SARS-CoV-2 lineage B.1.1.7 is associated with greater disease severity among hospitalised women but not men: multicentre cohort study

Oliver Stirrup 1, Florencia Boshier, 2 Cristina Venturini, 2 José Afonso Guerra-Assunção, 3,5 Adela Alcolea-Medina, 4,5 Angela Beckett, 6,7 Themoula Charalampous, 4 Ana da Silva Filipe, 8 Sharon Glaysher, 9 Tabassum Khan, 10 Raghavendran Kulasegaran Shylini 10,10 Beatrix Kele, 10 Irene Monahan, 11 Guy Mollett, 8 Matthew Parker, 12,13,14 Emanuela Pelosi, 15 Paul Randell, 16 Sunando Roy, 2 Joshua Taylor, 17 Sophie Weller, 18 Eleri Wilson-Davies, 15 Phillip Wade, 19,20 Rachel Williams, 3 The COG-UK-HOCI Variant substudy consortium, The COVID-19 Genomics UK (COG-UK) consortium, Andrew Copas, 1 Maria-Teresa Cutino-Moquiel, 10 Nick Freemantle, 21 Andrew C Hayward, 22 Alison Holmes, 23,24 Joseph Hughes, 8 Tabitha Mahungu, 18 Gaia Nebbia, 4,20 David Partridge, 19,20 Cassie Pope, 11,26 James Price 1,27 Samuel Robson, 6,28 Kordo Saeed, 29,30 Thushan de Silva, 19,20 Luke Snell, 4,25 Emma Thomson, 8 Adam A Witney, 11 Judith Breuer 2,31

ABSTRACT
Background SARS-CoV-2 lineage B.1.1.7 has been associated with an increased rate of transmission and disease severity among subjects testing positive in the community. Its impact on hospitalised patients is less well documented.

Methods We collected viral sequences and clinical data of patients admitted with SARS-CoV-2 and hospital-onset COVID-19 infections (HOCIs), sampled 16 November 2020 to 10 January 2021, from eight hospitals participating in the COG-UK-HOCI study. Associations between the variant and the outcomes of all-cause mortality and intensive therapy unit (ITU) admission were evaluated using mixed effects Cox models adjusted by age, sex, comorbidities, care home residence, pregnancy and ethnicity.

Findings Sequences were obtained from 2341 inpatients (HOCI cases=786) and analysis of clinical outcomes was carried out in 2147 inpatients with all data available. The HR for mortality of B.1.1.7 compared with other lineages was 1.01 (95% CI 0.79 to 1.28, p=0.94) and for ITU admission was 1.01 (95% CI 0.75 to 1.37, p=0.96). Analysis of sex-specific effects of B.1.1.7 identified increased risk of mortality (HR 1.30, 95%CI 0.75 to 2.25, p=0.32) and ITU admission (HR 1.50, 95%CI 1.15 to 1.95, p=0.0011) in females infected with the variant but not males (mortality HR 0.82, 95%CI 0.61 to 1.10, p=0.32; ITU HR 0.74, 95%CI 0.52 to 1.04, p=0.086).

Interpretation In common with smaller studies of patients hospitalised with SARS-CoV-2, we did not find an overall increase in mortality or ITU admission associated with B.1.1.7 compared with previously circulating variants. Lineage B.1.1.7 became the dominant lineage in the UK in winter 2020/2021. It has also been detected in over 120 countries outside the UK, and was assigned the label of variant of concern Alpha by WHO.

Key messages
► In a multicentre cohort we did not find an overall increase in mortality or intensive therapy unit admission associated with B.1.1.7 among hospitalised patients, but women with B.1.1.7 may be at an increased risk of admission to intensive care and at modestly increased risk of mortality.
► The impact of SARS-CoV-2 lineage B.1.1.7 on disease severity appears to be dependent on the patient’s sex, meaning that its health impacts and healthcare burden will differ from earlier lineages of the virus and that monitoring of disease course by sex should be considered for other viral lineages with increased transmissibility.

INTRODUCTION
The emergence of SARS-CoV-2 lineage B.1.1.7 in South East England has been found to be associated with an estimated 70% increase in mortality rate of community transmission compared with previously circulating variants. Lineage B.1.1.7 became the dominant lineage in the UK in winter 2020/2021. It has also been detected in over 120 countries outside the UK, and was assigned the label of variant of concern Alpha by WHO.

Lineage B.1.1.7 has acquired an unusually large number of mutations and deletions in a short period of time, specifically 14
non-synonymous single nucleotide polymorphisms and 3 amino acid deletions, with 8 of these 17 amino acid changes occurring in the spike protein, responsible for receptor binding and a major immunogenic target. At least three of the spike protein changes are associated with in vitro biological changes. A tyrosine substitution at position 501 in the spike protein receptor binding domain has been shown to increase binding to the ACE2 receptor, while deletion of spike protein amino acids 69/70 reduces antibody neutralisation by convalescent sera. The potential that so many mutations might change B.1.1.7 virulence has been examined epidemiologically using data largely from community-collected samples. However, there are few data on the impact of B.1.1.7 infection as compared with other variants on disease outcomes in hospitalised patients.

We investigated the potential associations between the B.1.1.7 variant and the outcomes of mortality and intensive therapy unit (ITU) admission both in patients admitted with COVID-19 and hospital onset COVID-19 infections (HOCI) in the COG-UK-HOCI study. The main objective was to estimate the overall effect of the variant on each of these outcomes, and we also evaluated whether the impact of the variant differed according to patient characteristics.

METHODS

Sequence and patient meta-data

Data were collected from five NHS hospitals within London and three outside. The first SARS-CoV-2-positive sample from all inpatients tested through hospital laboratories between 16 November 2020 and 10 January 2021 was sequenced. In addition, metadata were collected from clinical records on patient age, sex (as binary M/F), comorbidities as identified by the COVID-19 Greenbook (including obesity with body mass index ≥25 kg/m²), care home residence, pregnancy, ethnicity, date of hospital admission, ward location and first SARS-CoV-2-positive test for all samples plus dates of admission to the ITU and all-cause death where these events occurred.

Inpatients were classified as those admitted with SARS-CoV-2 plus cases which were identified after admission, with the latter termed HOCI cases and subdivided into indeterminate healthcare-associated infections (HCAs) diagnosed 3–7 days after admission and probable/definite HCAs diagnosed ≥28 days postadmission. The primary outcomes for analysis were the events of death and of ITU admission. Events were included in the analysis within 28 days of hospital admission for those admitted with COVID-19 and within 28 days of diagnosis for HOCI cases.

SARS-CoV-2 sequencing

Samples were sequenced by Nanopore or Illumina methods as part of the COVID-19 Genomics UK Consortium (COG-UK). To maximise success, 4/8 labs sequenced only those samples with quantitative PCR cycle thresholds (ct) values of ≤32 or equivalent. Sequences were assigned to lineages using COG-UK Pangolin.

Statistical analysis

Only patients with admission to hospital and HOCIs were included in the statistical analysis of the clinical outcomes of mortality and ITU admission. Mortality and ITU admission were modelled as time-to-event outcomes, from time of hospital admission for those admitted with COVID-19 and from time of diagnosis for HOCI cases, censored at 28 days. Analyses of ITU admission were also censored at patient death. Both outcomes were censored at date of data collection for these variables for each site (between 15 January and 22 February 2021). Mixed effects Cox models were used with adjustment for sex, patient age (using 5-knot restricted cubic spline), number of comorbidities (none, one, two, ≥three), care home residence, pregnancy, ethnicity (white, black, Asian, mixed or other) and sample week with separate parameters for London sites and for other sites grouped using the R package coxme V.2.2–16. A 5-knot restricted cubic spline was used for patient age in all analyses to allow flexibility in modelling the relationship with each outcome while maintaining a consistent model structure. Random intercept terms were included to reflect clustering of outcomes within hospital and weekly periods nested within hospitals. Cox models were stratified by HOCI status (allowing for different baseline hazard functions in patients admitted with COVID-19 vs HOCI groups).

Outcomes were analysed on a complete case basis with regard to patient characteristics. This decision was based on the availability of complete data for >90% of patients and the fact that Cox regression gives asymptotically unbiased estimates of an association of interest as long as the missingness is not dependent on both outcome (ie, death or ITU admission) and exposure (B.1.1.7 status). The variable of obesity was analysed as ‘morbid obesity’ versus ‘no record of morbid obesity’ on examination of case notes, and was included in statistical models within the ordinal comorbidities variable.

The primary aim of the analysis was to estimate the overall association between the B.1.1.7 vs non-B.1.1.7 strain and the risk of each of the outcomes considered. Exploratory secondary analyses also evaluated interactions between B.1.1.7 status and patient characteristics in estimating the effect on each outcome. Analyses were conducted in R V.4.0.2, using tidverse collection of packages with all plots generated using ggplot2 and survminer.
Results

Study dataset

Between 16 November 2020 and 10 January 2021, SARS-CoV-2 RNA-positive upper respiratory tract samples from 2341 inpatients were sequenced from the eight participating sites (table 1 and online supplemental figure S1). Analysis of clinical outcomes was carried out in 2147 inpatients with all data available. The prevalence of lineage B.1.1.7 was highest in London and Hampshire (South of England), but substantially increased at all sites over the study period (online supplemental figure S2).

Mortality outcome

Death within 28 days was reported in 527 (22.5%) of the 2341 patients. Death was recorded as having occurred following discharge with date of death missing in five, and these patients have been excluded from analyses. Death within 28 days was recorded in 494/2147 (23.0%) of the patients with all data available, with full 28 days of follow-up in 939/1653 (56.8%) of the other patients. On mixed effects multivariable Cox regression, the overall HR for mortality of lineage B.1.1.7 was 1.01 (95% CI 0.79 to 1.28, p=0.94) (figure 1, online supplemental table S1). Male sex was found to be a substantial risk factor for mortality (HR 1.46 vs female, 95% CI 1.22 to 1.75; p<0.001) and age was also strongly associated with the risk of death (figure 2). The risk of death was higher in care home residents (HR 1.39, 95% CI 1.02 to 1.90, p=0.04) and those with one or more significant comorbidities (HR 1.78 (95% CI 1.26 to 2.52) for one comorbidity, 2.03 (95% CI 1.43 to 2.88) for two and 2.89 (95% CI 2.04 to 4.08) for ≥three versus none; p<0.001). Those with ethnicity other than white were estimated to be at higher risk of death, but ethnicity was not a statistically significant predictor when evaluated over all categories (p=0.36). No pregnant women died and so this variable was dropped from the model as a perfect predictor.

The addition of an interaction term between B.1.1.7 status and patient age (p=0.48, LRT with 1 df), ethnicity (p=0.67, LRT with 3 df) or comorbidity category (p=0.33, LRT with 3 df).

A statistically significant interaction was found between the effect of B.1.1.7 and care home residence (p=0.03, LRT with 1 df), with those care home residents with B.1.1.7 infection estimated to be at lower risk of death (HR 0.52, 95% CI 0.27 to 1.02, p=0.057) with a non-significant conduct of this research and it would not be possible to disseminate the findings directly to participants.

Table 1 Proportion of SARS-CoV-2 due to lineage B.1.1.7 for all inpatient sequenced samples according to patient characteristics

<table>
<thead>
<tr>
<th>Sex</th>
<th>Lineage B.1.1.7 (n=1107)</th>
<th>Lineage B.1.1.7 (n=1234)</th>
<th>Total (n=2341)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>534 (46.1)</td>
<td>624 (53.9)</td>
<td>1158 (100)</td>
</tr>
<tr>
<td>Male</td>
<td>573 (48.4)</td>
<td>610 (51.6)</td>
<td>1183 (100)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample week starting</th>
<th>Age group (years)</th>
<th>Lineage B.1.1.7 (n=2337)</th>
<th>Lineage B.1.1.7 (n=2341)</th>
<th>Total (n=2341)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 November 2020</td>
<td>0–11</td>
<td>15 (62.5)</td>
<td>9 (37.5)</td>
<td>24 (100)</td>
</tr>
<tr>
<td></td>
<td>12–24</td>
<td>20 (58.8)</td>
<td>14 (41.2)</td>
<td>34 (100)</td>
</tr>
<tr>
<td></td>
<td>25–34</td>
<td>61 (62.9)</td>
<td>36 (37.1)</td>
<td>97 (100)</td>
</tr>
<tr>
<td></td>
<td>35–49</td>
<td>159 (56.2)</td>
<td>124 (43.8)</td>
<td>283 (100)</td>
</tr>
<tr>
<td></td>
<td>50–69</td>
<td>371 (53.2)</td>
<td>326 (46.8)</td>
<td>697 (100)</td>
</tr>
<tr>
<td></td>
<td>70–79</td>
<td>208 (42.2)</td>
<td>285 (57.8)</td>
<td>493 (100)</td>
</tr>
<tr>
<td></td>
<td>80+</td>
<td>273 (38.3)</td>
<td>440 (61.7)</td>
<td>713 (100)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient class</th>
<th>Probability/definite HCAI†</th>
<th>Indeterminate HCAI‡</th>
<th>CAI*</th>
<th>HCW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>199 (34.7)</td>
<td>54 (25.4)</td>
<td>847 (55.1)</td>
<td>7 (36.8)</td>
</tr>
<tr>
<td></td>
<td>374 (65.3)</td>
<td>159 (74.6)</td>
<td>689 (44.9)</td>
<td>12 (63.2)</td>
</tr>
<tr>
<td></td>
<td>573 (100)</td>
<td>239 (100)</td>
<td>1536 (100)</td>
<td>19 (100)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Ethnicity</th>
<th>Patient characteristics</th>
<th>Obesity (BMI ≥35)</th>
<th>Pregnant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glasgow</td>
<td>White</td>
<td>540 (39.4)</td>
<td>829 (60.6)</td>
<td>1669 (100)</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>174 (53.4)</td>
<td>152 (46.6)</td>
<td>326 (100)</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>118 (63.1)</td>
<td>69 (36.9)</td>
<td>187 (100)</td>
</tr>
<tr>
<td></td>
<td>Mixed or other</td>
<td>186 (67.1)</td>
<td>91 (32.9)</td>
<td>277 (100)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>89 (48.9)</td>
<td>93 (51.1)</td>
<td>182 (100)</td>
</tr>
<tr>
<td></td>
<td>Obese</td>
<td>122 (51)</td>
<td>117 (49)</td>
<td>239 (100)</td>
</tr>
<tr>
<td></td>
<td>Pregnant</td>
<td>25 (55.6)</td>
<td>20 (44.4)</td>
<td>45 (100)</td>
</tr>
</tbody>
</table>
The addition of an interaction term between B.1.1.7 status and patient sex for the effect on ITU admission led to an improvement in model fit (p=0.0004 interaction test, p=0.002 lineage B.1.1.7 effects by sex vs no B.1.1.7 effect, LRTs). The estimated HR for ITU admission for lineage B.1.1.7 vs non-B.1.1.7 was 0.74 (95% CI 0.52 to 1.04, p=0.086) in male patients and 1.82 (95% CI 1.15 to 2.90, p=0.011) in female patients. There was no evidence for an interaction of B.1.1.7 status with patient age (p=0.11, LRT with 4 df), ethnicity (p=0.74, LRT with 3 df), comorbidity category (p=0.79, LRT with 3 df), pregnancy (p=0.42, LRT with 1 df) or care home residence (p=0.24, LRT with 1 df) with ITU admission as the outcome. Kaplan-Meier plots of ITU admission in relation to B.1.1.7 status are presented according to patient sex and age categories in figure 4 (also provided separately for non-HOCI and HOCI inpatients in online supplemental figures S5 and S6, with HR estimates in online supplemental table S2).

### DISCUSSION

Our findings provide the largest dataset on disease severity in hospitalised patients with lineage B.1.1.7 and the only one based on routine sequencing of all specimens from multiple hospitals. The overall hazard of mortality and ITU were unchanged for patients with lineage B.1.1.7 in comparison to other viral variants (HR 1.01, 95% CI 0.79 to 1.28 and HR 1.01, 95% CI 0.75 to 1.37, respectively). These findings are in line with the results of a much smaller analysis of 341 (n=198 with B.1.1.7) hospital inpatients with viral sequencing over a similar time period in London, which found an adjusted mortality risk ratio for B.1.1.7 of 1.02 (95% CI 0.76 to 1.38). However, in contrast with this smaller study we also found evidence that B.1.1.7 infection appears to have a different impact on the disease course according to sex among hospitalised patients with SARS-CoV-2 infection, with increased hazard of both mortality and ITU admission associated with the variant for female but not male patients.

Several larger studies of disease severity in the UK have used PCR Spike (S) gene target failure (SGTF) as a surrogate for lineage B.1.1.7. These studies, based on community testing data, all found evidence of an overall increased risk of mortality associated with lineage B.1.1.7, with reported HRs of 1.64 (95% CI 1.32 to 2.04) by
Figure 1 Results of mixed effect Cox regression models for death and intensive therapy unit (ITU) admission, shown as HR (●, ■ for reference categories) with 95% CI. Results are displayed for outcomes of mortality (A, B) and ITU admission (C, D), both for the overall effect of B.1.1.7 variant (A, C) and with sex-specific effects of B.1.1.7 (B, D). Models were all also adjusted by age using natural cubic splines (as shown in figure 2).
Challen et al.8 1.55 (95% CI 1.39 to 1.72) by Davies et al.9 1.67 (95% CI 1.34 to 2.09) by Grint et al.10 and 1.59 (95% CI 1.25 to 2.03) by Patone et al.11 In the UK, SGTF is only available as a marker for a subset of those patients who were first positive for SARS-CoV-2 on testing within the community; most people who die of COVID-19 were not previously tested within the community9 and the relevant PCR assay is not used by all laboratories, meaning that SGTF status is only available for 8.6% of deaths.9 SGTF is an imperfect predictor of lineage B.1.1.7, and is much less accurate as a marker when prevalence of the variant is low (before mid-November 2020 in the UK).24

The apparent overall differences in mortality risk observed in the SGTF analyses in comparison to our study do not necessarily represent inconsistent findings. Studies that are limited to patients who test positive in the community may be subject to selection biases linked to propensity to present for testing or rapidity of disease progression, while analyses that include only data from inpatients will not reflect the characteristics of the population as a whole. For example, increased disease severity may result in a higher proportion of subjects reaching the threshold for admission to hospital but not affect the mortality rate among those who are admitted to hospital. Our study also includes a subset of patients with probable nosocomial infection, whose characteristics and comorbidity profile differs greatly from the UK population as whole.25

Individuals testing positive in the community for an SGTF-associated variant had higher risk of hospitalisation, with OR of 1.58 (95% CI 1.50 to 1.67).26 This result was confirmed by a study of national health register-data from Denmark including 18 499 patients with viral genomes available in the period 1 January to 9 February 2021 which found an adjusted OR of 1.64 (95% CI 1.32 to 2.04) for hospitalisation for B.1.1.7 compared with other lineages.27 Taken together with the findings regarding mortality in the UK,8–11 these results are consistent with an increased risk of mortality and hospitalisation among patients testing positive for B.1.1.7 in the community but no overall increase in mortality among the subset of patients admitted to hospital.

We found a significantly increased risk of both mortality (30%) and ITU admission (82%) in hospitalised female patients infected with B.1.1.7 but not in male patients. In contrast, studies of community-tested individuals found no interaction with sex for the effect of B.1.1.7 on mortality,9 11 critical care admission11 or risk of hospitalisation.26 However, these studies were all conducted among patients who first tested positive for SARS-CoV-2 within the community, and therefore they cannot rule out an interaction with sex for the impact of B.1.1.7 on disease severity among all people infected with the virus or among all those admitted to hospital. Nationally collated data show that females accounted for 33.2% of patients admitted to ITU with COVID-19 in London, East and South East England between 1 September and 30 November 2020 rising to 36.2%, between 1 December 2020 and 21 January 2021 when lineage B.1.1.7 predominated.28

There is evidence that the total number of pregnant women requiring intensive care in the UK was higher in late 2020/early 2021 in comparison to the first wave of the COVID-19 epidemic in Spring 2020,29 and that the proportion of symptomatic pregnant women requiring admission to ITU increased as lineage B.1.1.7 became dominant.30 This would be consistent with a differential impact of the B.1.1.7 variant on women in comparison to men. However, we should also note that our analyses found pregnancy itself to be negatively associated with...
ITU admission, conditional on age, sex, viral variant, comorbidities and other patient characteristics. This could possibly be due to unobserved or residual confounding not fully captured by our recording and analysis of patient characteristics.

An impact of lineage B.1.1.7 on females that is not observed in males could potentially be explained by physiological differences. For example, increased ACE2 expression in females has been posited as one explanation for the relatively lower mortality and morbidity observed for COVID-19 for women in comparison to men. Lineage B.1.1.7 has mutations that increase binding of the viral spike protein to ACE2, thereby providing a plausible mechanism by which the new variant might have a differential effect on disease severity in males and females.

Our results suggest a reduction in the risk of mortality or ITU admission associated with B.1.1.7 in comparison to other viral lineages among male inpatients, although this finding was not definitive with HR 95% CIs that included no effect for both outcomes. However, this is against a backdrop of the established increased risk for male versus female patients for pre-B.1.1.7 strains. Our results indicate an approximate equalisation of risk for otherwise equivalent male and female inpatients with B.1.1.7 infection.

Although ours is substantially the largest study of hospitalised patients with confirmed lineage B.1.1.7 and non-B.1.1.7 SARS-CoV-2 infection, it has a number of limitations. Primarily, while evaluation of disease severity among only hospital inpatients can give useful
information on disease course and progression, analysis of only these patients cannot provide information on disease severity across all SARS-CoV-2 infections in the population as a whole. In addition, ITU admission can be difficult to interpret as a measure of disease severity among inpatients. For instance, admission to ITU may reflect the presence of severe disease and local decisions around the benefit or lack thereof to frail patients, which may be influenced by bed numbers and availability of respiratory support in non-critical care settings. Our primary analysis also includes cases of hospital-acquired infection but exclusion of these HOCI cases from our analyses yielded similar findings (online supplemental table S2).

A further limitation of our analysis is that we do not have any information on vaccination status for individual patients. Our dataset covers a period in which a national vaccination programme was being initiated for HCWs and the elderly population in the UK, starting with those aged 80 years and above from 8 December 2020. This is a potential explanation for the observed protective interaction effect between care home residence and B.1.1.7 on mortality, as care home residents were prioritised for vaccination around the time that this viral variant was increasing in prevalence. Vaccine breakthrough infections are well described, particularly in partially vaccinated subjects.

Our analysis was focused on the effect of lineage B.1.1.7 as a single exposure of interest for each outcome variable.
However, we also considered 5 potential interactions with viral lineage for the outcome of mortality and 6 for ITU admission, giving a total of 13 hypothesis tests of interest. A conservative use of the Bonferroni correction gives a p value cut-off adjusted to 0.0038 from the commonly used threshold of 0.05, using which we would consider there to be robust statistical evidence for an interaction with sex on the outcome of ITU admission but not for the interactions between sex or care home status on mortality.

Implications
Although lineage B.1.1.7 was not associated with an overall increase in mortality among hospitalised patients, our investigation suggests that lineage B.1.1.7 may be associated with higher ITU admission and death in females compared with non-B.1.1.7 within this group. The dominance of lineage B.1.1.7 in the UK precluded further comparison with earlier non-B.1.1.7 variants, and there is now concern regarding the spread of other lineages in the UK and elsewhere.36 There is a need for ongoing large-scale sequencing of SARS-CoV-2 cases linked to data on patient characteristics and outcomes in order to generate timely information regarding the associations between viral lineages and disease severity. Monitoring of disease course by sex should be considered for other viral lineages with increased transmissibility.

Author affiliations
1Institute for Global Health, University College London, London, UK
2Department of Infection, Immunity and Inflammation, UCL Great Ormond Street Institute of Child Health, University College London, London, UK
3Department of Genetics & Genomic Medicine, UCL Great Ormond Street Institute of Child Health, University College London, London, UK
4Centre for Clinical Infection and Diagnostics Research, School of Immunology and Microbial Sciences, King’s College London, London, UK
5Infection Sciences, Viapath, London, UK
6Centre for Enzyme Innovation, University of Portsmouth, Portsmouth, UK
7School of Biological Sciences, University of Portsmouth, Portsmouth, UK
8MRC-University of Glasgow Centre for Virus Research, Glasgow, UK
9Portsmouth Hospitals University NHS Trust, Queen Alexandra Hospital, Portsmouth, UK
10Division of Infection, The Royal London Hospital, Barts Health NHS Trust, London, UK
11Institute for Infection and Immunity, St George’s University of London, London, UK
12Sheffield Bioinformatics Core, The University of Sheffield, Sheffield, UK
13Sheffield Institute for Translational Neuroscience, The University of Sheffield, Sheffield, UK
14Sheffield Biomedical Research Centre, The University of Sheffield, Sheffield, UK
15Southampton Specialist Virology Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK
16Department of Infection and Immunity, North West London Pathology, London, UK
17Department of Microbiology, South West London Pathology, St. George’s Hospital, London, UK
18Department of Virology, Royal Free London NHS Foundation Trust, London, UK
19Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
20The Florey Institute for Host-Pathogen Interactions & Department of Infection, Immunity and Cardiovascular Disease, Medical School, University of Sheffield, Sheffield, UK
21Institute of Clinical Trials and Methodology, University College London, London, UK
22Institute of Epidemiology and Health Care, University College London, London, UK
23Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, UK
24Hammersmith Hospital, Imperial College Healthcare NHS Trust, London, UK
25Department of Infectious Diseases, Guy’s and St Thomas’ Hospital NHS Foundation Trust, London, UK
26Infection Care Group, St George’s University Hospitals NHS Foundation Trust, London, UK
27Imperial College Healthcare NHS Trust, London, UK
28School of Pharmacy and Biomedical Sciences, University of Portsmouth, Portsmouth, UK
29Microbiology Innovation and Research Unit (MiRfU), Department of Microbiology, University Hospital Southampton NHS Foundation Trust, Southampton, UK
30Faculty of Medicine, Clinical and Experimental Sciences, University of Southampton, Southampton, UK
31Department of Microbiology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Collaborators
Composition of the COG-UK-HOCI Variant substudy consortium members (excluding those on main authorship): study oversight and quality assurance UCL Comprehensive Clinical Trials Unit: James Blackstone, Leanne Hockley, Georgia Morley. Collection of samples, sequence data and meta-data: Barts Health Trust, David Harrington, Anna Riddell; Queen Elizabeth University Hospital, NHS Greater Glasgow & Clyde (GEUH), Christine Peters; Guy’s and St Thomas’ NHS Foundation Trust and Centre for Clinical Infection & Diagnostics Research, King’s College London, Flavia Flaviani, Binid Patel, Tom G S Williams, Rahul Batra, Jonathan D Edgeworth; North West London Pathology, Pinglawate Madona, Alison Cox; Royal Free London NHS Foundation Trust, Jennifer Hart, Tanzia Haque, Dianne Irish; UCL Pathogen Genomics Unit, Juanita Pang, Charlotte Williams, Helena Tuttil, Nadua Bayzid, Marius Cotic; University of Sheffield and Sheffield Teaching Hospitals NHS Foundation Trust, Luke Green, Benjamin Lindsey, Amy State, Alison Cope, Katie Johnson, Adrienne Angyal, Peilun Zhang, Max Whiteley, Marta Gallis Ramalotta, Stella Christou, Stavroula Louka, Halley Hornsby, Benjamin Foulkes, Paige Wolverson, Joe Heffer, Nikki Smith; University of Portsmouth, Salman Goudarzi, Chris Fearn, Kate Cook, Katie Loveson; Portsmouth Hospitals University NHS Trust, Scott Elliott; University Hospital Southampton NHS Foundation Trust, Adhyanah Mahamana, Buddhini Samaraweera, Siona Silveira, Stephen Aplin, Sarah Jeremiah, Helen Uempleby, Helen Wheeler, Matthew Harvey, Thea Sass, Jacqui Prieto; St Georges University and Healthcare Trust, Kenneth Laing, Ngee Keong Tan, Claudia Cardoso Pereira; University College London Hospitals NHS Foundation Trust UCLH Advanced Pathogen Diagnostics Unit, Eleni Nasboul, Catherine F Houllian, Dan Frampton, Tommy Rampling, Matt Boytt, Judith Heaney, Gee Yen Shin, Moira Spyer, Malin Bergstrom, Emilie Sanchez, Stavroula M Paraskevopoulos, Marios Margaritis.

Contributors
OS, FB, CV, JAG-A, ACJ, NF, ACH and JB planned the analysis and drafted the first draft of the manuscript. AA-M, AB, AC, AdSF, SG, TK, RKS, BK, IM, GM, MP, EP, PR, SR, JT, SW, EW-D, PW, PW, M-TC-M, AH, JH, TM, GN, BP, CP, JP, SR, KS, TòS, LS, ET, AAW extracted and provided sequencing data and patient characteristics and outcome data. OS, FB, CV and JAG-A had full access to and verified the final collated dataset. FB, CV and JAG-A carried out phylogenetic lineage assignments and merged the final dataset for analysis, and OTS carried out statistical modelling. All authors reviewed the final manuscript and approved this for submission. OTS is the guarantor of the study. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding
This report was produced by members of the COG-UK-HOCI Variant substudy consortium. COG-UK-HOCI is part of COG-UK. COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute.

Disclaimer
The funder of the study had no role in the study design, the collection, analysis and interpretation of data, in the writing of the report or in the decision to submit the paper for publication.

Competing interests
All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: OTS has received funding for the submitted work through the COG-UK-HOCI study, funded by COG-UK consortium, supported by funding from UK Research & Innovation, National Institute of Health Research and Wellcome Sanger Institute; the COG-UK consortium
funded sequencing costs for the submitted work; NF reports grants from UKRI, during the conduct of the study; personal fees from Aimmune, personal fees from ALK, personal fees from AstraZeneca, personal fees from MSD, personal fees from Sanofi Aventis, personal fees from Novartis, personal fees from Ipsen, personal fees from Gedeon Richter, personal fees from Galderma, personal fees from Vertex, outside the submitted work. The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.

Ethics approval Ethical approval for the HOCl study, with an amendment to cover the substudy reported in this research paper, is provided by REC 20/EE/0118.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The sequence data analysed are included within publicly available datasets (https://www.cogconsortium.uk/data/). However, due to data governance restrictions it is not possible to share the associated patient data. Where the content to data governance restrictions it is not possible to share the associated patient publicly available datasets (https://www. 

The substudy reported in this research paper, is provided by REC 20/EE/0118.

The remaining authors do not have any declarations of interest. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not required.