Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study

Dan Frampton, Tommy Rampling*, Aidan Cross*, Heather Bailey*, Judith Heaney, Matthew Byott, Rebecca Scott, Rebecca Sconza, Joseph Price, Marios Margaritis, Malin Bergstrom, Moira J Spyer, Patricia B Miralhes, Paul Grant, Stuart Kirk, Chris Valerio, Zaheer Mangera, Thaventhran Prabhahar, Jeronimo Moreno-Cuesta, Nish Arulkumaran, Mervyn Singer, Gee Yen Shin, Emilie Sanchez, Stavroula M Paraskevopoulou, Deenan Pillay, Rachel A McKendry, Mariyam Mirfenderesky, Catherine F Houlihan†, Eleni Nastouli†*

Summary

Background Emergence of variants with specific mutations in key epitopes in the spike protein of SARS-CoV-2 raises concerns pertinent to mass vaccination campaigns and use of monoclonal antibodies. We aimed to describe the emergence of the B.1.1.7 variant of concern (VOC), including virological characteristics and clinical severity in contemporaneous patients with and without the variant.

Methods In this cohort study, samples positive for SARS-CoV-2 on PCR that were collected from Nov 9, 2020, for patients acutely admitted to one of two hospitals on or before Dec 20, 2020, in London, UK, were sequenced and analysed for the presence of VOC-defining mutations. We fitted Poisson regression models to investigate the association between B.1.1.7 infection and severe disease (defined as point 6 or higher on the WHO ordinal scale within 14 days of symptoms or positive test) and death within 28 days of a positive test and did supplementary genomic analyses in a cohort of chronically shedding patients and in a cohort of remdesivir-treated patients. Viral load was compared by proxy, using PCR cycle threshold values and sequencing read depths.

Findings Of 496 patients with samples positive for SARS-CoV-2 on PCR and who met inclusion criteria, 341 had samples that could be sequenced. 198 (58%) of 341 had B.1.1.7 infection and 143 (42%) had non-B.1.1.7 infection. We found no evidence of an association between severe disease and death and lineage (B.1.1.7 *vs* non-B.1.1.7) in unadjusted analyses (prevalence ratio [PR] 0·97 [95% CI 0·72–1·31]), or in analyses adjusted for hospital, sex, age, comorbidities, and ethnicity (adjusted PR 1·02 [0·76–1·38]). We detected no B.1.1.7 VOC-defining mutations in 123 chronically shedding immunocompromised patients or in 32 remdesivir-treated patients. Viral load by proxy was higher in B.1.1.7 samples than in non-B.1.1.7 samples, as measured by cycle threshold value (mean 28·8 [SD 4·7] *vs* 32·0 [4·8]; p=0·0085) and genomic read depth (1280 [1004] *vs* 831 [682]; p=0·0011).

Interpretation Emerging evidence exists of increased transmissibility of B.1.1.7, and we found increased virus load by proxy for B.1.1.7 in our data. We did not identify an association of the variant with severe disease in this hospitalised cohort.

Funding University College London Hospitals NHS Trust, University College London/University College London Hospitals NIHR Biomedical Research Centre, Engineering and Physical Sciences Research Council.

Lancet Infect Dis 2021

Published **Online** April 12, 2021 https://doi.org/10.1016/S1473-3099(21)00170-5

See **Online/Comment** https://doi.org/10.1016/S1473-3099(21)00201-2

*Joint first authors

†Authors contributed equally

Division of Infection and Immunity (D Frampton PhD, Prof D Pillay PhD, C F Houlihan PhD), **Institute for Global Health** (H Bailey PhD), **Great Ormond Street Institute of Child Health** (R Sconza MSc, M J Spyer PhD, E Nastouli FRCPath), **Bloomsbury Institute of Intensive Care Medicine** (N Arulkumaran PhD, Prof M Singer FRCP), **London Centre for Nanotechnology** (Prof R A McKendry PhD), **and Division of Medicine** (Prof R A McKendry), **University College London, London, UK; Advanced Pathogen Diagnostics Unit** (D Frampton, J Heaney PhD, M Byott MSc, M J Spyer, P Grant PhD, E Nastouli), **and Department of Clinical Virology** (T Rampling DPhil, M Margaritis DPhil, M Bergstrom BMBS, P B Miralhes MBBS, G Y Shin FRCPath, E Sanchez MD, S M Paraskevopoulou MBBS, C F Houlihan, E Nastouli), **University College London Hospitals NHS Foundation Trust, London, UK; The Francis Crick Institute, London, UK** (D Frampton, J Heaney, M Byott, M J Spyer, E Nastouli); **North Middlesex University Hospital NHS Trust, London, UK** (A Cross MBChB, R Scott MBChB, J Price MBChB, C Valerio MBChB, Z Mangera MA, T Prabhahar FFICM, J Moreno-Cuesta MD,

Introduction

On Dec 14, 2020, the UK Government was notified of the emergence of a SARS-CoV-2 variant under investigation.[1] This lineage had no obvious genetically close precursor within publicly available genomic datasets and is now defined as lineage B.1.1.7 within the COVID-19 Genomics UK Consortium (COG-UK) dataset.[2] Defining features include a deletion and several mutations within the key encoding the spike protein, notably Asn501Tyr (N501Y) in the receptor-binding domain. The effect of structural changes to the properties of the spike protein have sparked concern about transmissibility, pathogenicity, and effect of the variant on vaccine efficacy. Physical distancing restrictions were increased in London on Dec 21, 2020, in an effort to curb further spread of this variant. Further variants have since been reported with potentially similar properties to the B.1.1.7 variant,[3,4] increasing the urgency to understand the clinical relevance of the emerging variants.

We aimed to investigate the genomic characteristics and clinical outcomes associated with B.1.1.7 infection in patients admitted to our hospitals. We also assessed whether there was a difference in viral load, by proxy of PCR cycle threshold (Ct) values and whole-genome sequencing read depths, between patients infected with the B.1.1.7 variant and those infected with previously circulating lineages. Additionally, we aimed to investigate the frequency of variants of concern and frequency of escape mutations in a cohort of chronically shedding,

M Mirfenderesky FRCPath); **Health Services Laboratories, London, UK** (P Grant, S Kirk Msc); **London School of Hygiene & Tropical Medicine, London, UK** (M Margaritis)

Correspondence to: Dr Eleni Nastouli, Great Ormond Street Institute of Child Health, University College London, London, UK
e.nastouli@ucl.ac.uk

Research in context

Evidence before this study

On Jan 22, 2021, the UK Government's New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) published a document on the B.1.1.7 variant of SARS-CoV-2. An initial matched cohort study by Public Health England of patients infected with B.1.1.7 compared with other variants suggested no significant difference in risk of hospitalisation or death (risk ratio [RR] 1·00 [95% CI 0·58–1·73]). However the initial analysis had only limited time to follow up and ascertain deaths. The authors were later able to update the same matched cohort study with additional data on deaths to produce an RR of 1·65 (1·21–2·25). Additionally, two independent unpublished studies from the London School of Hygiene & Tropical Medicine and Imperial College London (both in the UK) detected a relative hazard of death of 1·35 (95% CI 1·08–1·68), and a case fatality rate of 1·35 (1·18–1·56). However, these population cohort studies, based predominately on community testing, were limited by the data available, with a low percentage of reported deaths, potential variation in case ascertainment, and transmission-setting bias. Further updates by NERVTAG include an unpublished analysis of patients admitted to hospital showing no significant increase in mortality associated with B.1.1.7. However, the authors recognise that this finding is not incompatible with an overall increase in disease severity, and they still conclude that B.1.1.7 is likely to be associated with increased hospitalisation and death. We searched PubMed for articles published between Sept 1, 2020, and Feb 1, 2021. We assessed all articles containting "B.1.1.7", "UK variant", "Kent variant", or "B117" in any language. We found 67 articles, with several focusing on genomic analyses and potential mechanisms of, and implications for, increased transmissibility. At the time of submission, we were not aware of any peer-reviewed articles combining virus genomic data and clinical outcomes of patients shown to have a B.1.1.7 infection as compared with other variants of SARS-CoV-2.

Added value of this study

Our report gives an early assessment of the B.1.1.7 variant's genomic characteristics and associated clinical outcomes, bridging the period over which B.1.1.7 became the predominant strain in two north-central London hospitals, mapping the start of the winter, 2020, surge in COVID-19 cases. By focusing on patients admitted to hospital we were able to gather comprehensive information on patient outcomes and confirm variant type by genome sequencing.

Implications of all the available evidence

The COVID-19 pandemic continues to represent a global health crisis. Analyses of the characteristics and clinical outcomes associated with this and other SARS-CoV-2 variants are likely to have important public health implications, both nationally and internationally. Our report complements previous analyses by providing further data on the outcomes of patients admitted to hospital with the B.1.1.7 variant, and it provides an exemplar of how similar studies can be rapidly conducted in the future.

mostly immunocompromised patients and remdesivir-treated patients, given reports suggesting that such populations can be the source of variants of concern.[5]

Methods

Study design and setting

Viral genomes were sequenced from combined nose and throat swab samples taken from patients with SARS-CoV-2 infection collected from Nov 9, 2020, for patients acutely admitted to a ward at either University College London Hospitals (UCLH) or North Middlesex University Hospital (NMUH) on or before Dec 20, 2020, for any clinical reason. The study dates were selected because the first hospitalised patient with the B.1.1.7 variant was admitted on Nov 9, 2020, and the B.1.1.7 variant became dominant in both hospitals by Dec 20, with this date coinciding with a surge in hospitalisations that stretched the capacity of the health services. All hospitalised patients with a positive PCR test during this time period were eligible and included in the study. .

Concerns have been raised around the emergence of variants of concern in long-shedding, immunocompromised or treated patients, especially when treatment modalities and prophylaxis target the spike protein (eg, convalescent plasma, monoclonal antibodies, and vaccination). Therefore, as part of the virological dataset, two pre-existing UCLH cohorts were analysed separately to investigate the prevalence of B.1.1.7 variant of concern (VOC)-defining mutations: 123 samples from a longitudinal study of 34 long-shedding patients, including immunocompromised patients who had remained PCR positive for more than 21 days and up to 196 days (median 33 days [IQR 26–64]), and 64 samples from a remdesivir-treated cohort of 32 patients (32 samples obtained before and 32 samples obtained after day 1 of treatment; samples were obtained a median of 5 days [IQR 3–10] before treatment and 13 days [6–19] after treatment).

To explore differences in the clinical severity associated with the B.1.1.7 and other lineages, we did a cohort study across our two centres. Inclusion criteria for this hospitalised cohort were individuals aged at least 18 years whose first PCR-positive SARS-CoV-2 result date and admission date met study criteria.

The clinical information and SARS-CoV-2 PCR samples were collected as part of routine clinical care. Data were extracted and analysed using permission granted by the National Health Service London Westminster Research Ethics Committee (IRAS 284088; REC 20/HRA/2505).

Viral detection for SARS-CoV-2

An array of SARS-CoV-2 RNA assays (Hologic Aptima TMA assay run on a Panther system [Hologic, San Diego,

CA, USA], a laboratory-developed PCR run using the open access functionality of the Panther Fusion System [Hologic], a laboratory-developed extraction-free PCR assay, and the Cepheid Xpert Xpress [Cepheid, Sunnyvale, CA, USA]) were used in the diagnostic laboratory, including non-PCR assays such as transcription-mediated amplification assay, which does not allow for Ct reporting (ie, not inferring on quantitation). Therefore, as Ct values were not always available, samples for sequencing were not pre-selected according to Ct.

Next-generation sequencing (NGS) and genomic data analysis

All SARS-CoV-2 RNA-positive samples underwent real-time whole-genome sequencing at the UCLH Advanced Pathogen Diagnostics Unit based in London, UK, and the data, used by both clinical teams, were discussed weekly at a multidisciplinary team meeting. RNA preparation and amplification were done in accordance with protocols published by the ARTIC network[6] using the V3 version of the ARTIC primer set from Integrated DNA Technologies (Coralville, IA, USA) to create tiled amplicons across the SARS-CoV-2 genome. Libraries were prepared using Nextera Flex and sequenced using Illumina MiSeq 500v2 kits (Nextera DNA Flex library preparation kit and MiSeq reagent cartridge V2 [Illumina, San Diego, CA, USA]).

Genomes were assembled using an in-house pipeline[7] and aligned to a selection of publicly available SARS-CoV-2 genomes[8] using Mafft.[9] A read depth cutoff of ten was applied after assembly; genomes with less than 75% alignment coverage were removed from subsequent analysis. Phylogenetic trees were generated from multiple sequence alignments using IQ-Tree[10] and FigTree, with lineages assigned (including VOC calls) using pangolin and confirmed by manual inspection of alignments using Aliview.[11] The COG-UK Mutation Explorer was used to identify potential mutations of concern.

Association with clinical severity

Severity was graded according to the WHO clinical progression ordinal scale (appendix p 2).[12] The scale provides a measure of illness severity from 0 (not infected) to 10 (dead). The highest value of the WHO ordinal scale that was reached by day 14 after symptom onset or after first positive SARS-CoV-2 PCR if asymptomatic was recorded. Severe disease was defined as that which requires positive pressure respiratory support, thereby reaching point 6 or higher on the WHO ordinal scale. Additionally, in-hospital mortality data by day 28 after the first positive test were collected. Clinical outcome was defined as severe if the score on the WHO scale by day 14 after symptom onset or after first positive SARS-CoV-2 PCR was at least 6 or the patient was known to have died within 28 days. Clinical outcome was defined as non-severe if the score on the WHO scale by day 14 was less than 6 and with no in-hospital death by day 28. Treatment escalation plans are a form of advanced directive used in the UK to communicate a ceiling of care around organ support treatments. Documentation of a treatment escalation plan is recommended but not mandatory for all acute hospital admissions in the UK. Because of the effect a treatment escalation plan might have on the maximum degree of organ support received and, therefore, maximum ordinal scale point reached, documentation of a valid treatment escalation plan and the relevant limitation on ordinal scale progression were recorded.

Potential confounders included age, sex, ethnicity (White *vs* any other ethnicity), body-mass index (BMI), and number of comorbidities. Comorbidities were categorised as the presence of no conditions, one condition, or at least two conditions as defined in the International Severe Acute Respiratory and Emerging Infection Consortium 4C Mortality Score, using a modified Charlson index (appendix p 1).[13] We used standard definitions[14] of community-acquired infection (symptoms or positive swab up to 2 days after admission), possible hospital-acquired infection (3–7 days after admission), probable hospital-acquired infection (8–14 days after admission) and definite hospital-acquired infection (≥15 days after admission). Admitted individuals were unlikely to have been vaccinated against COVID-19 because this study pre-dates the onset of the UK vaccination programme (appendix p 1).

Data were collected locally using a hospital laboratory information management system and electronic health record system and combined pseudo-anonymised for analysis. Logic, range, and missing data checks were done by the authors and queries verified against clinical records before analyses.

Statistical analysis

Statistical and sequence analysis was done using STATA SE (version 15), R (version 3.6.0), and in-house Perl scripts. Ct values, where available, were obtained for each sample from an in-house N-gene real time RT-PCR and compared by Welch's *t* test. Comparisons of genomic viral read depth were done using Welch's t test and two-factorial ANOVA, factorising by B.1.1.7 VOC classification and sequencing batch on: (1) the entire genomic dataset collected over the study period, (2) samples used in the cohort study, and (3) samples with whole-genome coverage greater than 50%. Analyses of correlation between median read depths and time from symptom onset to hospital admission, and between median read depths and time from symptom onset to sample collection, were done using Spearman's rank and Pearson's correlation tests. VOC and non-VOC prevalence data for London were obtained from the UK Office for National Statistics[15] (ONS), and comparisons were made with UCLH and NMUH data using standard linear regression.

Univariable comparisons of categorical variables were done using χ^2 or Fisher's exact tests, or χ^2 test for trend, and continuous variables were compared using

For **FigTree** see http://tree.bio.ed.ac.uk/software/figtree/

For **pangolin** see github.com/cov-lineages/pangolin

For the **COG-UK Mutation Explorer** see http://sars2.cvr.gla.ac.uk/cog-uk/

See **Online** for appendix

	Non-B.1.1.7	B.1.1.7	p value
Severe disease or death (n=339)	..	..	0·82
No	88/141 (62%)	126/198 (64%)	..
Yes (WHO level ≥6 or death)	53/141 (38%)	72/198 (36%)	..
Hospital (n=339)	..	..	0·043
NMUH	81/141 (57%)	135/198 (68%)	..
UCLH	60/141 (43%)	63/198 (32%)	..
Sex (n=339)	..	..	0·85
Female	74/141 (52%)	106/198 (54%)	..
Male	67/141 (48%)	92/198 (46%)	..
Age, years (n=339)	..	..	0·044
≤45	34/141 (24%)	44/198 (22%)	..
46–59	23/141 (16%)	65/198 (33%)	..
60–74	39/141 (28%)	49/198 (25%)	..
≥75	45/141 (32%)	40/198 (20%)	..
Ethnicity (n=292)	..	..	0·0004
White	85/120 (71%)	86/172 (50%)	..
Other	35/120 (29%)	86/172 (50%)	..
BMI, kg/m² (n=194)	..	..	0·088
<25	40/90 (44%)	35/104 (34%)	..
25–30	23/90 (26%)	27/104 (26%)	..
≥30	27/90 (30%)	42/104 (40%)	..
Origin of SARS-CoV-2 infection (n=339)	..	..	0·18
Community	126/141 (89%)	185/198 (93%)	..
Hospital (possible, probable, or definite)	15/141 (11%)	13/198 (7%)	..
Comorbidity score (n=336)	..	..	0·011
0	38/140 (27%)	68/196 (35%)	..
1	28/140 (20%)	57/196 (29%)	..
≥2	74/140 (53%)	71/196 (36%)	..

Data are n/N (%), unless otherwise indicated. p values were calculated using χ^2 or Fisher's exact tests, or χ^2 test for ordinal variables. BMI=body-mass index. NMUH=North Middlesex University Hospital. UCLH=University College London Hospitals.

Table 1: **Cohort characteristics by SARS-CoV-2 lineage**

the Wilcoxon-Mann-Whitney rank-sum test. Adjusted prevalence ratios (PRs) were estimated by fitting Poisson regression models with robust estimates to investigate associations between SARS-CoV-2 variant (B.1.1.7 *vs* non-B.1.1.7) and the outcome of severe disease or death, adjusting for potential confounders (hospital, age, sex, ethnicity, and comorbidity score). Wald tests were used to assess associations between the outcome and interaction terms between variant and hospital, age, and sex. Sensitivity analyses were done, first, limited to those without a treatment escalation plan documented, or whose treatment escalation plan was at WHO level 6 or higher; second, among those with symptoms or a positive test pre-dating hospital admission; and third, with inclusion of WHO level 5 (oxygen without positive pressure) in the outcome group.

Role of the funding source

The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Of 496 patients who were PCR positive for SARS-CoV-2 on a sample taken on or after Nov 9, 2020, and admitted up to Dec 20, 2020, 341 (69%) had samples that could be sequenced (appendix p 9). The intermittent use of 1 mL swabs hampered sequencing success because of inadequate sample. Median age of patients was 60 years (IQR 47–75), 242 (49%) of 496 were male, and 157 (42%) of 374 with ethnicity data available were of an ethnic minority group. The 155 (31%) patients without sequence data (75 [48%] with insufficient sample material for sequencing and 80 [52%] in whom sequencing failed to generate an adequate genome for analysis) were similar to those with sequence data with respect to hospital of admission (p=0·99), age (p=0·93), sex (p=0·15), and ethnicity (p=0·54), but patients with an available viral genetic sequence were more likely than those without to have severe disease or die (125 [37%] of 339 *vs* 37 [24%] of 154; p=0·0049). Looking only at deaths, this difference was also apparent (55 [16%] of 341 *vs* 14 [9%] of 153; p=0·039). We found no evidence of a trend in proportion of patients in whom a sequence could not be obtained over time (p=0·75; appendix p 12) or of a correlation between proportion of unsequenced samples and, among those sequenced, proportion with B.1.1.7 lineage in each week (p=0·486).

Of the 341 patients with sequence data available, 198 (58%) had B.1.1.7 VOC and 143 (42%) had other lineages. 339 patients (99%) had data available on WHO ordinal scale level of care or were reported to have died, of whom 72 (36%) of 198 patients with B.1.1.7 and 53 (38%) of 141 with non-B.1.1.7 met the outcome of severe disease (ie, WHO level 6 and above or death; p=0·82). By variant, the proportion of patients at level 6 or levels 7–9 on the WHO ordinal scale or who died were similar: in the non-B.1.1.7 group, 18% (26 of 141) were at level 6, 2% (three of 141) were at levels 7–9, and 17% (24 of 141) died; in the B.1.1.7 group, 15% (29 of 198) were at level 6, 6% (12 of 198) were at levels 7–9, and 16% (31 of 198) died (p=0·172 for a comparison of the distribution of these three categories within patients with clinically severe outcomes by variant). Although the proportion with non-severe disease was similar by variant overall (table 1), 88 (44%) of 198 patients with B.1.1.7 received oxygen by mask or nasal prongs (a category within the non-severe group) compared with 42 (30%) of 141 patients with non-B.1.1.7 lineage (p=0·0063 when compared as proportion of all patients with sequences; appendix p 2). The group with B.1.1.7 were younger overall, with fewer comorbidities, and were more likely to be of an ethnic minority group than those with non-B.1.1.7 (table 1). Figure 1 shows that the proportions with clinically severe outcomes were similar by variant, across age groups.

Characteristics of patients according to whether or not they had severe disease or died are shown in the appendix (p 7). 91 (73%) of 125 patients with severe disease or who died were at NMUH versus 125 (58%) of 214 patients

with a WHO ordinal scale level of less than 6 (p=0·079). Those with severe disease or who died were older (p<0·0001; figure 1) and more likely to have comorbidities than those without severe disease (p=0·0005). 67 (54%) of 125 patients with severe disease or who died were male versus 92 (43%) of 214 with less severe disease (p=0·059), but we found no difference by ethnicity (p=1·0) or BMI (p=0·92). BMI was not considered further in adjusted analyses due to substantial missing data (available for 194 patients only).

Overall, 92 (27%) of 339 patients had no identified treatment escalation plan in place, 221 (65%) had a treatment escalation plan with a specific maximum level, and 26 (8%) were missing information on the presence of a treatment escalation plan. 38 (17%) of 221 patients with a treatment escalation plan had restrictions limiting progression beyond ordinal scale level 5, of whom 24 (63%) of 38 had died.

We found no evidence of a difference in our main outcome of severe disease or death by SARS-CoV-2 lineage (B.1.1.7 *vs* non-B.1.1.7) in either unadjusted analyses (PR 0·97 [95% CI 0·72–1·31]) or analyses adjusted for hospital, sex, age, comorbidities, and ethnicity (adjusted PR 1·02 [0·76–1·38; table 2). We found no evidence of effect modification by hospital (p=0·81), sex (p=0·68), or age (p=0·47).

To explore the potential effect on outcome misclassification among the group with a treatment escalation plan in place, sensitivity analyses were done among 231 people without a treatment escalation plan or with a treatment escalation plan up to a maximum WHO level of 6, and for whom data were available on all variables included in the multivariable model in table 2. In this analysis, we found no association between SARS-CoV-2 lineage (B.1.1.7 *vs* non-B.1.1.7) and outcome of severe disease or death (unadjusted PR 0·99 [95% CI 0·68–1·43]; adjusted PR 0·99 [0·67–1·47]). In a sensitivity analysis among 209 patients with symptoms or a positive SARS-CoV-2 PCR test that pre-dated hospital admission, we found no association between variant and severe disease or death (unadjusted PR 0·89 [0·64–1·25]; adjusted PR 0·94 [0·67–1·33]). In a final sensitivity analysis, we found an association between infection with the B.1.1.7 variant and receipt of non-pressurised oxygen (WHO level 5) after adding this group to the outcome of severe disease or death and adjusting for other variables in table 2 (adjusted PR 1·19 [1·04–1·36]); however, as an isolated finding it was difficult to interpret.

31 (16%) of 198 patients with B.1.1.7 died within 28 days versus 24 (17%) of 141 with non-B.1.1.7 lineages (p=0·74). We found no excess mortality risk associated with B.1.1.7 compared with non-B.1.1.7 in unadjusted analyses (PR 0·85 [95% CI 0·52–1.41] for B.1.1.7 *vs* non-B.1.1.7), nor after adjusting for potential confounders listed in table 2 (adjusted PR 1·12 [95% CI 0·71–1·78]; n=289 for both).

From available sample Ct value data for 27 B.1.1.7 samples and 38 non-B.1.1.7 samples, we found significantly lower Ct values associated with B.1.1.7 compared with non-B.1.1.7 (mean Ct 28·8 [SD 4·7] *vs* 32·0 [4·8]; p=0·0085). Correspondingly, we found significantly higher median genomic read depths in B.1.1.7 samples than in non-B.1.1.7 samples (mean median depths 1445 [952] *vs* 782 [728]; p=0·0030).

In a larger analysis on the full genomic dataset (224 B.1.1.7 samples and 291 non-B.1.1.7 samples), we found significantly greater read depth for B.1.1.7 samples than for non-B.1.1.7 samples (mean median depth 1279 [SD 1004] *vs* 665 [693]; p=0·0002; appendix p 10). We obtained the same result when factoring both variant and sequencing batch (p<0·0001; appendix p 1) and when restricting the analysis to samples used in the cohort study (p=0·0030). Considering the selection bias inherent to classifying genomes as B.1.1.7 VOC, we removed all non-variant genomes beneath the 50% genome coverage cutoff proscribed by pangolin

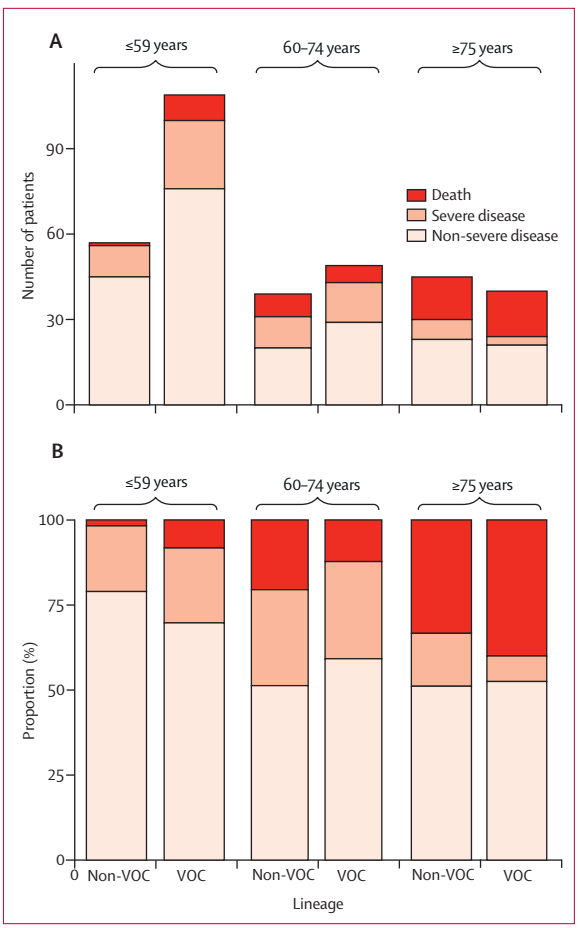

Figure 1: **Severity of illness across patient age groups and by presence of VOC or non-VOC SARS-CoV-2 infection**
Figure shows absolute counts (A) and proportion of patients (B). Non-severe disease was defined as reaching a WHO ordinal scale of less than 6 by day 14 after symptom onset. Severe disease was defined as reaching an ordinal scale point of 6 or higher. Death was defined as those who had died by day 28 after the first positive swab. VOC=variant of concern.

	Proportion with WHO level ≥6 or death	Crude prevalence ratio (95% CI; n=289)	p value	Adjusted prevalence ratio (95% CI; n=289)	p value
Lineage					
Non-B.1.1.7	46/119 (38·7%)	1 (ref)	..	1 (ref)	..
B.1.1.7	64/170 (37·7%)	0·97 (0·72–1·31)	0·86	1·02 (0·76–1·38)	0·88
Hospital					
NMUH	84/196 (42·9%)	1 (ref)	..	1 (ref)	..
UCLH	26/93 (28·0%)	0·65 (0·45–0·94)	0·022	0·72 (0·50–1·03)	0·075
Sex					
Female	51/154 (33·1%)	1 (ref)	..	1 (ref)	..
Male	59/135 (43·7%)	1·32 (0·98–1·77)	0·066	1·30 (0·97–1·74)	0·075
Age, years					
≤45	13/65 (20·0%)	1 (ref)	..	1 (ref)	..
46–59	24/73 (32·9%)	1·64 (0·91–2·96)	0·097	1·59 (0·88–2·88)	0·13
60–74	34/77 (44·2%)	2·21 (1·28–3·82)	0·0046	2·02 (1·12–3·66)	0·020
≥75	39/74 (52·7%)	2·64 (1·55–4·49)	0·0004	2·33 (1·28–4·26)	0·0059
Ethnicity					
White	65/170 (38·2%)	1 (ref)	..	1 (ref)	..
Other	45/119 (37·8%)	0·99 (0·73–1·33)	0·94	1·18 (0·87–1·60)	0·278
Comorbidity score					
0	23/86 (26·7%)	1 (ref)	..	1 (ref)	..
1	24/72 (33·3%)	1·25 (0·77–2·01)	0·37	1·02 (0·63–1·66)	0·939
≥2	63/131 (48·1%)	1·80 (1·21–2·66)	0·0034	1·22 (0·78–1·90)	0·387

Table 2: **Association of SARS-CoV-2 B.1.1.7 variant with disease severity**

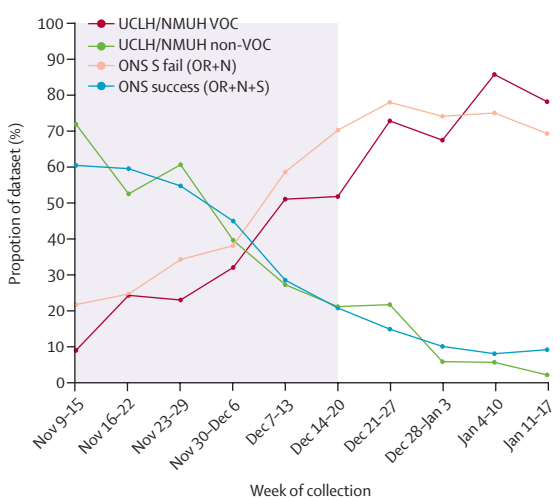

Figure 2: **Proportion of B.1.1.7 VOC and other lineages observed at UCLH and NMUH, with ONS S gene failure data for London, 2020–21**
The period covered by this study is shaded. Genomes were classified as either VOC (B.1.1.7 or VOC-202012/01), non-VOC (all other lineages), or were unclassifiable because of poor sequencing. Data for unclassifiable samples are not shown. All classifications were made using pangolin[12] followed by manual inspection of alignments. ONS=UK Office for National Statistics. NMUH=North Middlesex University Hospital. UCLH=University College London Hospitals. VOC=variant of concern.

and observed the same significant difference in read depth between B.1.1.7 and non-B.1.1.7 samples (mean median depth 1279 [1004] *vs* 831 [682]; p=0·0011).

We did not see a significant effect of time from symptom onset to hospital admission on variation in median read depth for the cohort study samples (p=0·71). We found no correlation between median read depth and time to hospital admission from symptom onset for these samples (r_s=–0·08; p=0·32; appendix p 13). However, although we found no significant effect of time from symptom onset to date of sample collection on variation in median read depth for the cohort study samples (p=0·77), we found a weak correlation between median read depth and time from symptom onset to date of sample collection as expected in the course of infection (r_s=–0·17; p=0·032; appendix p 14). Where data were available, we also found no correlation between Ct values and time from symptom onset to hospital admission (r_s=0·25; p=0·34).

Overall, time to hospital admission from symptom onset was longer for patients with B.1.1.7 than for patients without B.1.1.7 (median 6·0 days [IQR 4·0–8·0] or mean 5·1 days [SD 6·6] *vs* median 4·0 days [IQR 1·0–8·0] or mean 3·9 days [SD 6·8]; appendix p 15), but this difference was not significant (p=0·15).

We identified 224 (43%) B.1.1.7 VOC sequences in our dataset from 515 samples taken from UCLH, NMUH, and associated outpatient clinics (patients from associated outpatient clinics were not suitable for inclusion in the cohort study). The proportion of samples with B.1.1.7 sequences at UCLH and NMUH increased in line with data on S gene target failure from the ONS for the London region,[15] with corresponding reductions in non-B.1.1.7 sequences (linear regression, r^2=0·90;

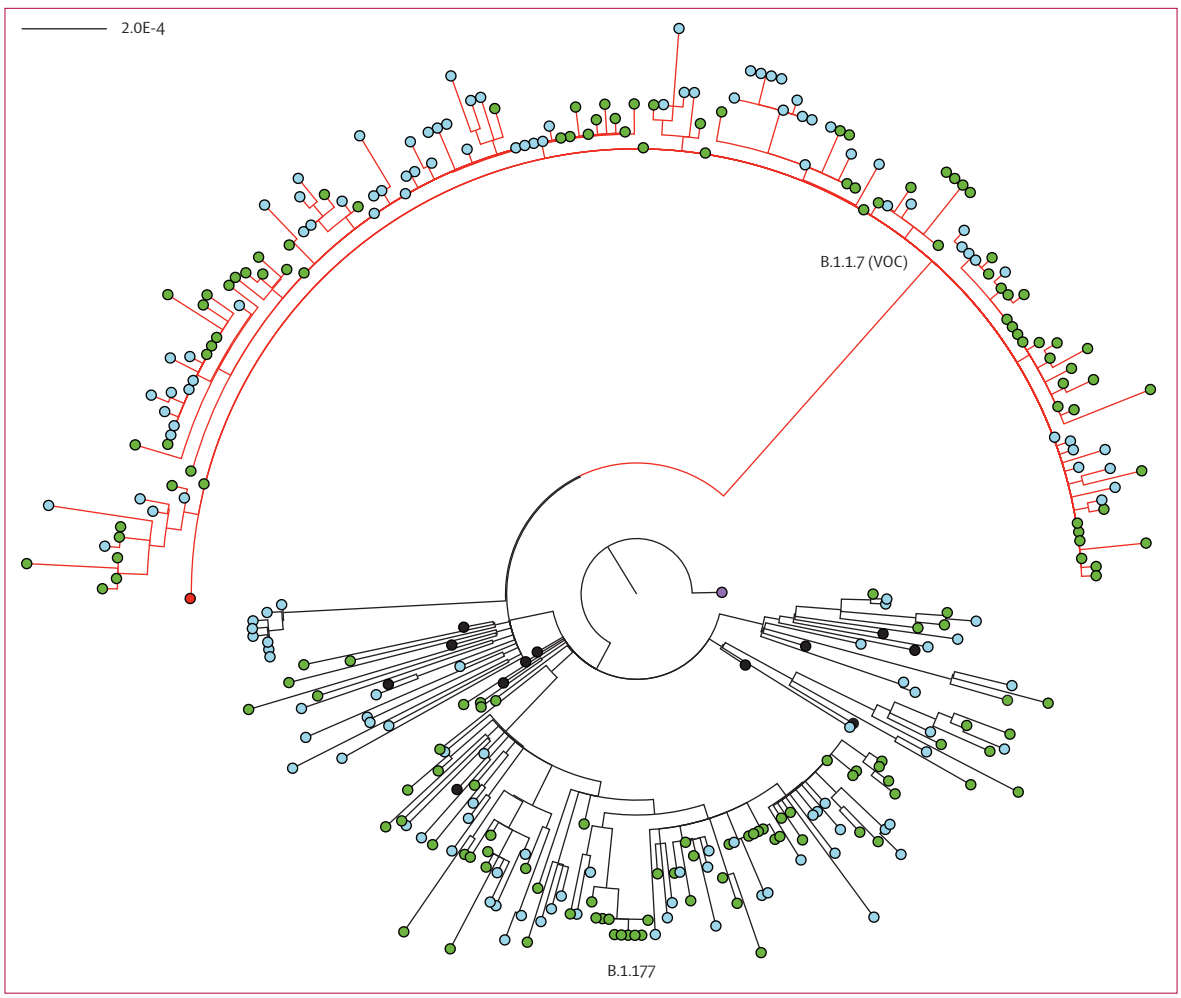

Figure 3: **A phylogenetic tree of UCLH and NMUH sequenced genomes**
UCLH samples are coloured blue (123 sequences) and NMUH samples (216 sequences) coloured green. The canonical B.1.1.7 VOC 2012012/01 sequence (GISAID accession EPI_ISL_601443) is highlighted in red. The tree is rooted on a historic SARS-CoV-2 sequence (Wuhan-Hu-1, NC_045512.2) shown in purple, and other representative lineages are shown in black (appendix pp 3–6). The B.1.1.7 VOC lineage is characterised by low within-clade sequence diversity relative to non-VOC strains, displaying a broad expansion of relatively shallow branches. The most frequently observed non-B.1.1.7 lineage in this study, B.1.177, is highlighted for comparison. NMUH=North Middlesex University Hospital. UCLH=University College London Hospitals. VOC=variant of concern.

p=0·0038) and PCR tests positive for ORF1ab, N, and S genes over time (r^2=0·88; p=0·0054; figure 2).

Sequence alignment of the 224 VOC sequences confirmed maximal concordance to the canonical B.1.1.7 VOC genome[16] (first isolated Sept 20, 2020, in Kent, UK; GISAID ID EPI_ISL_601443) with regards to B.1.1.7 VOC-defining single nucleotide polymorphisms (SNPs) and deletions, with the exception of samples where sequencing had failed for the region of interest. Of 13 non-synonymous SNPs, key mutations included Asn501Tyr, A23063T) and P681H (C23604A, Pro681His) in the spike protein, a co-occurrence not previously observed. Asn501Tyr is a key contact residue in the receptor-binding domain, which has been shown to enhance angiotensin-converting enzyme 2 receptor affinity.[17] P681H forms part of a quartet of residues involved in creating a furin cleavage site between S1 and S2, promoting entry into lung cells and primary human airway epithelial cultures.[18] All B.1.1.7 genomes contained a deletion at S 69–70, which causes reproducible S-gene target failure in the TaqPath assay and in conjunction with N501Y might account for increased transmissibility of the variant.[16]

Non-spike B.1.1.7 VOC-defining mutations included SNPs in N, ORF1ab, and ORF8 (including a premature stop codon at position 27) and six synonymous mutations observed across all samples.

Phylogenetic analysis revealed UCLH and NMUH B.1.1.7 genomes form a distinct cluster with the B.1.1.7 VOC canonical sequence at the root of the sub-clade and a long branch length relative to other clades (figure 3). Diversity within the cluster was low. Samples were at most nine SNPs different from the B.1.1.7 VOC reference. Nine samples were identical to the reference, despite it being isolated over 2 months earlier.

Further sequence analysis confirmed this observation. Nucleotide diversity within all UCLH and NMUH VOC samples was approximately 4·5 nt, with 5·2 nt for UCLH and 3·9 nt for NMUH variants as individual sets, most having no clear epidemiological linkage. Similar analysis of B.1.177 lineage samples at UCLH during the same time period estimated their nucleotide diversity to be approximately two to three times higher (11·8 nt; 64 sequences). 156 (85%) of 182 B.1.1.7 samples linked by pairwise comparison (distance ≤2 nt) to another sample at the same hospital could also be linked to another sample at the other hospital.

Genome analysis of 123 longitudinal samples from a cohort of 34 patients with protracted viral shedding at UCLH who each had remained SARS-CoV-2-positive on PCR for more than 21 days revealed no B.1.1.7 VOC-defining SNPs or deletions, nor evidence of additional mutations at the respective genomic positions. 28 (82%) of these patients had underlying immunocompromise (appendix p 16).

We obtained similar results from a genome analysis of 32 patients treated with a 5-day course of remdesivir (appendix p 16). We found no VOC-defining mutations or deletions in this cohort, nor evidence of additional mutations at the same sites, either before treatment (32 samples, median time before treatment was 5 days [IQR 3–10]) or after treatment (32 samples, median 12 days after day 1 of remdesivir treatment [6–19]). For both cohorts, we also found no evidence of B.1.1.7 VOC-defining mutations or deletions present at frequencies greater than 5% of the total viral population in these samples—ie, we found no minority variant mutations containing these mutations in these cohorts.

More generally, we found no lineages of concern in either cohort, although two potentially notable mutations in the spike protein (L18F [C21614T, Leu18Phe] and H146Y [C21998T, His146Tyr]) were identified by COG-UK Mutation Explorer, both posited to potentially affect antibody binding on the basis of changes to protein secondary structure[19] but classified by COG-UK Mutation Explorer as there being low confidence in their effect. However, UK B.1.1.7 lineages have shown a rapid rise in acquisition of Leu18Phe,[20] and it is a conserved VOC-defining mutation for the P.1 VOC lineage first isolated in Brazil.[21] In this study, Leu18Phe was found in seven samples from three treated patients and one protracted viral shedder. In each case, the Leu18Phe mutations were present in the first successfully sequenced sample for each patient and were observed in all subsequent samples. We found no evidence that the mutations arose over the course of infection or in response to treatment.

Discussion

The emergence of novel VOCs in the ongoing SARS-CoV-2 pandemic requires rapid genomic, virological, epidemiological, and clinical characterisation to inform public health, clinical, and research responses. This study was done contemporaneously with the emergence and spread of the B.1.1.7 variant throughout the south of England and offers a unique and well characterised cohort of hospitalised patients. Within this cohort, which represents a substantial proportion of the hospitalised patients with COVID-19 in north-central London during this period, we found no evidence that the B.1.1.7 variant was associated with severe disease or death. One of the strengths of this study lies in its timing, which was several weeks before the peak of hospital admissions in London, and before any substantial resource limitation or strain on clinical care.

Investigating the emergence of strains and variants of RNA viruses of pandemic potential, including influenza and coronaviruses, is a key component of pandemic preparedness.[22] In the COVID-19 pandemic, two advances have facilitated this surveillance: the wider use of deep sequencing techniques and the availability of advanced bioinformatic tools and digital platforms giving immediate access for near real-time analysis.[23,24] Of particular concern are mutations relating to cross-species transmission, in the case of SARS-CoV-2 allowing for potential establishment of new animal reservoirs.[25] International travel adds further complexity because population movement offers opportunities for variants to transmit worldwide. Variants must be rapidly assessed for their potential to increase transmission, to result in resistance to antiviral treatments and vaccines, and to alter the clinical phenotype, disease severity, and mortality.

Following concerns that the B.1.1.7 VOC has enhanced transmissibility,[1,16] we investigated whether this characteristic is reflected by an increase in viral load, using Ct values from an in-house N-gene real-time RT-PCR assay and genomic read depths as surrogates. Although our Ct value analysis was limited by data availability, other studies have shown that NGS read counts can be used as a reliable predictor of viral load.[26] Given that we found significant differences for Ct values and genomic read depths between B.1.1.7 and non-B.1.1.7 samples, we believe that B.1.1.7 infections were associated with higher viral loads than were non-B.1.1.7 infections in this study. This finding is in keeping with results from similar independent analyses, including that of approximately 1400 genomes assembled as part of the UK test and trace programme, which reported a 0·5 increase in median $\log_{10}$-inferred viral load in B.1.1.7 relative to non-B.1.1.7 samples.[27] Our observed higher read depths are equivalent to a 0·2–0·3 increase in $\log_{10}$ read depth in B.1.1.7 relative to non-B.1.1.7, a smaller increase than observed in the previous study, which might be a consequence of sampling patients at later stages of infection than was done for test and trace swabs, which are typically derived from recently symptomatic individuals when viral loads are likely to be high.[28]

Although our data show that B.1.1.7 was associated with an increased viral load by proxy of PCR Ct values and NGS read depth in the nasopharynx, we saw no association between B.1.1.7 and severity. Previous studies have

suggested an association between viral load and mortality.[29] In our study, a greater proportion of the sequenced group had severe disease than of the unsequenced group (who were unsequenced as a result of having insufficient sample collected or failed sequencing). An underlying association between B.1.1.7 and disease severity in the hospitalised population overall might have been unobserved in our analyses, if those with B.1.1.7 were more likely to have a successful sequence because of a higher viral load or a greater number of samples available. However, we did not detect a trend in the proportion of sequences failing over time, as B.1.1.7 began to predominate in the population, or any correlation between proportion of B.1.1.7 samples among those sequenced and proportion of unsequenced samples in each week of the study.

The currently identified B.1.1.7 VOC possesses the del21765–21770 (del69–70 HV) deletion and mutations across the spike protein, importantly including the receptor-binding domain. Lineage B.1.1.7 has largely replaced the previously circulating variant in our centres and across the UK. The emergence of the lineage in the presence of a circulating variant is suggestive of natural selection of a more transmissible virus,[30] consistent with other UK data and ours. The accumulation of 17 mutations suggests possible emergence of the lineage in an immunocompromised host, and although our data from immunocompromised and remdesivir-treated patients do not confirm this hypothesis, these populations will need to be monitored intensely as they receive vaccines, monoclonal antibodies, and other preventive and treatment modalities in the near future. However, our findings suggest that B.1.1.7 VOC-defining mutations do not arise solely in response to remdesivir treatment and are not more likely in immunocompromised patients in the absence of additional treatment. The L18F mutations we observed in several patients most likely reflect a higher prevalence of the mutation at the time of sampling before B.1.1.7 became the dominant lineage in the UK. The canonical B.1.1.7 VOC does not contain this mutation. Nevertheless, variants identified in South Africa and Brazil pose further concern, especially because they carry mutations with the potential to escape antibodies or vaccines and have been emerging in populations with presumed high seroprevalence. Indeed, an analysis of our data identified two B.1.1.7 isolates with the Glu484Lys (E484K) substitution (both otherwise identical to the canonical B.1.1.7 VOC reference genome), causing concern that the VOC is acquiring this mutation while circulating in the UK and might further spread.

The lower observed diversity within the SARS-CoV-2 sequenced genomes included in this study is consistent with B.1.1.7 transmission occurring more readily during early infection. A broader phylodynamic analysis over a longer timeframe accounting for sampling bias (eg, local outbreaks) would confirm whether the underlying rate of nucleotide substitution is genuinely lower for B.1.1.7 or, more likely, simply a reflection of a more recent most common ancestor.

Most B.1.1.7 samples were linked by pairwise comparison (distance ≤2 nucleotide difference) to another sample at the same hospital or the other hospital. Although these linkages are associative, the finding highlights the necessity for timely and reliable epidemiological data, in addition to genomic data, to rule out potential nosocomial transmissions given the prevalence of B.1.1.7 in the UK. Reassuringly, the proportion of B.1.1.7 infections arising from hospital acquisition was similar to that of non-B.1.1.7, confirming the variant had not yet become established in the hospital setting and removing any bias of a difference in mortality associated with nosocomial infection.[31] B.1.1.7 was seen more frequently in one of the two hospitals and significantly more frequently in ethnic minority groups than in White people. This finding could be explained by a difference in demographics and socioeconomic factors between patients in the two hospitals, suggesting a founder effect in this population at the time of B.1.1.7 VOC emergence.

Patients with B.1.1.7 were younger and had fewer comorbidities than those with non-B.1.1.7 infection, possibly representing the widespread and potential increased transmission of this variant in the community or differences in probability of hospital admission, which we were not able to explore in this hospital-based cohort. An unpublished ecological study cited in a report by the UK Government's New and Emerging Respiratory Virus Threats Advisory Group found an increased risk of hospitalisation per case associated with B.1.1.7 on a population level.[32] In our study, older age remained associated with severe outcome or death in adjusted analyses, although no difference between lineages was reported. Further community-based studies should be done to allow a larger denominator unselected by disease severity, to investigate any association between B.1.1.7 and the probability of hospitalisation or small differences in virulence that might occur in individuals with pauci-symptomatic or asymptomatic infection. This association might be of particular relevance when investigating effects potentially confounded by age because minimally symptomatic infections occur more frequently in younger individuals than in older individuals. This study was able to rule out a difference of 1·85 or greater increased odds of severe disease. More subtle associations with severity have been reported[32] but in different types of community cohorts that do not allow for direct comparison.

Level of severity on a validated scale for COVID-19[12] was captured within 14 days after a positive test or onset of symptoms in this study, allowing sufficient time for deterioration, given the median time to clinical deterioration in a large observational study was 4 days (IQR 1–9) after admission.[13] Some patients might have deteriorated after day 14 and the outcome missed, but this possibility was mitigated by capturing death at

day 28 for hospitalised patients. Some individuals might have been discharged and died either at home or another site and their outcome not captured. Finding B.1.1.7 more commonly in younger versus older individuals gives a subtle hint of more severe disease if patients with B.1.1.7 are hospitalised more often compared with patients with other lineages, although difference in disease severity by B.1.1.7 was not found in this hospitalised cohort in the main analysis. In sensitivity analyses further exploring the non-severe group, compared with those with non-B.1.1.7, those with B.1.1.7 were more likely to receive oxygen without positive pressure, and this difference persisted in adjusted analyses. However, we are cautious in the interpretation of this finding because of the limitations of oxygen without positive pressure as a measure of disease severity (with its use possibly being selected by reasons for hospitalisation unrelated to COVID-19 or residual confounding by other patient characteristics). Further, we found no clear pattern towards more severe disease in the other ordinal scale levels in the B.1.1.7 group. We acknowledge comparison of outcomes between groups have not corrected for treatments including use of steroids, antiviral medications, tocilizumab, and convalescent plasma. Also, some patients possibly met our outcome definition by receiving oxygen with positive pressure or ventilation for reasons other than COVID-19.

Rapid collection of good quality clinical data with the appropriate granularity, in combination with whole-genome sequencing of SARS-CoV-2, is imperative in deciding whether variants are associated with altered clinical outcomes. These data, in conjunction with in-vitro investigation of neutralisation capacity of sera from individuals following vaccination and natural infection, are essential in the public health response and clinical management of COVID-19. Large readily available datasets will be key in enabling rapid clinical assessment of variants. Our data, within the context and limitations of a real-world study, provide initial reassurance that severity in hospitalised patients with B.1.1.7 is not markedly different from severity in those without, and this study provides a model to answer the same question again as we move into an era of emerging variants.

Contributors

EN, DF, CFH, TR, AC, and HB drafted the first version of the manuscript. JH did all sample extractions and sequencing. TR, AC, RScot, JP, and CFH did the clinical data extraction, verification, and curation. DF and MBy designed and did the bioinformatic analyses. CFH, HB, RScon, and EN designed and did the cohort study analysis. All authors provided data or contributed to the writing of the manuscript and approved the final version. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

MS reports funding support for at-cost development and manufacture of UCL-Ventura continuous positive airway pressure device for patients with COVID-19 from the UK Department of Health and Social Care, during the conduct of the study; grants and advisory board fees paid to institution research fund from NewB; grants from DSTL; advisory board and speaking fees paid into institutional research fund from Amormed, Biotest, General ElectricBaxter, Baxter, Roche, Bayer, and Shionogi; and grants from Critical Pressure and Apollo Therapeutics, outside the submitted work. All other authors declare no competing interests.

Data sharing

Individual-level data of patients included in this manuscript after de-identification are considered sensitive and will not be shared. The study method and statistical and bioinformatic analyses are all described in detail in the Methods and throughout the manuscript. Virus genomic data are uploaded on the COG-UK servers and subsequently shared on GISAID.

Acknowledgments

UCLH Advanced Pathogens Diagnostic Unit receives funding from the University College London Hospitals NHS Trust. The APDU Illumina sequencing platform was funded by a grant from the UCLH/UCL UK National Institute for Health Research Biomedical Research Centre in 2017. APDU participates in COG-UK as a sequencing site but receives no funding from COG-UK. APDU whole-genome sequencing data are uploaded in real time on the COG-UK servers. DF was funded by the i-sense Engineering and Physical Sciences Research Council Interdisciplinary Research Collaborations in Agile Early Warning Sensing Systems for Infectious Diseases and Antimicrobial Resistance (EP/R00529X/1). EN, MJS, and JH receive funding from the EU Horizon 2020 programme (734857 and 862840). We thank Nathan Lea, information governance officer at UCLH, and Maurice Griffin at the UCLH/UCL Joint Research Office for their help with ethics applications; Lisa Levett, at Health Services Laboratories, and Jerome Nicod, at Francis Crick Institute, for their support for APDU; and Mike Brown, Kirit Ardeshna, and Alun Marc Henry, all UCLH and NMUH clinicians, including Sarjana Jain and Teh Li Chin, for their critical reading of the manuscript.

References

1 Wise J. Covid-19: new coronavirus variant is identified in UK. *BMJ* 2020; **371:** m4857.

2 Rambaut A, Loman N, Pybus O, et al. (CoG-UK). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Dec 18, 2020. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed Jan 31, 2021).

3 Tegally H, Wilkinson E, Giovanetti M, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike proteins in South Africa. *medRxiv* 2020; published online Dec 22. https://doi.org/10.1101/2020.12.21.20248640 (preprint).

4 Faria N, Claro IM, Candido D, et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Jan 12, 2021. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586 (accessed Jan 31, 2021).

5 Kemp SA, Collier DA, Gupta RK. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 2021; published online Feb 5. https://doi.org/10.1038/s41586-021-03291-y.

6 Quick J, Loman N. hCoV-2019/nCoV-2019 version 3 AmpliconSet. ARTIC Consortium; 2020. https://artic.network/resources/ncov/ncov-amplicon-v3.pdf (accessed Jan 31, 2021).

7 Harvala H, Frampton D, Grant P, et al. Emergence of a novel subclade of influenza A(H3N2) virus in London, December 2016 to January 2017. *Euro Surveill* 2017; **22:** 30466.

8 Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 2017; **1:** 33–46.

9 Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; **30:** 772–80.

10 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; **32:** 268–74.

11 Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014; **30:** 3276–78.

12 WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection. A minimal common outcome measure set for COVID-19 clinical research. *Lancet Infect Dis* 2020; **20:** e192–97.

13 Gupta RK, Harrison EM, Ho A, et al. Development and validation of the ISARIC 4C deterioration model for adults hospitalised with COVID-19: a prospective cohort study. *Lancet Respir Med* 2021; published online Jan 11. https://doi.org/10.1016/S2213-2600(20)30559-2.

14 European Centre for Disease Prevention and Control. Surveillance definitions for COVID-19. July 20, 2020. https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions (accessed Feb 4, 2021).

15 UK Office for National Statistics. Coronavirus (COVID-19) infection survey, UK: 22 January 2021. Office for National Statistics. Jan 22, 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/22january2021 (accessed Jan 31, 2021).

16 Public Health England. Investigation of novel SARS-CoV-2 variant of concern 202012/01. Dec 21, 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/947048/Technical_Briefing_VOC_SH_NJL2_SH2.pdf (accessed Jan 31, 2021).

17 Starr T, Greaney AJ, Hilton SK, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 2020; **182:** 1295–310.e20.

18 Peacock TP, Goldhill DH, Zhou J, et al. The furin cleavage site of SARS-CoV-2 spike protein is a key determinant for transmission due to enhanced replication in airway cells. *bioRxiv* 2020; published online Aug 30. https://doi.org/10.1101/2020.09.30.318311 (preprint).

19 McCallum M, De Marco A, Lempp F, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *bioRxiv* 2021; published online Jan 14. https://doi.org/10.1101/2021.01.14.426475 (preprint).

20 Public Health England. Investigation of SARS-CoV-2 variants of concern in England: technical briefing 6. Feb 13, 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961299/Variants_of_Concern_VOC_Technical_Briefing_6_England-1.pdf (accessed March 3, 2021).

21 Toovey OTR, Harvey KN, Bird PW, Tang JWW. Introduction of Brazilian SARS-CoV-2 484K.V2 related variants into the UK. *J Infect* 2021; published online Feb 3. https://doi.org/10.1016/j.jinf.2021.01.025.

22 Lam TT-Y, Zhu H, Guan Y, Holmes EC. Genomic analysis of the emergence, evolution, and spread of human respiratory RNA viruses. *Annu Rev Genomics Hum Genet* 2016; **17:** 193–218.

23 Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018; **34:** 4121–23.

24 Argimón S, Abudahab K, Goater RJE, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016; **2:** e000093.

25 Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 2020; **581:** 221–24.

26 Bonsall D, Golubchik T, de Cesare M, et al. A comprehensive genomics solution for HIV surveillance and clinical monitoring in low-income settings. *J Clin Microbiol* 2020; **58:** e00382–20.

27 Golubchik T, Lythgoe KA, Hall M, et al. Early analysis of a potential link between viral load and the N501Y mutation in the SARS-COV-2 spike protein. Dec 24, 2020. https://www.coronavirus-fraser-group.org/files/files/updated_report_to_nervtag_oxford_20201222.pdf (accessed Jan 31, 2021).

28 Singanayagam A, Patel M, Charlett A, et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill* 2020; **25:** 2001483.

29 Fajnzylber J, Regan J, Coxen K, et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat Commun* 2020; **11:** 5493.

30 Lauring AS, Hodcroft E. Genetic variants of SARS-CoV-2—what do they mean? *JAMA* 2021; **325:** 529–31.

31 Rickman H M, Rampling T, Shaw K, et al. Nosocomial transmission of coronavirus disease 2019: a retrospective study of 66 hospital-acquired cases in a London teaching hospital. *Clin Infect Dis* 2020; **72:** 690–93.

32 The New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG). Update note on B.1.1.7 severity, 11 February. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961042/S1095_NERVTAG_update_note_on_B.1.1.7_severity_20210211.pdf (accessed March 3, 2021).