The effect of SARS-CoV-2 variant B.1.1.7 on symptomatology, re-infection and transmissibility

Mark S. Graham1*, Carole H. Sudre1,2,3*, Anna May4, Michela Antonelli1, Benjamin Murray1, Thomas Varsavsky1, Kerstin Kläser1, Liane S. Canas1, Erika Molteni1, Marc Modat1, David A. Drew5, Long H. Nguyen5, Lorenzo Polidori4, Somesh Selvachandran6, Christina Hu4, Joan Capdevila Pujol4, The COVID-19 Genomics UK (COG-UK) consortium6+, Alexander Hammers1, Andrew T. Chan5, Jonathan Wolf4, Tim D. Spector7, Claire J. Steves7+, Sebastien Ourselin1+

1. School of Biomedical Engineering & Imaging Sciences, King’s College London, London, UK
2. MRC Unit for Lifelong Health and Ageing, Department of Population Science and Experimental Medicine, University College London, UK
3. Centre for Medical Image Computing, Department of Computer Science, University College London, UK
5. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
6. https://www.cogconsortium.uk
7. Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK

* Equal contribution
+Equal contribution
+Full list of consortium names and affiliations are in the appendix

Corresponding Author:
Mark Graham, PhD
School of Biomedical Engineering and Imaging Sciences
King’s College London,
Lambeth Palace Road,
SE1 7EH
mark.graham@kcl.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

The new SARS-CoV-2 variant B.1.1.7 was identified in December 2020 in the South-East of England, and rapidly increased in frequency and geographic spread. While there is some evidence for increased transmissibility of this variant, it is not known if the new variant presents with variation in symptoms or disease course, or if previously infected individuals may become reinfected with the new variant. Using longitudinal symptom and test reports of 36,920 users of the Covid Symptom Study app testing positive for COVID-19 between 28 September and 27 December 2020, we examined the association between the regional proportion of B.1.1.7 and reported symptoms, disease course, rates of reinfection, and transmissibility. We found no evidence for changes in reported symptoms, disease severity and disease duration associated with B.1.1.7. We found a likely reinfection rate of around 0.7% (95% CI 0.6-0.8), but no evidence that this was higher compared to older strains. We found an increase in R(t) by a factor of 1.35 (95% CI 1.02-1.69). Despite this, we found that regional and national lockdowns have reduced R(t) below 1 in regions with very high proportions of B.1.1.7.

Introduction

In early December 2020, a phylogenetically distinct cluster of SARS-CoV-2 was genetically characterised in the South-East of England. The majority of cases had been detected in November with a small number detected as early as September\(^1\). Genomic surveillance reveals that this new variant, termed B.1.1.7, has a number of mutations of immunologic significance and is growing rapidly in frequency and spread.\(^2\)

Preliminary evidence from epidemiological studies suggests the new strain is more transmissible. Davies et. al. found the new strain is 56% (95% CI 50-74) more transmissible\(^3\) and Volz et. al. found the new strain increases the effective reproduction number R(t) by a factor of 1.4-1.8\(^4\). There is early data to suggest B.1.1.7 increases risk of death by \(\sim1.3\)^. However, there is much that is still unknown. Little is known about the disease course of infections due to the new variant. Early evidence suggested that the new variant does not affect rates of hospitalisation\(^3\), but it is crucial to assess whether the new variant alters the symptomatology, duration, and severity of disease. It is also important to understand whether B.1.1.7 alters the rate of asymptomatic infection and reinfection. Furthermore, early estimates of the new transmissibility of B.1.1.7 are uncertain and there is a need for additional estimates using independent data sources.

We make use of data from the COVID Symptom Study (CSS)\(^6\) to investigate the symptomatology, disease course, and transmissibility of the new variant. The longitudinal dataset provides symptom reports and test results from a population of over 4 million adults living in the UK using the mobile application. By combining these data with surveillance data from the Covid-19 UK Genetics Consortium (COG-UK)\(^7\) and a spike-gene target failure correlate in community testing data, we performed associative studies to study the symptoms, disease course, rates of reinfection, and transmissibility of the new variant.
Methods

Symptom study data

Longitudinal data were prospectively collected using the CSS app, developed by Zoe Global with input from King’s College London (London, UK), the Massachusetts General Hospital (Boston, MA, USA), and Lund and Uppsala Universities (Sweden). The app guides users through a set of enrolment questions, establishing baseline demographic and health information. Users are asked to record each day whether they feel physically normal, and if not, to log any symptoms. Users are also asked to maintain a record of any COVID-19 tests, their type, and their results in the app. Users are able to record the same data on behalf of others, such as family members, to increase data coverage amongst those unlikely to use mobile applications, such as the elderly. More details about the app can be found in a study by Drew and colleagues.

Genomic data

We used data released on 13 January 2020 from COG-UK to extract time-series of the percentage of daily cases that came from the B.1.1.7 lineage in Scotland, Wales, and each of the seven National Health Service (NHS) regions in England. Northern Ireland was excluded due to the low number of samples in the COG-UK dataset. These data are produced by sequencing a sample of polymerase chain reaction (PCR) tests carried out in the community. Due to the delay of approximately two weeks between PCR and genomic sequencing, we only used data from samples taken up to 31 December to avoid censoring effects.

Additionally, we used data from Public Health England (PHE) on the probable new variant captured in community cases in England using spike gene target failure (SGTF). It has been observed that one of the spike gene mutations in B.1.1.7 causes an SGTF in the test used in three of England’s large laboratories used for analysis of community cases. This failure results in a marker that is sensitive to B.1.1.7, but not necessarily specific, as other circulating variants also contain the mutation leading to an SGTF. Comparison to genomic data finds that from 30 November 2020 onwards more than 96% of cases with the SGTF were from lineage B.1.1.7. The proportion of SGTF cases is made available in England for each of the 316 “Lower Tier” Local Authorities. We grouped these data to each NHS region using a population-weighted average to enable integration with other data sources.

Disease symptoms and course

In order to assess whether the symptomatology of infection from B.1.1.7 differed from previous variants, we investigated the change in symptom reporting from 28 September to 27 December 2020, covering 15 complete weeks over the period when the proportion of B.1.1.7 grew most notably in London, South East and East of England. We took the symptom reports from users reporting a positive swab test (PCR or lateral flow) in this period and examined the association between the proportion of B.1.1.7 in each region and the proportion of reports per week for each
symptom, accounting for age, sex, and two seasonal environmental confounders: regional temperature and humidity, in a linear regression. Seasonal confounders were calculated each day as the average of the temperature and relative humidity at two meters above the surface, averaged across each region considered.\textsuperscript{9}

We also examined the relation between proportion of B.1.1.7 and disease burden, measured here as the total number of different symptoms reported over a period of two weeks before and two weeks after the test, and the relation with asymptomatic infection, defined as users reporting a positive test result but no symptoms in the two weeks before or after the test. Using similar corrections for demographic and seasonal environmental confounders, we investigated the rate of self-reported hospital visits. We also investigated the proportion of individuals reporting long symptom duration using a previously published definition of continuous symptoms reported for at least 28 days.\textsuperscript{10} To avoid censoring effects, both hospitalisation and long duration analyses included symptom reports extended up to 18 January, and the long duration analysis only considered reports of positive tests up to 21 December.

Reinfection
We defined possible reinfection as the presence of two reported positive tests separated by more than 90 days with a period of reporting no symptoms for more than seven days before the second positive test. We calculated the proportion of possible reinfection among individuals reporting their first positive test before 1 October 2020 and the correlation between number of possible reinfections and number of reported positive tests. To assess whether the risk of reinfection was stronger in the presence of the new variant, we calculated the correlations between the number of possible reinfections and the proportion of B.1.1.7 cases regionally over time.

Transmissibility
Daily incidence for Scotland, Wales, and each of the seven NHS regions in England were produced from the period 1 October 2020 to 27 December 2020 using data from the CSS app and previously described methodology\textsuperscript{11}. These data were used to determine the number of new daily cases from both old variants and from B.1.1.7 in each region. $R(t)$ was estimated separately for the old and new variants using methods described in\textsuperscript{11}. We compared both multiplicative and additive differences of the new and old $R$ values for days when the proportion of B.1.1.7 in a region was greater than 3%. While data is not available for the proportion of B.1.1.7 in January, we also computed total incidence and $R$ from 1 October to 16 January to see the effect of national lockdown in England on these measures.
Results

Symptom study data

Table 1 shows the demographic data for the cohort studied. From 24 March to 27 December 2020, 4,327,245 participants from the UK signed up to use the app. We excluded users living in Northern Ireland due to the low number of sign-ups (38,976), 383,352 users lacking information on sex, and 2,175,979 who had not logged in the app during the period 28 September to 27 December 2020, leaving a total of 1,767,914 users. Between them, these users recorded 65,606,869 logs in the app between 28 September and 27 December. In this period, 497,989 users reported a swab test. 55,192 of these reported a positive test, and we investigated the symptom reports of 36,920 of those whose region was known and who reported as healthy on app sign-up.

<table>
<thead>
<tr>
<th>Overall</th>
<th>Tested</th>
<th>Tested positive</th>
<th>Signed up healthy with reporting around positive test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Users</td>
<td>1,767,914</td>
<td>---</td>
<td>497,989</td>
</tr>
<tr>
<td>Daily reports*</td>
<td>65,613,697</td>
<td>19,154,601</td>
<td>1,514,244</td>
</tr>
<tr>
<td>Age in years mean (std)</td>
<td>48.4 (19.3)</td>
<td>---</td>
<td>46.06 (17.8)</td>
</tr>
<tr>
<td>≤18</td>
<td>163,112</td>
<td>9.2</td>
<td>40,717</td>
</tr>
<tr>
<td>19 - 64</td>
<td>1,234,259</td>
<td>69.8</td>
<td>381,900</td>
</tr>
<tr>
<td>≥ 65</td>
<td>370,543</td>
<td>20.9</td>
<td>72,741</td>
</tr>
<tr>
<td>Invalid</td>
<td>5,576</td>
<td>0.3</td>
<td>2,631</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1,046,074</td>
<td>59.2</td>
<td>315,875</td>
</tr>
<tr>
<td>Male</td>
<td>720,562</td>
<td>40.8</td>
<td>181,110</td>
</tr>
<tr>
<td>Intersex</td>
<td>79</td>
<td>&lt;0.1</td>
<td>21</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>1,199</td>
<td>0.1</td>
<td>983</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South East</td>
<td>342,881</td>
<td>19.4</td>
<td>97,143</td>
</tr>
</tbody>
</table>
Table 1. Characteristics of GB app users active in the period 28 September - 27 December 2020
* Reports logged between 28 September - 27 December. For some analyses we took further reports from an extended time period 14 September 2020 - 18 January 2021
**May be more than one test per individual as the overall number contains failed tests and unknown results

<table>
<thead>
<tr>
<th>Region</th>
<th>Total</th>
<th>Report</th>
<th>Positive</th>
<th>Positive</th>
<th>Negative</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>East of England</td>
<td>196,063</td>
<td>11.1</td>
<td>57,680</td>
<td>11.6</td>
<td>5,373</td>
<td>9.8</td>
</tr>
<tr>
<td>London</td>
<td>227,004</td>
<td>12.8</td>
<td>81,940</td>
<td>16.5</td>
<td>9,733</td>
<td>17.8</td>
</tr>
<tr>
<td>Midlands</td>
<td>198,350</td>
<td>11.2</td>
<td>57,582</td>
<td>11.6</td>
<td>6,695</td>
<td>12.2</td>
</tr>
<tr>
<td>North East and Yorkshire</td>
<td>156,999</td>
<td>8.9</td>
<td>42,986</td>
<td>9.1</td>
<td>5,292</td>
<td>9.7</td>
</tr>
<tr>
<td>North West</td>
<td>123,201</td>
<td>7.0</td>
<td>45,156</td>
<td>9.1</td>
<td>6,180</td>
<td>11.3</td>
</tr>
<tr>
<td>South West</td>
<td>186,372</td>
<td>10.5</td>
<td>46,780</td>
<td>9.4</td>
<td>3,685</td>
<td>6.7</td>
</tr>
<tr>
<td>Scotland</td>
<td>87,263</td>
<td>4.9</td>
<td>13,793</td>
<td>2.8</td>
<td>1,589</td>
<td>2.9</td>
</tr>
<tr>
<td>Wales</td>
<td>82,886</td>
<td>4.7</td>
<td>16,471</td>
<td>3.3</td>
<td>3,092</td>
<td>5.6</td>
</tr>
<tr>
<td>Not known</td>
<td>165,164</td>
<td>9.3</td>
<td>38,458</td>
<td>7.5</td>
<td>4,638</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Genomic data

In the period between 27 September 2020 and 31 December 2020, 98,170 sequences were made available by COG-UK, corresponding to 4.4% of the 2,207,476 cases recorded in this period.\textsuperscript{12} 16,224 sequences (16.5%) were variant B.1.1.7. Considering the mean of the rolling average across December, the three regions with the largest proportion of B.1.1.7 are the South East, London, and East of England. The three regions with the lowest proportion are Wales, the North East and Yorkshire, and the North West. SGTF data was made available in England on a weekly basis from 10 November 2020 to 29 December 2020. Of the 700,590 cases reported in this period, 295,404 (42.2%) caused an SGTF. Examining the COG-UK data from England in the same time period, we find 34.6% cases are B.1.1.7. The difference is in part attributable to the SGTF being a nonspecific marker of B.1.1.7: in the week from 9-15 November 81% of cases with an SGTF were B.1.1.7, while from 30 November at least 96% of cases with the SGTF were from B.1.1.7. Figure 1 shows how the proportion of the new variant changed over time in regions of the UK using COG-UK and the SGTF data.
Figure 1. Presence of B.1.1.7 in each of the 7 NHS regions in England, and Scotland and Wales, as measured using genomic surveillance data (COG-UK) and SGTF data. SGTF data are not available for Scotland or Wales.

Disease symptoms and course

Figure 2 illustrates the variation of symptom occurrence over time considering a one-week window smoothed over 3 time points as a function of time, and Supplementary Figure 1 shows how these symptoms vary as a function of the proportion of B.1.1.7. These results show no change in the proportion of users reporting each symptom with the new variant.

Figure 3 shows the variation of total number of symptoms reported, the total number of asymptomatic infections, self-reported hospital visits, and symptoms of long duration over time; Supplementary Figure 2 shows how these plots vary with proportion of B.1.1.7. When correcting for mean age, sex, ambient temperature and humidity there was no evidence of an association between B.1.1.7 and either the number of symptoms reported over a 4-week window, the number of hospitalisations, long symptom duration, or proportion of asymptomatic case (Supplementary Table 1).
**Figure 2.** Regional plots of the frequency of reporting of symptoms over time for each reported symptom. Drop in fever reporting in early November was caused by a change in the question wording; this wording was subsequently reverted a week later.
Figure 3. Regional plot of hospitalisation report, proportion of asymptomatic, report of long duration and number of experienced symptoms around test. For the study of long duration, tests are only considered up to 21 December counting reports up to 18 January 2021 to limit right censoring. Only symptomatic individuals for which duration can be ascertained are included.

Reinfection

Overall, we identified 304 individuals reporting two positive tests with more than 90 days between the two. Among these individuals, symptom reporting allowed us to identify 249 for which there is a period of at least 7 symptom-free days in between positive tests among the 36,509 individuals having reported a positive swab test before 1 October 2020 (0.7%, 95% CI 0.6-0.8). Among those 249, daily reports were available in the periods around both of the positive tests for 173. There was no difference in reinfection reporting rates across the different NHS regions (p=0.1). Figure 4 shows the evolution in the number of possible reinfections along
with reported positive cases (red line) and proportion of B.1.1.7 (green line). For all regions (except Scotland), reinfection occurrences were more positively correlated with the overall regional rise in cases rather than the regional rise in the new variant percentage (Number of cases: reinfection, Spearman rho 0.55 to 0.69 \([p<0.05]\) for South East, London and East of England; \% new variant: reinfection, Spearman rho 0.37 to 0.55 in the same regions). Supplementary Table 2 shows the bootstrapped median values of correlation compared across the different regions and the outcome of a Mann-Whitney U test across the bootstrapped distributions.

**Figure 4.** Number of reinfection reports by region according to week of second infection, along with the total number of positive tests reported through the app and the proportion of B.1.1.7 in circulation.
Transmissibility

Figure 5 shows incidence and R(t) for the old and new variants in the three regions in England with the highest proportions of the new variant. Results consistently show the R(t) of B.1.1.7 to be greater than that of other variants. The mean (95% CI) of the additive increase in R for B.1.1.7 was 0.34 (0.02-0.66), and the multiplicative increase was 1.35 (1.02-1.69). England exited its second national lockdown on 2 December, leading to a change in behaviour and R(t). When considering only the period after the second lockdown ended, we find 0.28 (0.01-0.61) for the additive and 1.28 (1.02-1.61) for the multiplicative increases. Supplementary Figure 3 shows the same using the SGTF data, with analysis limited to the period after 1 December when at least 95% of all SGTF cases were B.1.1.7. These data are provided weekly, and linear interpolation was used to obtain daily estimates, leading to smoother estimates for variant-specific incidence and R(t). Using these values, we find R(t) of B.1.1.7 has an additive increase of 0.26 (0.15-0.37) and a multiplicative increase of 1.25 (1.17-1.34).

On 19 December 2020 London and much of the South East and East of England were placed in ‘Tier 4’ restrictions, enforcing stricter rules for social distancing and decreased human-to-human contact that stopped short of nationwide measures. On 5 January 2021 the whole of England was placed in national lockdown. Figure 6 shows overall incidence and R(t) for the longer period from 1 October 2020 to 16 January 2021 in the three regions with the largest proportion of B.1.1.7. The proportion of B.1.1.7 in these regions in January is at least 80%, assuming the proportion has not decreased from the end of December. The combination of Tier 4 and national lockdown measures were able to bring R(t) to \( \sim 0.8 \) in all three of these regions.
Figure 5. Incidence and R(t) for the old and new variants, along with the ratio between these R values, for the three regions in England with the largest proportion of B.1.1.7. Dark grey regions indicate national lockdowns, light grey the period where London and much of the South East and East of England were placed in Tier 4 restrictions.

Figure 6. Total incidence and R(t) for the three regions with the highest proportion of B.1.1.7 in December, extended to capture the third national lockdown beginning 5 January 2021. Dark grey regions indicate national lockdowns, light grey indicate the period where London and much of the South East and East of England were placed in Tier 4 restrictions.

Discussion

Using data collected through community reporting of symptoms and tests via the COVID Symptom Study app, we investigated whether the appearance of the variant B.1.1.7, first detected in a sample from England in September 2020, was related to differences in symptom reporting, disease duration, hospitalisation, asymptomatic infection, risks of reinfection, and transmissibility for users reporting a positive test result between 28 September and 27 December 2020.

We did not find associations between the proportion of B.1.1.7 in circulation and disease severity, either measured by the number of different reported symptoms over the 4-week window around each positive test, hospitalisations, any of the different symptoms, or the proportion of individuals with long symptom duration when correcting for variations in demographic characteristics (age, sex) and seasonal variables (temperature, humidity). The proportion of individuals with duration of symptoms ≥28 days without a break of more than seven days did not change in association with the presence of variant B.1.1.7. Likewise the proportion of users with asymptomatic disease did not significantly change as B.1.1.7 increased in prevalence.
A recent study reported that individuals infected with B.1.1.7 were more likely to report a cough, sore throat, fatigue, myalgia and fever in the seven days preceding the test, and less likely to report a loss of taste or smell.\(^{13}\) It is not clear if this report adjusted for age, sex, and environmental factors. If we do not correct for these factors we find some significant changes in symptom reporting but in our view these are not likely to be due to B.1.1.7 (Supplementary Figure 4). The periods considered also differed; we considered symptoms reported both two weeks before and after the positive test result. Further opportunity to study symptoms with B.1.1.7 in different contexts are required to be definitive.

We observed, based on 249 potential cases, a very low rate of possible reinfection of 0.7\% (95\% CI 0.6- 0.8). This rate is consistent with another study of 6614 healthcare workers that had previously tested positive for Covid-19, finding 44 possible reinfections (0.66\%).\(^{14}\) Our reinfection rate did not vary consistently across regions or time, which would be consistent with the hypothesis that reinfection is no more likely in the context of B.1.1.7. This may mean that if adequate immunity is built over the first infection it may be sufficient to protect against reinfection in the presence of B.1.1.7. Ultimately this is a positive sign that the immunity built through vaccination against the old variants could also be useful against B.1.1.7. This is in line with initial reports regarding the efficacy of vaccines designed for early strains against this newer variant.\(^{15,16}\)

We found an increase in the reproduction number R(t) in association with the B.1.1.7 variant: we found a multiplicative increase in R(t) of \(\sim 1.35\) (95\% CI 1.02-1.69), compatible with estimates from Volz et al.\(^{3,4}\), of 1.4-1.8, and Davies et al. who estimated a transmissibility increase of 1.56 (95\% CI 1.50-1.74).\(^{3,4}\) These increases in transmissivity have worrying implications for the ability of lockdown measures to control B.1.1.7, given R(t) was estimated to be 0.7-0.9 during the first national lockdown in England.\(^{17}\) Despite this, we found R(t) to be \(\sim 0.8\) in the three regions in England with at least 80\% of B.1.1.7, with very clear response to lockdown measures. This could indicate that the true increase in transmissivity is at the lower end of the available estimates, or that the increase in transmissivity estimated outside of lockdown cannot be extrapolated to lockdown, perhaps due to B.1.1.7 responding differently to lockdown measures than the old variants.

**Strengths**

The large, longitudinal nature of the CSS data, with good coverage of the UK population, provides a unique opportunity to study potential changes in symptomatology, symptom severity, and disease duration. The ability to match tests and symptom reports over long periods further allows us to measure possible reinfection rates. Our data also offers the ability to provide a valuable complementary measure to existing measurements of the increased transmissibility of B.1.1.7: we were able to use real-time, representative incidence estimates to measure R(t), whilst other studies have relied on deaths and hospitalisations, which are lagged, or community case numbers which do not reflect true infection numbers.

**Limitations**
Our study is limited by reliance on self-report of symptoms and test results, although previous publications from our group show that our figures triangulate well with other study designs. Despite the ability of the app users to correct any wrong input of their test results, errors still may be made. We make the assumption that testing positive for SARS CoV2 after an interval of 90 days with at least seven days of freedom from symptoms in the interval is consistent with reinfection. Repeated positive testing has been reported shortly after hospital discharge and showed that PCR positivity could be detected up to 28 days post symptom resolution. While the chosen cut-off of 90 days between two positive tests is unlikely to be due to prolonged PCR positivity, this cannot be ruled out, but would only affect a small number of cases. Viral sequencing of the two infections would ideally be required to confirm reinfection. Despite correcting for changes in temperature and humidity, a possible limitation in the study is that comparisons in symptoms are made across time, and seasonal effects (e.g. on symptoms) may not have been fully taken into account. As we lack information on the disease strain of individual positive infections reported through the app, the study is associative in nature; and we cannot account for the effects of other potentially circulating variants.

Conclusions

We examined the effect of SARS-CoV-2 variant B.1.1.7 on the symptoms, disease course, rates of reinfection, and transmissibility in the UK. We found no change in symptoms and no increase in overall disease severity. We found a low rate of reinfection (0.7%) and no evidence of increased rates associated with B.1.1.7. We found an increase in R(t) of ~ 1.38 (95% CI 1.06-1.71), but evidence that lockdown measures are effective even in regions with very high (>80%) proportions of B.1.1.7.

Ethics

Ethics has been approved by KCL Ethics Committee REMAS ID 18210, review reference LRS-19/20-18210 and all participants provided consent.

Data sharing

Data collected in the COVID Symptom Study smartphone application are being shared with other health researchers through the UK National Health Service-funded Health Data Research UK (HDRUK) and Secure Anonymised Information Linkage consortium, housed in the UK Secure Research Platform (Swansea, UK). Anonymised data are available to be shared with researchers according to their protocols in the public interest (https://web.www.healthdatagateway.org/dataset/fddcb382-3051-4394-8436-b92295f14259). US investigators are encouraged to coordinate data requests through the Coronavirus Pandemic Epidemiology Consortium (https://www.monganinstitute.org/cope-consortium).

Acknowledgements

ZOE Global provided in kind support for all aspects of building, running and supporting the app and service to all users worldwide. 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. Support for this study was provided by the NIHR-funded Biomedical Research Centre based at GSTT NHS Foundation Trust. Investigators also received support from the Wellcome Trust, the MRC/BHF, Alzheimer’s Society, EU, NIHR, CDRF, and the NIHR-funded BioResource, Clinical Research Facility and BRC based at GSTT NHS Foundation Trust in partnership with KCL, the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare, the Wellcome Flagship Programme (WT213038/Z/18/Z), the Chronic Disease Research Foundation, and DHSC. ATC was supported in this work through a Stuart and Suzanne Steele MGH Research Scholar Award. The Massachusetts Consortium on Pathogen Readiness (MassCPR) and Mark and Lisa Schwartz supported MGH investigators (DAD, LHN, ATC).

Declaration of interests

AM, LP, SS, JCP, CH, JW are employees of Zoe Global Ltd. TDS is a consultant to Zoe Global Ltd. DAD and ATC previously served as investigators on a clinical trial of diet and lifestyle using a separate smartphone application that was supported by Zoe Global.

References


5. Peter Horby, Catherine Huntley, Nick Davies, John Edmunds, Neil Ferguson, Graham Medley, Andrew Hayward, Muge Cevik, Calum Semple. NERVTAG note on B.1.1.7 severity. NERVTAG, 2021.


16 Wu K, Werner AP, Moliva JI, et al. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. DOI:10.1101/2021.01.25.427948.


Supplementary material

Supplementary tables and figures
Supplementary Figure 1. Regional plots of the frequency of reporting of symptoms over time for each reported symptom, against the proportion of B.1.1.7. Drop in fever reporting in early November was caused by a change in the question wording; this wording was subsequently reverted a week later.
Supplementary Figure 2. Regional plot of hospitalisation report, proportion of asymptomatic, report of long duration and number of experienced symptoms around test against proportion of B.1.1.7. For the study of long duration, tests are only considered up to 21 December counting reports up to 18 January 2021 to limit right censoring. Only symptomatic individuals for which duration can be ascertained are included.
Supplementary Figure 3. Incidence and R(t) for the old and new variants, along with the ratio between these R values, for the three regions in England with the largest proportion of B.1.1.7, using SGTF data. Dark grey regions indicate national lockdowns, light grey shaded the period where London and much of the South East and East of England were placed in Tier 4 restrictions.

Supplementary Figure 4: Colour plot of beta values and associated p-values for each region and symptoms when investigating association between symptom report (in a 4 week window around the test) and proportion of variant B.1.1.7 and without any correction for personal characteristic or seasonal feature. Note that the p-values are capped at 0.1. Beta-values are presented for an increase of 0.1 in the proportion of variant B.1.1.7.

<table>
<thead>
<tr>
<th>Region</th>
<th>Proportion of fully asymptomatic</th>
<th>Number of symptoms reported over 4 weeks around test</th>
<th>Proportion of hospital reports</th>
<th>Proportion of individuals with symptom duration &gt;= 28 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>South East</td>
<td>0.001 [-0.015;0.017] ; 0.901</td>
<td>-0.021 [-0.163;0.121] ; 0.733</td>
<td>-0.002 [-0.011;0.007] ; 0.624</td>
<td>-0.003 [-0.009;0.004] ; 0.37</td>
</tr>
<tr>
<td>East of England</td>
<td>0.002 [-0.008;0.012] ; 0.588</td>
<td>-0.012 [-0.153;0.13] ; 0.851</td>
<td>-0.002 [-0.01;0.006] ; 0.52</td>
<td>-0.002 [-0.015;0.01] ; 0.689</td>
</tr>
<tr>
<td>London</td>
<td>-0.005 [-0.014;0.005] ; 0.298</td>
<td>0.031 [-0.055;0.116] ; 0.423</td>
<td>-0.002 [-0.007;0.003] ; 0.298</td>
<td>-0.002 [-0.013;0.009] ; 0.682</td>
</tr>
<tr>
<td>Midlands</td>
<td>-0.016 [-0.028;0.004] ; 0.14</td>
<td>0.02 [-0.133;0.173] ; 0.766</td>
<td>-0.002 [-0.007;0.003] ; 0.328</td>
<td>0.002 [-0.01;0.015] ; 0.671</td>
</tr>
<tr>
<td>North East and Yorkshire</td>
<td>-0.011 [-0.046;0.023] ; 0.462</td>
<td>-0.086 [-0.444;0.272] ; 0.586</td>
<td>-0.011 [-0.04;0.019] ; 0.426</td>
<td>0.015 [-0.022;0.052] ; 0.349</td>
</tr>
<tr>
<td>North West</td>
<td>-0.005 [-0.023;0.013] ; 0.512</td>
<td>-0.053 [-0.218;0.111] ; 0.468</td>
<td>-0.009 [-0.015;0.004] ; 0.005</td>
<td>0 [-0.031;0.031] ; 0.98</td>
</tr>
<tr>
<td>South West</td>
<td>0.015 [-0.011;0.04] ; 0.217</td>
<td>-0.261 [-0.437;0.085] ; 0.01</td>
<td>-0.001 [-0.02;0.018] ; 0.902</td>
<td>-0.048 [-0.091;0.004] ; 0.036</td>
</tr>
<tr>
<td>Scotland</td>
<td>0.022 [-0.013;0.058] ; 0.177</td>
<td>-0.4 [-0.711;-0.088] ; 0.019</td>
<td>-0.018 [-0.037;0.002] ; 0.073</td>
<td>-0.012 [-0.027;0.003] ; 0.107</td>
</tr>
<tr>
<td>Wales</td>
<td>-0.002 [-0.05;0.047] ; 0.943</td>
<td>-0.041 [-0.683;0.601] ; 0.884</td>
<td>-0.008 [-0.045;0.028] ; 0.602</td>
<td>-0.053 [-0.141;0.035] ; 0.192</td>
</tr>
</tbody>
</table>
**Supplementary Table 1**: Beta coefficient of the variant proportion when evaluating association with number of reported symptoms, asymptomatic rate, proportion of hospital report and proportion of individuals with duration >28 days (among symptomatic) across the different regions when correcting for age, sex, temperature and humidity. All values are presented for an increase in 0.1 in the proportion of variant B.1.1.7. All results are presented in the form mean [CI]; p-value

<table>
<thead>
<tr>
<th>Region</th>
<th>Correlation Variant/Reinfection</th>
<th>Correlation New cases/Reinfection</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>South East</td>
<td>0.55</td>
<td>0.69</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>East of England</td>
<td>0.51</td>
<td>0.56</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>London</td>
<td>0.46</td>
<td>0.62</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Midlands</td>
<td>0.28</td>
<td>0.75</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>North East and Yorkshire</td>
<td>-0.02</td>
<td>0.30</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>North West</td>
<td>0.06</td>
<td>0.43</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>South West</td>
<td>-0.35</td>
<td>0.05</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Scotland</td>
<td>0.59</td>
<td>-0.15</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Wales</td>
<td>0.07</td>
<td>0.26</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Supplementary Table 2**: Comparison of regional correlation over time between proportion of B.1.1.7 and number of possible reinfections and between new reported cases and number of possible reinfections. Medians over 100 bootstrapped samples are calculated for each and compared using a Mann-Whitney U test.
COG-UK authorship list

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Samuel C Robson 13.

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Nicholas J Loman 41 and Thomas R Connor 10, 69.

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Tanya Golubchik 5.

Funding acquisition, Metadata curation, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Rocio T Martinez Nunez 42.

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:
Catherine Ludden 88.

Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:
Sally Corden 69.

Funding acquisition, Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:
Ian Johnston 99 and David Bonsall 5.

Funding acquisition, Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:

Funding acquisition, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Giselda Bucca 87.

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
M. Estee Torok 22, 101.

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Visualisation:
Kordo Saeed 81, 110 and Jacqui A Prieto 83, 109.
Leadership and supervision, Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:
David K Jackson

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
William L Hamilton

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Visualisation:
Luke B Snell

Funding acquisition, Leadership and supervision, Metadata curation, and Samples and logistics:
Catherine Moore

Funding acquisition, Leadership and supervision, Project administration, and Samples and logistics:
Ewan M Harrison

Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:
Sonia Goncalves

Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:
Ian G Goodfellow, Derek J Fairley, Matthew W Loose and Joanne Watkins

Leadership and supervision, Metadata curation, Samples and logistics, and Software and analysis tools:
Rich Livett

Leadership and supervision, Metadata curation, Samples and logistics, and Visualisation:
Samuel Moses

Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:
Roberto Amato, Sam Nicholls and Matthew Bull

Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:
Darren L Smith

Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:
Jeff Barrett, David M Aanensen

Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
Martin D Curran, Surendra Parmar, Dinesh Aggarwal, and James G Shepherd.
Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:
Matthew D Parker 93.

Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:
Sharon Glaysher 61.

Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:
Matthew Bashton 37, 58, Anthony P Underwood 14, 114, Nicole Pacchiarini 69 and Katie F Loveson 77.

Project administration, Sequencing and analysis, Software and analysis tools, and Visualisation:
Alessandro M Carabelli 88.

Funding acquisition, Leadership and supervision, and Metadata curation:
Kate E Templeton 53, 90.

Funding acquisition, Leadership and supervision, and Project administration:
Cordelia F Langford 99, John Sillitoe 99, Thushan I de Silva 93 and Dennis Wang 93.

Funding acquisition, Leadership and supervision, and Sequencing and analysis:
Dominic Kwiatkowski 96, 107, Andrew Rambaut 90, Justin O’Grady 70, 89 and Simon Cottrell 69.

Leadership and supervision, Metadata curation, and Sequencing and analysis:

Leadership and supervision, Project administration, and Samples and logistics:
Husam Osman 64, 36, Monique Andersson 59, Anoop J Chauhan 61 and Mohammed O Hassan-Ibrahim 6.

Leadership and supervision, Project administration, and Sequencing and analysis:
Mara Lawniczak 99.

Leadership and supervision, Samples and logistics, and Sequencing and analysis:

Leadership and supervision, Sequencing and analysis, and Software and analysis tools:

Leadership and supervision, Sequencing and analysis, and Visualisation:
Andrew R Bassett 99.

Metadata curation, Project administration, and Samples and logistics:

Metadata curation, Project administration, and Sequencing and analysis:
Martin P McHugh 93 and Rebecca Dewar 93.
Metadata curation, Samples and logistics, and Sequencing and analysis:
Aminu S Jahun 24, Claire McMurray 41, Sarojini Pandey 84, James P McKenna 3, Andrew Nelson 58, 105, Gregory R Young 37, 58, Clare M McCann 58, 105 and Scott Elliott 61.

Metadata curation, Samples and logistics, and Visualisation:
Hannah Lowe 25.

Metadata curation, Sequencing and analysis, and Software and analysis tools:
Ben Temperton 91, Sunando Roy 82, Anna Price 10, Sara Rey 69 and Matthew Wyles 93.

Metadata curation, Sequencing and analysis, and Visualisation:
Stefan Rooke 90 and Sharif Shaaban 68.

Project administration, Samples and logistics, Sequencing and analysis:
Mariateresa de Cesare 98.

Project administration, Samples and logistics, and Software and analysis tools:
Laura Letchford 99.

Project administration, Samples and logistics, and Visualisation:
Siona Silveira 81, Emanuela Pelosi 81 and Eleri Wilson-Davies 81.

Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Myra Hosmillo 24.

Sequencing and analysis, Software and analysis tools, and Visualisation:

Funding acquisition, and Leadership and supervision:

Funding acquisition, and Project administration:
David Buck 98 and Michaela John 9.

Leadership and supervision, and Metadata curation:
Gemma L Kay 70.

Leadership and supervision, and Project administration:
Steve Palmer 99, Sharon J Peacock 88, 64 and David Heyburn 69.

Leadership and supervision, and Samples and logistics:
Danni Weldon 99, Esther Robinson 64, 36, Alan McNally 41, 86, Peter Muir 64, Ian B Vipond 64, John BoYes 29, Venkat Sivaprakasam 46, Tranprit Salluja 75, Samir Dervisevic 54 and Emma J Meader 54.

Leadership and supervision, and Sequencing and analysis:

**Leadership and supervision, and Visualisation:**
Jane AH Masoli 73, 91.

**Metadata curation, and Samples and logistics:**
Bridget A Knight 73, 91, Christopher R Jones 73, 91, Cherian Koshy 1, Amy Ash 1, Anna Casey 71, Andrew Bosworth 84, 36, Liz Ratcliffe 71, Li Xu-McCrae 36, Hannah M Pymont 64, Stephanie Hutchings 64, Lisa Berry 84, Katie Jones 84, Fenella Halstead 46, Thomas Davis 21, Christopher Holmes 16, Miren Iturriza-Gomara 92, Anita O Lucaci 92, Paul Anthony Randell 38, 104, Alison Cox 38, 104, Pinglawathee Madona 38, 104, Kathryn Ann Harris 30, Julianne Rose Brown 30, Tabitha W Mahungu 74, Dianne Irish-Tavares 74, Tanzina Haque 74, Jennifer Hart 74, Eric Witele 74, Melisa Louise Fenton 75, Steven Liggett 78, Clive Graham 56, Emma Swindells 57, Jennifer Collins 50, Gary Eltringham 50, Sharon Campbell 17, Patrick C McClure 97, Gemma Clark 15, Tim J Sloan 60, Carl Jones 15 and Jessica Lynch 2, 111.

**Metadata curation, and Sequencing and analysis:**
Ben Warne 8, Steven Leonard 99, Jillian Durham 99, Thomas Williams 90, Sam T Haldenby 92, Nathaniel Storey 30, Nabil-Fareed Alikhan 70, Nadine Holmes 18, Christopher Moore 18, Matthew Carlile 18, Malorie Perry 69, Noel Craine 69, Ronan A Lyons 60, Angela H Beckett 13, Salman Goudarzi 77, Christopher Fearn 77, Kate Cook 77, Hannah Dent 77 and Hannah Paul 77.

**Metadata curation, and Software and analysis tools:**
Robert Davies 89.

**Project administration, and Samples and logistics:**

**Project administration, and Sequencing and analysis:**
Nazreen F Hadjirin 88 and Joshua Quick 41.

**Project administration, and Software and analysis tools:**
Radoslaw Poplawski 41.

**Samples and logistics, and Sequencing and analysis:**

**Samples and logistics, and Software and analysis tools:**
Igor Starinskij 48.

Sequencing and analysis, and Software and analysis tools:

Software and analysis tools, and Visualisation:

Leadership and supervision:

Metadata curation:

Project administration:

Samples and logistics:
Sequencing and analysis:

Software and analysis tools:
James Bonfield, Christoph Puethe, Andrew Whitham, Jennifer Liddle, Will Rowe, Igor Siveroni, Thanh Le-Viet and Amy Gaskin.

Visualisation:
Rob Johnson.
Medical Centre, Clinical Microbiology, University Hospitals of Leicester NHS Trust, County Durham and Darlington NHS Foundation Trust, Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, Department of Infection Biology, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, Department of Infectious Diseases, King's College London, Department of Microbiology, Kettering General Hospital, Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust; Cambridge, UK, Department of Zoology, University of Oxford, Division of Virology, Department of Pathology, University of Cambridge, East Kent Hospitals University NHS Foundation Trust, East Suffolk and North Essex NHS Foundation Trust, Gateshead Health NHS Foundation Trust, Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, Gloucestershire Hospitals NHS Foundation Trust, Great Ormond Street Hospital for Children NHS Foundation Trust, Guy's and St. Thomas' BRC, Guy's and St. Thomas' Hospitals, Hampshire Hospitals NHS Foundation Trust, Health Data Research UK Cambridge, Health Services Laboratories, Heartlands Hospital, Birmingham, Hub for Biotechnology in the Built Environment, Northumbria University, Imperial College Hospitals NHS Trust, Imperial College London, Institute of Biodiversity, Animal Health & Comparative Medicine, Institute of Microbiology and Infection, University of Birmingham, King's College London, Liverpool Clinical Laboratories, Maidstone and Tunbridge Wells NHS Trust, Manchester University NHS Foundation Trust, Microbiology Department, Wye Valley NHS Trust, Hereford, MRC Biostatistics Unit, University of Cambridge, MRC-University of Glasgow Centre for Virus Research, National Infection Service, PHE and Leeds Teaching Hospitals Trust, Newcastle Hospitals NHS Foundation Trust, Newbury University, NHS Greater Glasgow and Clyde, NHS Lothian, Norfolk and Norwich University Hospital, Norfolk County Council, North Cumbria Integrated Care NHS Foundation Trust, North Tees and Hartlepool NHS Foundation Trust, Northumbria University, Oxford University Hospitals NHS Foundation Trust, PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, Portsmouth Hospitals University NHS Trust, Princess Alexandra Hospital Microbiology Dept., Public Health Agency, Public Health England, Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK, Public Health England, Colindale, Public Health England, Colindale, Public Health Scotland, Public Health Wales NHS Trust, Quadram Institute Bioscience, Queen Elizabeth Hospital, Queen's University Belfast, Royal Devon and Exeter NHS Foundation Trust, Royal Free NHS Trust, Sandwell and West Birmingham NHS Trust, School of Biological Sciences, University of Portsmouth (PORT), School of Pharmacy and Biomedical Sciences, University of Portsmouth (PORT), Sheffield Teaching Hospitals, South Tees Hospitals NHS Foundation Trust, Swansea University, University Hospitals Southampton NHS Foundation Trust, University of Birmingham Turnkey Laboratory, University of Brighton, University of Cambridge, University of East Anglia, University of Edinburgh, University of Exeter, University of Liverpool, University of Sheffield, University of Warwick, University of Cambridge, Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Wellcome Sanger Institute, West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, Department of Medicine, University of Cambridge, Ministry of Health, Sri Lanka, NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, North West London Pathology, NU-OMICS, Northumbria University, University of Kent, University of Oxford, University of Southampton, University of Southampton School of Health Sciences, University of Surrey, Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCFW NHS Trust, Wellcome Africa Health Research Institute Durban and Wellcome Genome Campus.