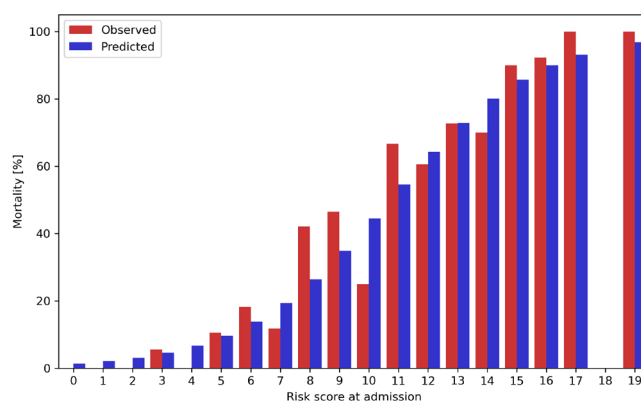


Figure 1 Performance of weighted risk score at admission as predictor of mortality. Red bars: predicted mortality based on a univariate logistic regression model of mortality according to weighted risk score at admission. Blue bars: actual observed mortality rate for a given risk score at admission.

Descriptive data

Patient characteristics

Patient characteristics are described in [table 1](#). The median age of the analysis population was 71 years (interquartile range (IQR) : 53–83). The majority of patients (60% overall) were male. The most common comorbidities (>25% overall) were cardiac disease, renal disease, obesity and diabetes. The majority of patients (208) were classified as high or very high risk at admission using the modified 4C risk score.⁸ Thirty-one patients were missing one or more of the parameters required to calculate 4C weighted risk scores.

Outcome data

Disease progression

Across all risk groups, the majority of patients were hospitalised within 1 week of symptom onset ([table 1](#)). Patients in the high and very high risk groups were admitted sooner after symptom onset compared with the low and intermediate risk groups. Time from admission to final outcome showed a bimodal relationship to risk ([figure 2](#)). Most patients in the low risk group (30/45, 67%) were discharged within 7 days of admission; most patients in the very high risk group (22/29, 76%) died within 7 days of admission. Most patients in the intermediate and high risk categories (48/75, 64% and 113/179, 63%, respectively) took longer than 7 days to reach final outcome. The longest time from admission to final outcome was 143 days.

Biochemical characteristics at admission

Based on first available values up to and including the day after admission, high CRP, D-dimer, Ferritin, LDH levels and low lymphocyte counts were common among all patients regardless of final outcome (online supplemental figure 1). The following abnormal results were more common among those who died compared with those who were discharged: high urea, high creatinine and low haemoglobin.

Abnormal values (below LLN or above ULN) for the following laboratory tests at admission were more common among higher risk groups: creatinine (eg, 40% of low-risk patients vs 86% of very high risk patients), CRP (38% vs 100%), ferritin (54% vs 95%), glucose (12% vs 68%), haemoglobin (13% vs 55%), neutrophils (18% vs 62%), urea (7% vs 97%) and white cell count (18% vs 45%). Within risk groups, differences between patients who were discharged versus died were broadly consistent (online supplemental figure 2).

Prognostic biochemical features at admission

Associations between laboratory values at admission (first available value, up to and including the day after admission) and outcome were evaluated for patients who were admitted for two or more days and had a final outcome within 28 days of admission (n=308). Potential associations with final outcome were evaluated separately for each parameter using multivariable logistic regression, adjusted for age, gender and number of comorbidities. P

values were adjusted for multiple testing using the Holm-Bonferroni method. After correction for multiple testing, only CRP:lymphocyte (p=0.011) and neutrophil:lymphocyte (p=0.0189) ratios at admission were significantly associated with outcome, with higher values associated with increased mortality (online supplemental table 2). Associations between the temporal change of laboratory values from admission (first available value, up to and including the day after admission) to the nearest available result to outcome were evaluated for patients who were admitted for 2 or more days and had a final outcome within 28 days of admission, and the last result was taken no more than 4 days prior final outcome (n=209). After correction for multiple testing, increases in CRP, LDH, neutrophils, neutrophil:lymphocyte ratio, urea and white cell count were each significantly associated with increased mortality, whereas increases in platelets were associated with reduced mortality. For example, a 1 mg/L increase in CRP was associated with approximately 2%–4% increase in the odds of death and vice-versa for a 1 mg/mL decrease ([table 2](#)).

Unsurprisingly, among patients who were discharged, CRP levels at discharge were significantly lower than at admission (p<0.001), whereas among those who died, CRP levels at death were significantly higher than at admission (p<0.001). Similar patterns were observed for CRP:lymphocyte ratio and neutrophil counts ([figure 3](#)). Platelet counts at discharge were higher than at admission (p<0.001) but were unchanged among patients who died. However, decreased urea, white cell counts and neutrophil:lymphocyte ratios were seen at discharge compared with admission, in patients who were discharged but not those who died.

Timing of prognostic changes in biochemical parameters

Clusters of patients with similar trajectories for a given biochemical parameter over the first week after admission were identified using dynamic time warping. Low, stable CRP (below 100 mg/L) was associated with reduced mortality (cluster 3, 31% mortality) compared with high and/or rising CRP (eg, cluster 2, 69% mortality, [figure 4A](#)). Low and stable urea levels were associated with reduced mortality (cluster 2, 18% mortality), whereas high and/or rising urea was associated with increased mortality (eg, cluster 4, 92% mortality, [figure 4B](#)).

In contrast, low and stable platelet counts were associated with increased mortality (eg, cluster 2, 59% mortality) compared with high and/or rising platelet counts (eg, cluster 1: 23% mortality; cluster 3: 30% mortality) ([figure 4C](#)). Increasing LDH values were associated with higher mortality (cluster 3, 67%) compared with low and stable LDH (clusters 1 and 4, 35% and 17%) or declining LDH (cluster 2, 33%, [figure 4D](#)). Neutrophil counts that persisted above normal range were associated with higher mortality ([figure 4E](#)). No obvious relationships between patterns of white blood cell counts and lymphocyte counts were observed ([figure 4F,G](#)).

Table 1 Known risk factors at admission and disease progression, categorised into risk groups based on the modified 4C risk score⁸

	Risk group at admission*						P value (unadjusted)	Adjusted p value†
	Total n=359	Data missing n=31	Low n=45	Intermediate n=75	High n=179	Very high n=29		
Age, median (IQR)	71 (53–83)	53 (32.5–61.5)	42 (34–46)	57 (48.5–64.5)	79 (72–86)	83 (80–90)	<0.001§	<0.001
Male, n (%)	217 (60)	19 (61)	27 (60)	38 (51)	110 (61)	23 (79)	0.116¶	>1
Comorbidities, median (IQR)	2 (1–4)	1 (0–2)	0 (0–1)	1 (0–3)	3 (2–5)	4 (3–5)	<0.001§	<0.001
Cardiac disease, n (%)	106 (30)	4 (13)	0	9 (12)	79 (44)	14 (48)	<0.001¶	<0.001
COPD, n (%)	68 (19)	2 (6)	0	9 (12)	47 (26)	10 (34)	<0.001¶	<0.001
Diabetes, n (%)	97 (27)	6 (19)	2 (4)	17 (23)	62 (35)	10 (34)	<0.001¶	0.014685
Dementia, n (%)	20 (6)	1 (3)	0	2 (3)	14 (8)	3 (10)	0.1276¶	>1
HIV, n (%)	4 (1)	2 (6)	0	1 (1)	1 (1)	0	0.0522¶	>1
Cancer, n (%)	17 (5)	0	0	3 (4)	10 (6)	4 (14)	0.05122¶	>1
Neural disease, n (%)	89 (25)	4 (13)	1 (2)	9 (12)	66 (37)	32	<0.001¶	<0.001
Obesity, n (%)	96 (27)	6 (19)	7 (16)	21 (28)	54 (30)	8 (28)	0.2997¶	>1
Renal disease, n (%)	100 (28)	5 (16)	0	4 (5)	74 (41)	17 (59)	<0.001¶	<0.001
Thromboembolism, n (%)	1 (<1)	0	0	1 (1)	0	0	0.4341¶	>1
O ₂ saturation*, median (IQR), %	96 (94–97)	97.5 (96–99.25)	97 (96–98)	96 (94–97)	96 (94–97)	94 (91–96)	<0.001§	<0.001
Respiration rate*, median (IQR)	20 (18–23)	22.5 (18–30.5)	19 (18–20)	20 (18–22.5)	20 (18–22)	24 (22–30)	<0.001§	<0.001
CRP*, median (IQR)	88 (31–158.8)	168.5 (42.3–236)	33 (14–66)	105 (33.5–164)	82 (31–141.5)	158 (112–240)	<0.001§	<0.001
Urea*, median (IQR)	7 (4.9–11.8)	5.25 (3–7.8)	4.4 (3–5.4)	5.2 (4.15–6.4)	9.4 (6.9–13.8)	19.8 (15–23.7)	<0.001§	<0.001
Deaths, n (%)	149 (42)	6 (19)	1 (2)	14 (19)	101 (56)	27 (93)	<0.001¶	<0.001
Discharges, n (%)	210 (58)	25 (81)	44 (98)	61 (81)	78 (44)	2 (7)	<0.001¶	<0.001
Days symptoms to admission, median (IQR)	4 (1–8)	5 (1–8.5)	6 (2–9)	7 (3.5–11)	3 (1–7)	4 (1–7)	<0.001§	0.015169
Days admission to outcome, median (IQR)	8 (4–16)	3 (0–10.5)	4 (1–9)	9 (6–19.5)	10 (6–20.5)	4 (2–7)	<0.001§	<0.001
Days symptoms to outcome, median (IQR)	15 (8–24)	10 (3–21)	12 (7–19)	17 (12–26)	16 (9–30)	8 (6–16)	<0.001§	<0.001

*Based on first available values within 1 day after admission. Participants were classified based on these scores as either: low risk (weighted risk score 0–3); intermediate (score 4–8); high (score 9–14); or very high (score >14) risk; .

†Features needed to derive risk score at admission were missing for some patients.

‡Adjusted for multiple testing using Bonferroni correction.

§Kruskal-Wallis rank sum test.

¶Pearson's χ^2 test.

CRP, C reactive protein.

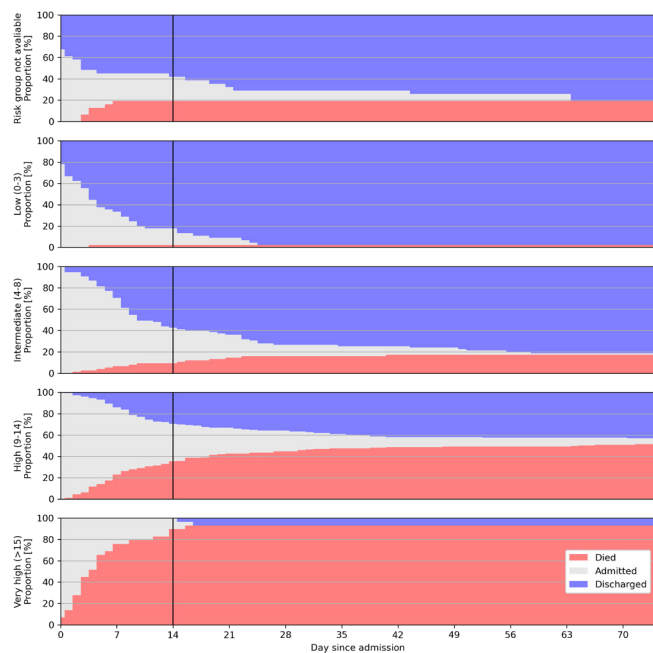


Figure 2 Time from admission to outcome according to risk group at admission. Patients were grouped based on weighted risk scores at admission: low (0–3), intermediate,^{4 5 7–9} high^{10–15} and very high (>14). Risk scores at admission could not be calculated for some patients due to one or more missing features ('data missing' group).

DISCUSSION

Key results

We apply a dynamic time warping approach to deeply understand clinical characteristics and biomarker trajectory patterns of patients with COVID-19. We demonstrate that increasing comorbidity and age are associated with and predictive of adverse outcome with SARS-CoV-2 infection, in line with the 4C risk score,^{8 23 24} providing reassurance of the translatability of our findings from a small, single centre cohort, despite the omission of Glasgow Coma Scale. Through our analysis, we highlight that serial measurement of some routine blood tests may be a useful prognostic tool for monitoring treatment response in hospitalised patients with COVID-19. However, others may provide limited clinical benefit, which may have important implications for resources and healthcare costs within the pandemic.

Symptom onset has a role in increasing understanding of the natural history of SARS-CoV-2 infection, yet many studies focus on day of hospital admission rather than symptom onset for calculation of time to outcome.²⁵ We demonstrate that the majority of patients requiring hospitalisation for SARS-CoV-2 infection present to hospital within 7 days of symptom onset. This is in line with findings from a Belgian study, demonstrating a mean time from symptom onset to hospitalisation of 5.74 days.²⁶ A potential explanation for this could be found in the description of disease phases in SARS-CoV-2 infection, with rapid progression through early infection, pulmonary phase to hyperinflammation stage driving

presentation relatively early in their disease course for those patients who require hospitalisation.²⁷ This may be driven by a higher viral load in those patients who develop severe disease more rapidly, with support for this hypothesis coming from work demonstrating a higher and more persistent viral shedding in those patients with severe disease when compared with those with mild disease.²⁸ In terms of time to outcome in those patients requiring hospitalisation, those with a low-risk score or a very high-risk score reach their outcome generally within 7 days of admission. This is reflective of the literature, in terms of meta-analyses findings of a median length of stay of 5 days outside of China.²⁵ This is likely due to those in the low-risk group having sufficient physiological reserve to overcome the infection rapidly and those in the very high risk group having minimal reserve and therefore succumbing quickly, due to a combination of changes in the immune cell repertoire, epigenome and inflammasome response to infection.²⁹ Those in the medium-risk and high-risk groups demonstrate a more variable and prolonged time to outcome, highlighting the groups in which the risk scores may be less accurate in predicting outcome. This is of clinical relevance as healthcare services become more stretched as the current wave progresses.

Consistent with previous reports, most patients presented with high levels of CRP, D-dimer, ferritin and LDH and low lymphocyte counts.^{30 31} However, when adjusted for repeated measures, age, comorbidity and how unwell patients were at admission, only neutrophil lymphocyte ratio was predictive of outcome. The literature is suggestive of a greater range of tests being predictive of outcome, and this may be a result of our more complex statistics correcting for confounders more accurately. It may also be that within a relatively small sample size, these additional predictive tests did not reach statistical significance. The neutrophil:lymphocyte ratio as predictive of outcome raises interesting questions as to the pathobiology driving morbidity and mortality in SARS-CoV-2 infection. It may be that the rise in neutrophils is associated with subclinical bacterial coinfection, which is supported by a higher CRP:lymphocyte ratio at admission and rising through the course of admission predicting worse outcome. However, a number of reviews have demonstrated detectable bacterial infection to be relatively low.^{32–34} The increased neutrophil:lymphocyte ratio may also be a direct response to the combination of viral infection and increased inflammation that are seen with worse outcomes with SARS-CoV-2 infection¹³ and have been demonstrated in other human coronavirus infections.³⁵ The viral response is thought to directly drive increasing neutrophil numbers and the reduction in lymphocytes thought to be a subsequent response to resulting oxidative stress and inflammation.³⁶

The granularity of our data allowed an in-depth investigation of change over time that adds novelty to this work. During the course of hospitalisation, changes in CRP, LDH, neutrophil counts, neutrophil:lymphocyte ratio, urea, white cell count and platelet counts were all

Table 2 Prognostic biochemical changes between admission and outcome

Parameter	Units	Total, n	Deaths, n	Discharges, n	OR for mortality per unit increase (95% CI)	P value	Adjusted p value
ALT increase	U/L	158	55	103	1.003 (0.996 to 1.01)	0.369	1.000
AST increase	U/L	56	14	42	1.023 (1.002 to 1.044)	0.029	0.264
Bilirubin increase	µmol/L	162	56	106	1.147 (1.023 to 1.287)	0.019	0.186
Creatinine increase	mmol/L	185	62	123	1.004 (0.999 to 1.009)	0.095	0.650
CRP increase	mg/L	182	61	121	1.028 (1.018 to 1.037)	0.000	0.000***
CRP:lymphocyte increase	–	166	58 (35)	108 (65)	1.015 (1.009 to 1.021)	0.000	0.000***
D-dimer increase	–	67	25	42	1.003 (1.001 to 1.004)	0.004	0.052
Eosinophils increase	10e9/L	188	65	123	0.559 (0.141 to 2.218)	0.408	1.000
Ferritin increase	ng/L	101	36	65	1.0 (1.0 to 1.0)	0.803	1.000
Glucose increase	mmol/L	122	58	64	1.176 (1.032 to 1.34)	0.015	0.166
Hb increase	g/L	189	65	124	0.976 (0.95 to 1.003)	0.084	0.650
LDH increase	U/L	92	30	62	1.006 (1.003 to 1.009)	0.000	0.005**
Lymphocytes increase	10e9/L	189	65	124	0.864 (0.51 to 1.465)	0.588	1.000
Neutrophils increase	10e9/L	189	65	124	1.353 (1.164 to 1.574)	0.000	0.002**
Neutrophil:lymphocyte ratio increase	–	166	58	108	1.343 (1.175 to 1.534)	0.000	0.000***
Platelets increase	10e9/L	189	65	124	0.989 (0.984 to 0.993)	0.000	0.000***
Potassium increase	mmol/L	178	59	119	1.193 (0.592 to 2.405)	0.621	1.000
Sodium increase	mmol/L	185	62	123	1.112 (1.026 to 1.205)	0.010	0.114
Triglycerin increase	–	51	14	37	0.091 (0.006 to 1.347)	0.081	0.650
Troponin increase	ng/L	96	29	67	1.441 (1.123 to 1.849)	0.004	0.053
Urea increase	mmol/L	185	62	123	1.165 (1.074 to 1.263)	0.000	0.003**
WCC increase	10e9/L	189	65	124	1.219 (1.093 to 1.358)	0.000	0.005**

Investigated using logistic regression adjusted for age, number of comorbidities and gender. P values were adjusted for multiple testing using Holm-Bonferroni method. ALT, alanine transaminase; AST, aspartate aminotransferase; CRP, C-reactive protein; Hb, Haemoglobin; LDH, lactate dehydrogenase; WCC, white cell count

*Significance based on adjusted p value: * when p value <0.05, **when p value <0.01, ***when p value <0.001. WCC, white cell count.

significantly associated with outcome. This adds novelty and depth to existing larger but cross-sectional work in the literature. The platelet change is likely reflective of the coagulopathy and microthrombosis that has been described in SARS-CoV-2 infection³⁷ and reflective of the literature,³⁸ although the role of anticoagulation in SARS-CoV-2 infection remains under investigation^{39, 40} and exact pathological mechanisms are unclear.⁴¹ The prognostic capability of this is of relevance to clinical practice and adds novelty to existing data, with many hospitals repeating a broad panel of ‘COVID bloods’ on a regular basis with little to guide interpretation of results. This is of relevance in the context of the current national shortage of blood tubes and also has cost per admission implications. We demonstrate that simple and inexpensive full blood count (FBC), renal profile and CRP are able to provide good prognostic information, and this is useful in the context of escalating healthcare costs of the pandemic.⁴²

Generalisability

The study period largely predates the finding that dexamethasone treatment improves survival: only 2% of the study population received dexamethasone. Furthermore, anecdotal evidence suggests that treatment pathways for COVID-19 infection have evolved and become more standardised since the start of the pandemic. It is not yet clear whether these findings are applicable in a postdexamethasone era and in relation to new variants and vaccination statuses.

Limitations and future potential

There are limitations to our study. Omission of Glasgow Coma Scale (GCS) values systematically underestimates risk compared with the ISARIC study validation data for the 4C mortality score, the maximum risk score in this analysis was 19, compared with 21 for 4C mortality score. However, we demonstrate good correlation between our adapted risk score and the 4C risk score, suggesting

21/10/2021, 14:55

Figure_4.svg

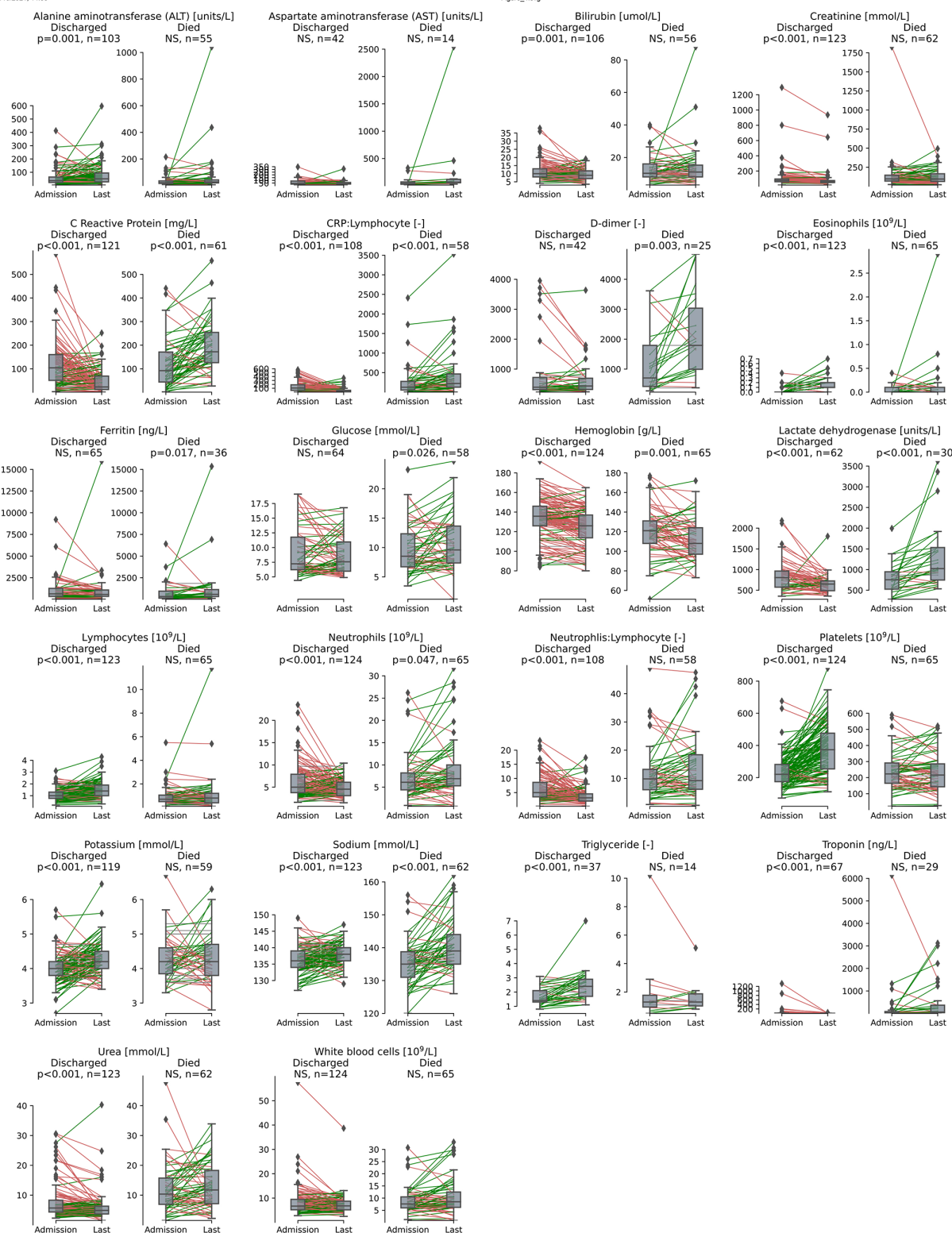


Figure 3 Changes in biochemical parameters between admission and last available specimen. Green line: increase; red line: decrease of parameter value for each individual patient. Biochemical parameters measured at admission were compared with the last available records using Wilcoxon signed-rank test within groups stratified by outcome.

reasonable validity in our results. This analysis does not adjust for differences in treatments, in particular for changes in treatment pathways as understanding of the disease and treatment effectiveness evolved over the course of the study period. With regards to symptom onset, there may be recall bias. The single-centre, relatively small

data set is a limitation to this work; however, the reflection of published findings in our data set provide reassurance that additional novel findings would be extrapolatable to a larger population. The pragmatic nature of data collection alongside clinical treatment is both a strength and a weakness; it may result in missing data points, but

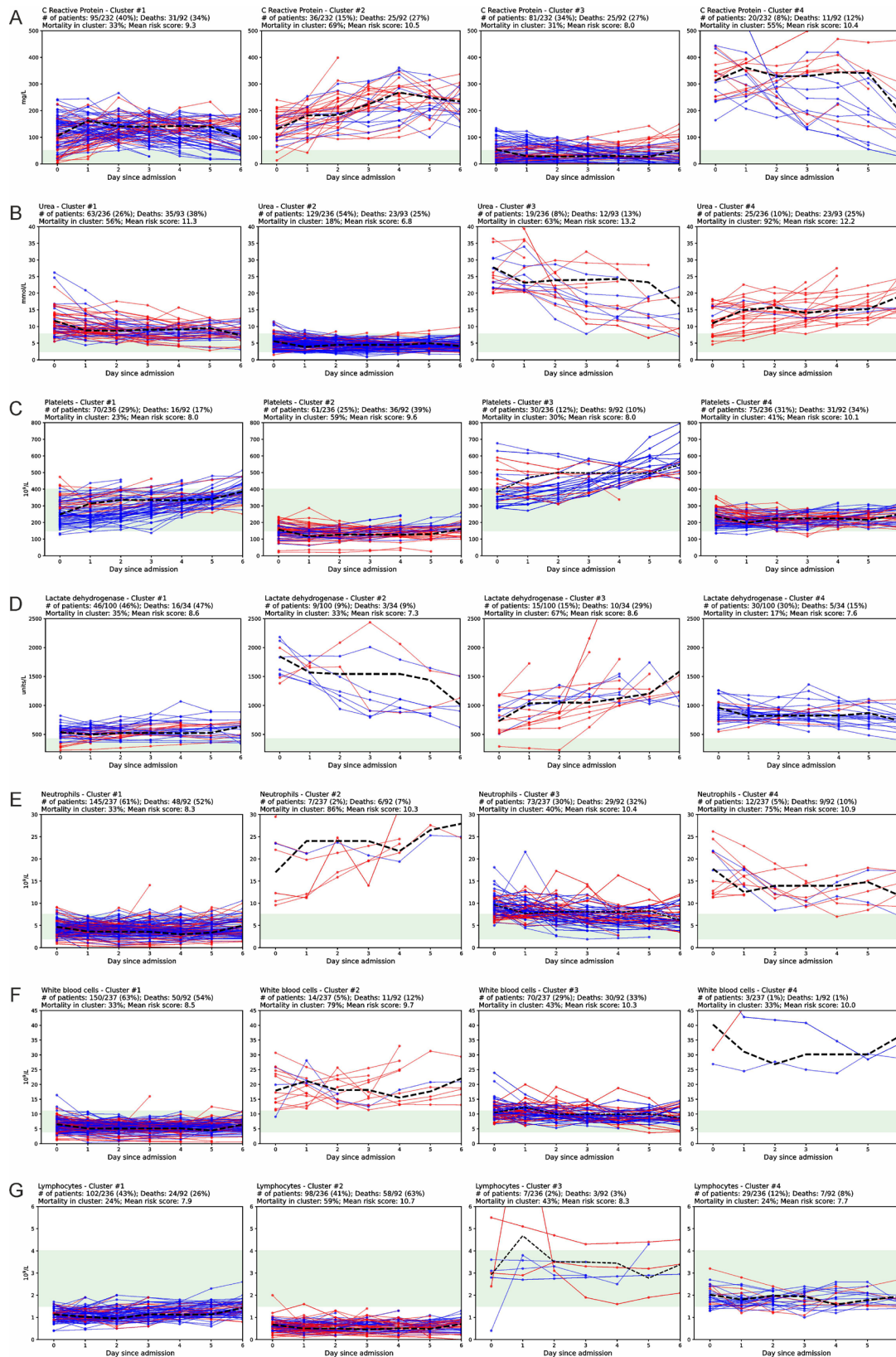


Figure 4 Clustered trajectories of parameter values over the first 7 days after admission: trajectories of biochemical parameters were clustered using the k-means clustering algorithm together with dynamic time warping^{18 19} used as a distance metric, pragmatically setting the number of cluster $k=4$: (A) CRP; (B) urea; (C) platelet counts; (D) LDH; (E) neutrophil counts; (F) white cell counts; and (G) lymphocyte counts. Each line represents an individual patient: blue: discharged; red: died; black dashed line is a centre of the cluster and green area is normal range. Note: clustering was performed for each parameter separately, that is, CRP cluster #1 does not contain the same patients as urea cluster #1. CRP, C reactive protein; LDH, lactate dehydrogenase.



adjustments have been made for these inconsistencies in the analysis. Furthermore, there may be systematic biases in the collection of data; for example, sicker patients may have more laboratory results, especially close to their date of final outcome. However, the fact that data are collected alongside routine clinical management allows for more directly translatable findings. Future work would involve investigation of the role of viral load in the context of these findings in order to better understand the pathological mechanisms that are key in driving a more severe disease phenotype and with serial cytokine sampling of airways and blood to investigate the role of hyperinflammation in this. A mechanistic modelling of the pathobiology of SARS-CoV-2 infection would allow greater understanding of the pathological processes driving these prognostic responses and may also facilitate more targeted novel treatment trials for this disease^{43, 44–49}. Importantly, we have demonstrated the utility of a new approach, based on dynamic time warping analysis, for rapidly characterising emerging COVID-19 clinical cohorts and their trajectories. Further characterisation and validation of this approach is now required to understand whether it has potential to be used to promptly delineate the impact of vaccines, treatments and emerging SARS-CoV-2 strains on clinical characteristics of COVID-19 disease and outcomes in the future.

CONCLUSIONS

We implement a new dynamic time warping approach to gain a deeper understanding of patient clinical trajectories and outcomes in a COVID-19 hospital cohort. We demonstrate that serial monitoring of specific biochemical parameters does provide additional information to single testing at admission that is immediately clinically translatable but could be limited to less expensive, readily available testing, adding novelty to existing literature. The demonstration of factors predictive for outcome raise questions as to the mechanism of severe illness in SARS-CoV-2 infection and demonstrates the need for further mechanistic modelling, which is now underway.

Author affiliations

¹Faculty of Medicine, University of Southampton, Southampton, UK

²Digital Experimental Cancer Medicine Team, Cancer Biomarker Centre, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK

³Clinical Informatics Research Unit, University of Southampton Faculty of Medicine, Southampton, UK

⁴University of Southampton, Southampton, UK

⁵Institute for Life Sciences, University of Southampton, Southampton, UK

⁶University Hospital Southampton NHS Foundation Trust, Southampton, UK

⁷University of Manchester, Cancer Biomarker Centre, Cancer Research UK Manchester Institute, Manchester, UK

⁸Digital Experimental Cancer Medicine Team, University of Manchester, Cancer Biomarker Centre, Cancer Research UK Manchester Institute, Alderley Edge, UK

Twitter Hannah Burke @AlastairWatson and Anna Freeman @DrAnnaFreeman

Contributors HB and AFree designed the protocol and, with OW and PO'R, drafted the manuscript; AD and MC were involved in protocol design; HP, FB and PF were involved in the realisation of data extraction, integration, transformation and upload processes. DL, PF, HP and JB were involved in the design and adaptation of the

Real-time Analytics for Clinical Trials platform. PO'R, OW and AFree designed and PO'R and OW performed the data analysis. AFree and DL contributed in the project management. NS and SW are involved in manual data collection processes, and AW was involved in manuscript drafting and formatting. TW was involved in study conception and protocol design. All authors contributed to editing and reviewed the final manuscript. AFree guarantees and is responsible for the overall content of the finished work and has full access to the data, and controlled the decision to publish.

Funding The REACT platform has been supported by the digital Experimental Cancer Medicine Team free of charge (grant number not applicable). The biobanking sub-cohort is supported by the NIHR Southampton CRF and NIHR Southampton Biomedical Research Centre at University Hospital Southampton NHS Foundation Trust and as part of a broader effort (Enabling New Treatment Approaches for COVID-19 Treatment) by the University of Southampton (UoS) charity (Office of Development and Alumni Relations) (grant number not applicable). In addition, the Clinical Informatics Research Unit, UoS has supported infrastructure costs (Grant number not applicable). The support described previously was not provided from a specific award or grant. The digital Experimental Cancer Medicine Team are supported by the AstraZeneca iDECIDE Programme (grant number: 119106), awarded to Manchester Cancer Research Centre and by Cancer Research UK via an Accelerator Award (award number: A29374) through the CRUK Manchester Institute (award number: C147/A25254).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Data were collected for COVID-19 positive patients admitted to University Hospital Southampton between 7 March and 4 September 2020. Ethical approval for the study was obtained from HRA specific review board (REC 20/HRA/2986). Ethical approval was obtained from the HRA Specific Review Board (REC 20/HRA/2986) for waiver of informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement A REACT COVID Data Access Management Committee has been established to prioritise and ensure appropriate governance of requests to access linked anonymised clinical data, which may be shared with other research centres. Further details of this have been published elsewhere. [please refer ref. 15]

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Andre Freitas <http://orcid.org/0000-0002-6735-2567>

Dónal Landers <http://orcid.org/0000-0001-8376-9779>

REFERENCES

- 1 Guan W-J, Liang W-H, Zhao Y, *et al*. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J* 2020;55:2000547.
- 2 W-j G, Z-y N, Hu Y. Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine* 2020;382:1708–20.
- 3 Thanh Le T, Andreadakis Z, Kumar A, *et al*. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* 2020;19:305–6.
- 4 Ramasamy MN, Minassian AM, Ewer KJ. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *Lancet* 2020.
- 5 Docherty AB, Harrison EM, Green CA, *et al*. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical

- Characterisation Protocol: prospective observational cohort study. *BMJ* 2020;369:m1985.
- 6 Watson A, Öberg L, Angermann B, *et al.* Dysregulation of COVID-19 related gene expression in the COPD lung. *Respir Res* 2021;22:1–13.
 - 7 Clift AK, Coupland CAC, Keogh RH, *et al.* Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. *BMJ* 2020;371:m3731.
 - 8 Knight SR, Ho A, Pius R, *et al.* Risk stratification of patients admitted to hospital with covid-19 using the ISARIC who clinical characterisation protocol: development and validation of the 4C mortality score. *BMJ* 2020;370:m3339.
 - 9 Moutchia J, Pokharel P, Kerri A, *et al.* Clinical laboratory parameters associated with severe or critical novel coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis. *PLoS One* 2020;15:e0239802.
 - 10 Williams S, Sheard N, Stuart B, *et al.* Comparisons of early and late presentation to hospital in COVID-19 patients. *Respirology* 2021;26:204–5.
 - 11 Wu C, Chen X, Cai Y, *et al.* Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020;180:934–43.
 - 12 Zhou F, Yu T, Du R, *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62.
 - 13 Burke H, Freeman A, Cellura DC, *et al.* Inflammatory phenotyping predicts clinical outcome in COVID-19. *Respir Res* 2020;21:245.
 - 14 Stübinger J, Schneider L. Epidemiology of coronavirus COVID-19: Forecasting the future incidence in different countries. *Healthcare* 2020;8. doi:10.3390/healthcare8020099. [Epub ahead of print: 15 04 2020].
 - 15 Jin Q. *Time Warping clustering for the forecast and analysis of COVID-19*. California Institute of Technology, 2020.
 - 16 Freeman A, Watson A, O'Regan P, *et al.* Wave comparisons of clinical characteristics and outcomes of COVID-19 admissions - Exploring the impact of treatment and strain dynamics. *J Clin Virol* 2022;146:105031.
 - 17 Burke H, Freeman A, Dushianthan A, *et al.* Research evaluation alongside clinical treatment in COVID-19 (react COVID-19): an observational and biobanking study. *BMJ Open* 2021;11:e043012.
 - 18 Cuturi M, Blondel M. Soft-DTW: a differentiable loss function for time-series. Proceedings of the 34th International Conference on machine learning, in PMLR 70:894–903, 2017. Available: <http://proceedings.mlr.press/v70/cuturi17a.html>
 - 19 Tavenard R, Faouzi J, Vandewiele G. 'Tslearn, a machine learning toolkit for time series data'. *Journal of Machine Learning Research* 2020;21:1–6.
 - 20 Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987;20:53–65.
 - 21 Satopaa V, Albrecht JR, Irwin DE. Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. 2011 31st International Conference on Distributed Computing Systems Workshops 2011:166–71.
 - 22 Azim A, Mistry H, Freeman A, *et al.* Protocol for the Wessex asthma cohort of difficult asthma (Watch): a pragmatic real-life longitudinal study of difficult asthma in the clinic. *BMC Pulm Med* 2019;19:99.
 - 23 Sperrin M, McMillan B. Prediction models for covid-19 outcomes. *BMJ* 2020;371:m3777.
 - 24 Watson A, Wilkinson TMA. Respiratory viral infections in the elderly. *Thorax* 2021;76:1000–1005.
 - 25 Rees EM, Nightingale ES, Jafari Y, *et al.* COVID-19 length of hospital stay: a systematic review and data synthesis. *BMC Med* 2020;18:270.
 - 26 Faes C, Abrams S, Van Beckhoven D, *et al.* Time between symptom onset, hospitalisation and recovery or death: statistical analysis of Belgian COVID-19 patients. *Int J Environ Res Public Health* 2020;17:7560.
 - 27 Siddiqi HK, Mehra MR. COVID-19 illness in native and immunosuppressed states: a clinical-therapeutic staging proposal. *J Heart Lung Transplant* 2020;39:405–7.
 - 28 Wang Y, Zhang L, Sang L, *et al.* Kinetics of viral load and antibody response in relation to COVID-19 severity. *J Clin Invest* 2020;130:5235–44.
 - 29 Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? *Aging* 2020;12:9959–81.
 - 30 Ye W, Chen G, Li X, *et al.* Dynamic changes of D-dimer and neutrophil-lymphocyte count ratio as prognostic biomarkers in COVID-19. *Respir Res* 2020;21:169.
 - 31 Lagadinou M, Salomou EE, Zareifopoulos N, *et al.* Prognosis of COVID-19: changes in laboratory parameters. *Infez Med* 2020;28:89–95.
 - 32 Lansbury L, Lim B, Baskaran V, *et al.* Co-Infections in people with COVID-19: a systematic review and meta-analysis. *J Infect* 2020;81:266–75.
 - 33 Rawson TM, Moore LSP, Zhu N, *et al.* Bacterial and fungal coinfection in individuals with coronavirus: a rapid review to support COVID-19 antimicrobial prescribing. *Clin Infect Dis* 2020;71:2459–68.
 - 34 Langford BJ, So M, Raybardhan S, *et al.* Bacterial co-infection and secondary infection in patients with COVID-19: a living rapid review and meta-analysis. *Clin Microbiol Infect* 2020;26:1622–9.
 - 35 Channappanavar R, Perlman S. Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. *Semin Immunopathol* 2017;39:529–39.
 - 36 Yang A-P, Liu J-P, Tao W-Q, *et al.* The diagnostic and predictive role of NLR, d-NLR and PLR in COVID-19 patients. *Int Immunopharmacol* 2020;84:106504.
 - 37 Wool GD, Miller JL. The impact of COVID-19 disease on platelets and coagulation. *Pathobiology* 2021;88:15–27.
 - 38 Liao D, Zhou F, Luo L, *et al.* Haematological characteristics and risk factors in the classification and prognosis evaluation of COVID-19: a retrospective cohort study. *Lancet Haematol* 2020;7:e671–8.
 - 39 Angus DC, Berry S, Lewis RJ, *et al.* The REMAP-CAP (randomized embedded multifactorial adaptive platform for community-acquired pneumonia) study. rationale and design. *Ann Am Thorac Soc* 2020;17:879–91.
 - 40 van Haren FMP, Richardson A, Yoon H-J, *et al.* Inhaled nebulised unfractionated heparin for the treatment of hospitalised patients with COVID-19 (INHALE-HEP): protocol and statistical analysis plan for an investigator-initiated international metatrial of randomised studies. *Br J Clin Pharmacol* 2021;87:3075–91.
 - 41 Haimee MA. Pathogenesis and treatment strategies of COVID-19-Related Hypercoagulable and thrombotic complications. *Clin Appl Thromb Hemost* 2020;26:107602962094449.
 - 42 Abbasi K. The NHS: COVID-19's next casualty. *J R Soc Med* 2020;113:291.
 - 43 Killingley B, Mann A, Kalinova M. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge. *Researchsquare* 2022.
 - 44 van Haren FMP, Page C, Laffey JG, *et al.* Nebulised heparin as a treatment for COVID-19: scientific rationale and a call for randomised evidence. *Crit Care* 2020;24:454.
 - 45 Tree JA, Turnbull JE, Buttigieg KR, *et al.* Unfractionated heparin inhibits live wild type SARS-CoV-2 cell infectivity at therapeutically relevant concentrations. *Br J Pharmacol* 2021;178:626–35.
 - 46 Monk PD, Marsden RJ, Tear VJ, *et al.* Safety and efficacy of inhaled nebulised interferon beta-1a (SNG001) for treatment of SARS-CoV-2 infection: a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Respir Med* 2021;9:196–206.
 - 47 Watson A, Spalluto CM, McCrae C, *et al.* Dynamics of IFN- β responses during respiratory viral infection. insights for therapeutic strategies. *Am J Respir Crit Care Med* 2020;201:83–94.
 - 48 Wilkinson T, Dixon R, Page C, *et al.* Accord: a multicentre, seamless, phase 2 adaptive randomisation platform study to assess the efficacy and safety of multiple candidate agents for the treatment of COVID-19 in hospitalised patients: a structured summary of a study protocol for a randomised controlled trial. *Trials* 2020;21:691.
 - 49 Watson A, Madsen J, Clark HW. SP-A and SP-D: dual functioning immune molecules with antiviral and immunomodulatory properties. *Front Immunol* 2020;11:622598.