PRELIMINARY – NOT PEER REVIEWED

Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England

Nicholas G. Davies[1], Rosanna C. Barnard*, Christopher I. Jarvis*, Adam J. Kucharski*, James Munday*, Carl A. B. Pearson*, Timothy W. Russell*, Damien C. Tully*, Sam Abbott, Amy Gimma, William Waites, Kerry LM Wong, Kevin van Zandvoort, CMMID COVID-19 Working Group, Rosalind M. Eggo, Sebastian Funk, Mark Jit, Katherine E. Atkins, W. John Edmunds

Centre for Mathematical Modelling of Infectious Diseases
London School of Hygiene and Tropical Medicine
Keppel St, London, WC1E 7HT
United Kingdom

[1] Corresponding author
* Equal contribution

A novel SARS-CoV-2 variant, VOC 202012/01, emerged in southeast England in November 2020 and appears to be rapidly spreading towards fixation. We fitted a two-strain mathematical model of SARS-CoV-2 transmission to observed COVID-19 hospital admissions, hospital and ICU bed occupancy, and deaths; SARS-CoV-2 PCR prevalence and seroprevalence; and the relative frequency of VOC 202012/01 in the three most heavily affected NHS England regions (South East, East of England, and London). We estimate that VOC 202012/01 is 56% more transmissible (95% credible interval across three regions 50-74%) than preexisting variants of SARS-CoV-2. We were unable to find clear evidence that VOC 202012/01 results in greater or lesser severity of disease than preexisting variants. Nevertheless, the increase in transmissibility is likely to lead to a large increase in incidence, with COVID-19 hospitalisations and deaths projected to reach higher levels in 2021 than were observed in 2020, even if regional tiered restrictions implemented before 19 December are maintained. Our estimates suggest that control measures of a similar stringency to the national lockdown implemented in England in November 2020 are unlikely to reduce the effective reproduction number R_t to less than 1, unless primary schools, secondary schools, and universities are also closed. We project that large resurgences of the virus are likely to occur following easing of control measures. It may be necessary to greatly accelerate vaccine roll-out to have an appreciable impact in suppressing the resulting disease burden.

In late December 2020, evidence began to emerge that a novel SARS-CoV-2 variant, Variant of Concern 202012/01 (henceforth VOC 202012/01), was prevalent and rapidly outcompeting preexisting variants in at least three regions of England: the South East, London and the East of England (*1*). This variant seems to have increased in incidence during the last period of national lockdown (November 5th - December 2nd 2020), and continued to increase following the lockdown despite many of the affected areas being in the (then) highest level of restrictions ("Tier 3"). Concern over this variant led the UK government to place parts of these three regions under "Tier 4" restrictions on 20th December, a package of control measures broadly equivalent to those enacted during the nationwide lockdown in England in November. Our current understanding of effective pharmaceutical and non-pharmaceutical control of SARS-CoV-2 does not reflect potential epidemiological and clinical characteristics of VOC 202012/01. Early estimates of the transmissibility and disease severity for this novel variant are crucial for informing rapid policy responses to this potential threat.

Details of emergent variant

VOC 202012/01 is defined by 17 mutations (14 non-synonymous mutations and 3 deletions), among which eight are located in the spike protein. At least three mutations have a potential biological significance. Mutation N501Y is one of the key contact residues in the receptor

binding domain and has been shown to enhance binding affinity to human ACE2 (*2*, *3*). The function of mutation P681H is unclear, but it is located immediately adjacent to the furin cleavage site in spike, a known region of importance for infection and transmission (*4*, *5*). The deletion of two amino acids at positions 69-70 in spike has arisen in multiple independent circulating lineages of SARS-CoV-2, is linked to immune escape in immunocompromised patients and enhances viral infectivity in vitro (*6*, *7*). This deletion is also responsible for certain commercial diagnostic assays failing to detect the spike glycoprotein gene (S gene drop-out), with genomic data confirming these S gene target failures are primarily due to the new variant (*1*). Accordingly, molecular evidence is consistent with a potentially altered infectiousness phenotype for this variant.

The proportion of COVID-19 cases caused by VOC 202012/01 is growing rapidly in the South East, East of England and London regions (**Fig. 1A**) and is associated with an increase in the estimated reproduction number R_t (**Fig. 1B**). Social contacts and mobility data suggest that the rise in relative prevalence of VOC 202012/01 within England is unlikely to be caused by founder effects: that is, if certain regions had higher levels of transmission as a result of more social interactions, genetic variants that were more prevalent within these regions could become more common overall. However, we did not find evidence of differences in social interactions between regions of high and low VOC 202012/01 prevalence, as measured by Google mobility (*8*) and social contact survey data (*9*) (**Fig. 1C, D**), despite the fact that changes in measured contact patterns appear to closely correlate with changes in the reproduction number inferred from community infection prevalence (**Fig 1E**) and that regionally-differentiated mobility data have previously informed accurate predictions for COVID-19 dynamics in England (*10*). This apparent decoupling between social contacts and transmission during November and December could therefore indicate a change in characteristics of the virus.

To evaluate this possibility, we extended an age- and regionally-structured mathematical model of SARS-CoV-2 transmission (*10*, *11*) to consider two co-circulating variants of SARS-CoV-2 (**Fig. S1**). The model is fitted across 7 NHS England regions to observed hospital admissions, hospital and ICU bed occupancy, deaths within 28 days of a positive SARS-CoV-2 test, PCR prevalence, seroprevalence, and, specifically for the East of England, London and South East NHS England regions, to the relative frequency of VOC 202012/01 identified by genomic sequencing surveillance from the COVID-19 Genomics UK Consortium (*12*). We use this model to capture the emergence of VOC 202012/01 and consider its potential impact.

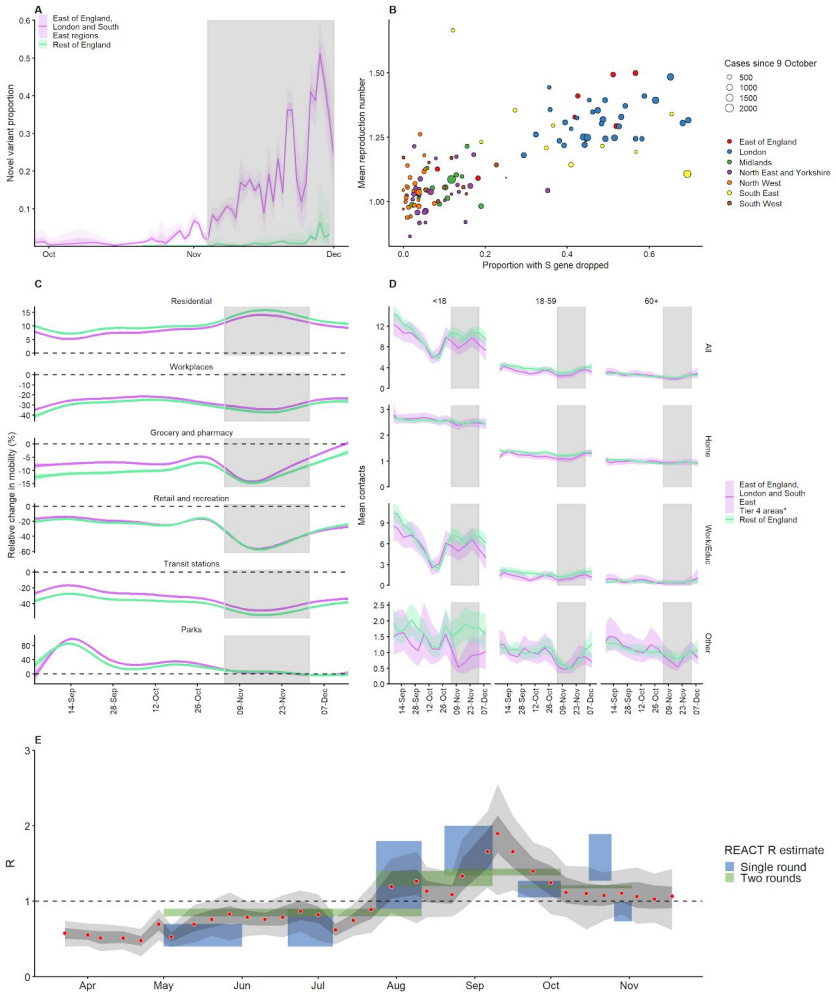

Fig. 1. (A) Proportion of VOC 202012/01 in South East, East of England, and London NHS England regions versus the rest of England from 28 September – 1 December 2020 (mean and 95% CI). Grey shaded areas (panels A, C, D) reflect the period of time when England was in a second national lockdown. We cut off the data after 1 December 2020 due to a substantial decrease in representativeness after this time (**Fig. S4**). **(B)** Proportion of S gene drop-outs (5 – 11 December) versus mean reproduction number (27 November – 4 December) by local authority in England. The one-week lag accounts for delays from infection to test. **(C)** Percentage change (95% CI) in Google Mobility indices relative to baseline over time and **(D)** setting-specific mean contacts (95% CI) from the CoMix study (*9*) over time and by age for local authorities that went into Tier 4 compared to the rest of England. Educ = education setting. *Some local authorities that were within the South East, East of England, and London NHS England regions did not go into Tier 4 and were therefore included in the rest of England for panels C and D. **(E)** Estimates of R_0 (50% and 95% CI) from CoMix social contact survey (*9*) compared to R_t estimates from REACT-1 SARS-CoV-2 prevalence survey (*13*) for England. R_t estimates based on single and aggregated REACT-1 survey rounds are shown.

Characterising the transmissibility and severity of VOC 202012/01

To understand possible biological mechanisms for the observed dynamics associated with VOC 202012/01, we considered four alternative hypotheses for why the new variant might be spreading more efficiently: increased infectiousness; immune escape; increased susceptibility among children; and shorter generation time.

First, we modelled increased infectiousness as an increase in the risk of transmission of VOC 202012/01 per contact, relative to preexisting variants. This model was best able to capture the data parsimoniously among the four hypotheses tested (Deviance Information Criterion DIC = 9395, ΔDIC = 0, **Fig 2A**). Such a mechanism is consistent, in principle, with observations of lower Ct values (i.e., higher viral load) for VOC 202012/01 (*14*).

Second, we modelled immune escape by assuming individuals previously infected with preexisting variants had a degree of susceptibility to reinfection by VOC 202012/01. Such a mechanism is consistent with the ΔH69/ΔV70 deletion contributing to immune escape in an immunocompromised patient (*6*) and could have implications for vaccine effectiveness. However, this model performed substantially worse (DIC = 10217, ΔDIC = 822, **Fig 2B**) and underestimated the observed relative growth rate of VOC 202012/01 even when assuming no cross-protection.

Third, we modelled increased susceptibility among children (aged 0-19) by assuming they could be as susceptible as adults (aged 20+) to infection by VOC 202012/01 given exposure. Evidence suggests children are typically less susceptible to SARS-CoV-2 infection than adults (*15*, *16*), possibly due to immune cross-protection resulting from infection by other human coronaviruses (*17*). Our baseline model assumes that 0–19-year-olds are approximately half as susceptible to SARS-CoV-2 infection as 20+-year-olds (*15*); if this were to change for the new variant, it could have implications for the effectiveness of school closures as a control measure. This model was the least parsimonious model we evaluated (DIC = 11718, ΔDIC = 2323, **Fig 2C**), and underestimated the observed relative growth rate of COV 202012/01.

Finally, we modelled a shorter generation time by assuming individuals could become infectious more quickly after exposure, but maintained the same infectious period. A shorter generation time results in a higher growth rate when $R_t > 1$, and would have implications for the effectiveness of control measures against this variant, because holding the growth rate of an epidemic constant, a shorter generation time implies a lower reproduction number and hence $R_t < 1$ is easier to achieve. This final model performed poorly (DIC = 10478, ΔDIC = 1083, **Fig. 2D**), largely because it predicted that VOC 202012/01 should have decreased in relative

frequency during the lockdown in November 2020. When $R_t < 1$ for both variants, a shorter generation time is a selective disadvantage, because infections with this variant decline faster compared to a variant with the same R_t but transmitting on a longer timescale.

The four models evaluated here to explain infection resurgence generate further testable hypotheses. For example, an increase in susceptibility among children will, all else being equal, generate a marked increase in cases in children, and reductions across young and middle-aged adults (**Fig. S2**). Limited cross-protection between the variants would entail a high reinfection rate, while a shorter generation time could be corroborated with epidemiologic investigation. Additional data, when available, could therefore help verify our early findings as well as detect the possibility of combinations of multiple mechanisms that were not independently identifiable in the inference process given the data available at the time of writing. There could be other plausible alternatives for the biological mechanism underlying an increased growth rate which we have not evaluated here. For example, the duration of infectiousness could be increased for VOC 202012/01, which would likely produce similar results to our first model (increased infectiousness).

The fitted model based upon increased transmissibility, which reproduces observed epidemiological dynamics and increases in relative prevalence of VOC 202012/01 (**Figs. 3, S3**), suggests no clear evidence of a difference in odds of hospitalisation or relative risk of death, but finds strong evidence of higher relative transmissibility (**Fig. 2A**), estimated at 56% higher than preexisting variants (95% CrI across three regions: 50-74%). This estimate is consistent with a previous estimate of 70% increased transmissibility for VOC 202012/01 (*14*). By contrast, a model without these differences in transmissibility between VOC 202012/01 and preexisting variants was unable to reproduce observed patterns in the data, particularly for December 2020 (**Fig. 3**). This further highlights that changing contact patterns cannot explain the growth of VOC 202012/01.

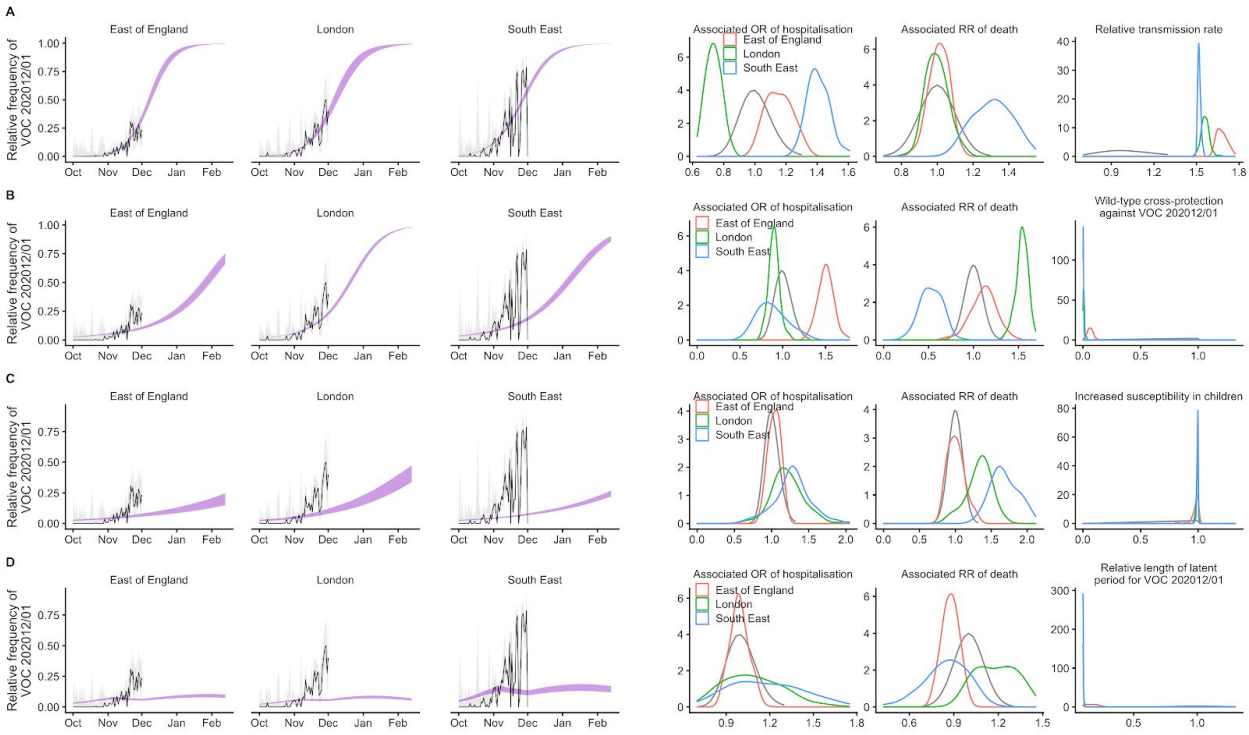

Fig. 2. Comparison of possible biological mechanisms underlying the rapid spread of VOC 202012/01. Each row shows a different assumed mechanism: **(A)** increased transmissibility; **(B)** immune escape; **(C)** increased susceptibility among children; **(D)** shorter generation time. The first three panels in each row show the relative frequency of VOC 202012/01 (black line shows data with 95% binomial confidence interval; purple ribbon shows model fit). The last three panels in each row show posterior estimates for odds of hospitalisation, relative risk of death, and the parameter that defines the hypothesised mechanism (grey line shows prior distributions for each parameter).

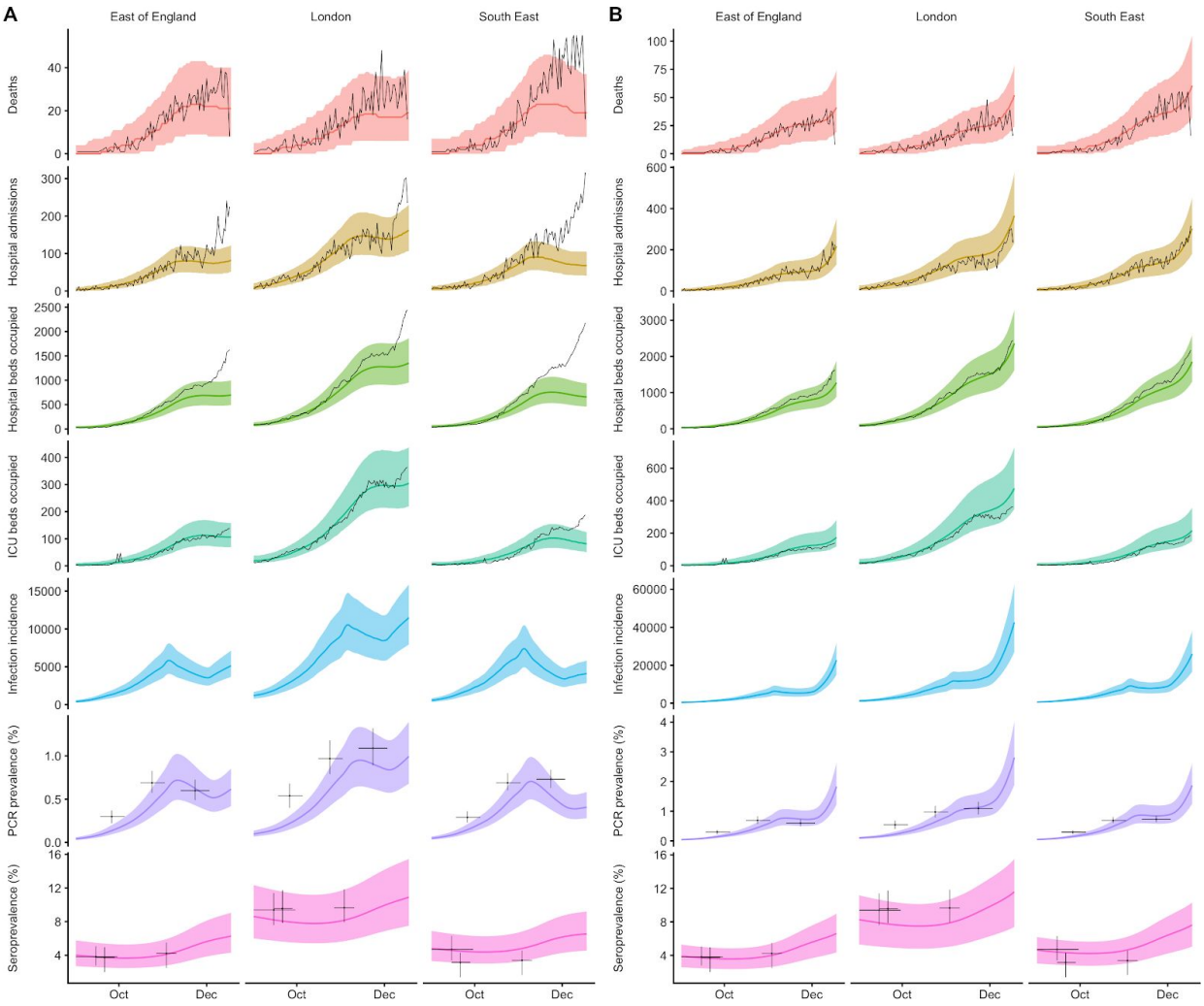

Fig 3. Comparison of fitted model and data across East of England, London, and South East NHS England regions. Black lines show observed data, while coloured lines and shaded regions show median and 95% credible intervals from the fitted model. Note that the deaths data are lagged due to delays from death to notification, so sharp drops in the number of deaths for the most recent 2-4 days are likely to be underestimates. **(A)** Model fit without introduction of VOC 202012/01 variant. **(B)** Model fit with introduction of VOC 202012/01 variant. Full model fit for all regions is shown in **Fig. S3**.

Projections of future dynamics

Using the best-fitting transmission model (increased transmissibility), we compared epidemic dynamics under different assumptions about control measures from mid-December 2020 to the end of June 2021. We compared three main scenarios for non-pharmaceutical interventions: (i) a counterfactual scenario with Tiers 1–3 only, i.e. without additional Tier 4 restrictions that were first introduced on 20 December 2020; (ii) Tier 4 introduced from 20 December 2020 in East of England, London, and the South East, with Tier 4 restrictions introduced from 26 December 2020 in all other regions of England, lasting until 31 January 2021 and with schools and universities opening from 4 January 2021; (iii) scenario ii, but with schools and universities remaining closed until 31 January 2021. We also examined two vaccination scenarios: (iv) 200,000 vaccinations per week and (v) 2,000,000 vaccinations per week. Both vaccination scenarios occurred against a backdrop of non-pharmaceutical interventions as in scenario iii. We assumed that vaccine rollout started on 1 January 2021 and that the vaccine exhibited 95% efficacy against disease and 60% efficacy against infection (**Fig. S1**). For simplicity of modelling, we assumed that vaccine protection was conferred immediately upon receipt of one vaccine dose. Because VOC 202012/01 has now been detected in all parts of England (*14*), for the purposes of these projections we assume that community transmission of VOC 202012/01 begins 30 days later in the rest of England than in the East of England, London, and the South East.

We found that regardless of control measures simulated, all NHS regions are projected to experience a subsequent wave of COVID-19 cases and deaths, peaking in spring 2021 for London, South East and East of England, and in summer 2021 for the rest of England (**Fig. 4**). In the absence of substantial vaccine roll-out, cases, hospitalisations, ICU admissions and deaths in 2021 may exceed those in 2020 (**Table 1**). School closures in January 2021 may delay the peak (**Fig. 4**) and decrease the total burden in the short term. However, implementation of more stringent measures now with a subsequent lifting of these restrictions in February 2021 leads to a bigger rebound in cases, particularly in those regions that have been least affected so far (**Fig. 4** and **Table 1**). However, these delaying measures may buy time to reach more widespread population immunity through vaccination. Vaccine roll-out will further mitigate transmission, although the impact of vaccinating 200,000 people per week—similar in magnitude to the rates reached in December 2020—may be relatively small (**Fig. 5**). An accelerated uptake of 2 million people vaccinated per week is predicted to have a much more substantial impact. The most stringent intervention scenario with Tier 4 England-wide and schools closed during January, and 2 million individuals vaccinated per week, is the only scenario we considered which reduces peak ICU burden below the levels seen during the first wave (**Table 1**).

Table 1. Summary of projections for England, 15 Dec 2020 – 30 June 2020.

indicator	Tiers 1-3 only	Tier 4, schools open	Tier 4, schools closed	Plus 200k immunised per week	Plus 2M immunised per week
Peak ICU (rel. to 1st wave)*	162% (146 - 181%)	113% (90 - 151%)	114% (85 - 158%)	104% (82 - 148%)	84% (75 - 91%)
Peak ICU requirement*	4,750 (4,300 - 5,320)	3,310 (2,660 - 4,450)	3,360 (2,510 - 4,650)	3,060 (2,430 - 4,340)	2,460 (2,200 - 2,660)
Peak deaths*	942 (831 - 1,090)	784 (622 - 1,040)	822 (655 - 1,060)	629 (461 - 835)	413 (383 - 450)
Total admissions	426,000 (379,000 - 471,000)	394,000 (340,000 - 449,000)	375,000 (319,000 - 432,000)	335,000 (288,000 - 387,000)	147,000 (130,000 - 172,000)
Total deaths	118,000 (110,000 - 126,000)	107,000 (98,300 - 118,000)	102,000 (90,800 - 115,000)	83,300 (73,800 - 93,800)	35,700 (31,800 - 40,400)
Weeks in Tier 2	3 (1.83 - 3.96)	3.04 (2.16 - 4.06)	4.12 (3.15 - 5)	4.48 (3.45 - 5.59)	3.72 (2.55 - 5.52)
Weeks in Tier 3	10.8 (10.2 - 12)	9.1 (7.89 - 10)	8.6 (7.61 - 9.87)	8.17 (6.88 - 9.39)	2.76 (1.76 - 4.18)
Weeks in Tier 4	0 (0 - 0)	5.52 (5.52 - 5.52)	5.52 (5.52 - 5.52)	5.52 (5.52 - 5.52)	5.52 (5.52 - 5.52)

* Peak ICU requirement and peak deaths are higher under the “Tier 4, schools closed” than under the “Tier 4, schools open” scenario because closing schools shifts the peak later in East of England, South East, and London NHS regions, so that it coincides with the projected peak in other NHS regions.

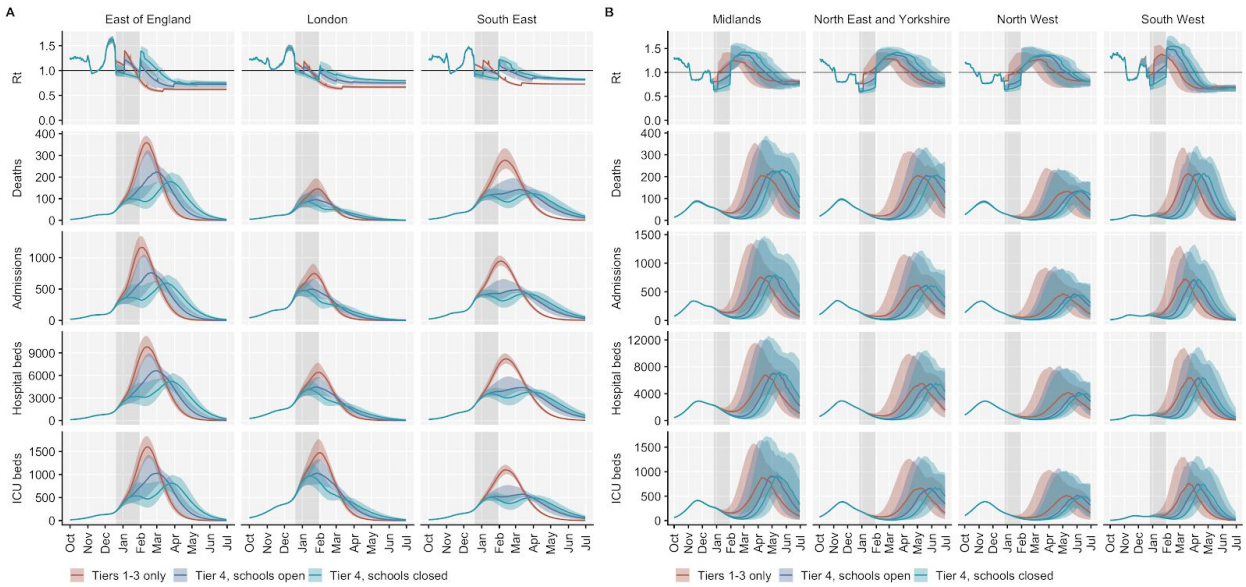

Fig. 4. Projections of epidemic dynamics under different control measures, in the absence of widespread vaccination. Results shown for: **(A)** NHS England regions with data suggesting extensive spread of VOC 202012/01 **(B)** regions without data suggesting extensive spread of VOC 202012/01. Grey shaded areas correspond to the time each region spends under Tier 4 control measures. Note the wider uncertainty in projections for those regions that have, as yet, had few cases of VOC 202012/01.

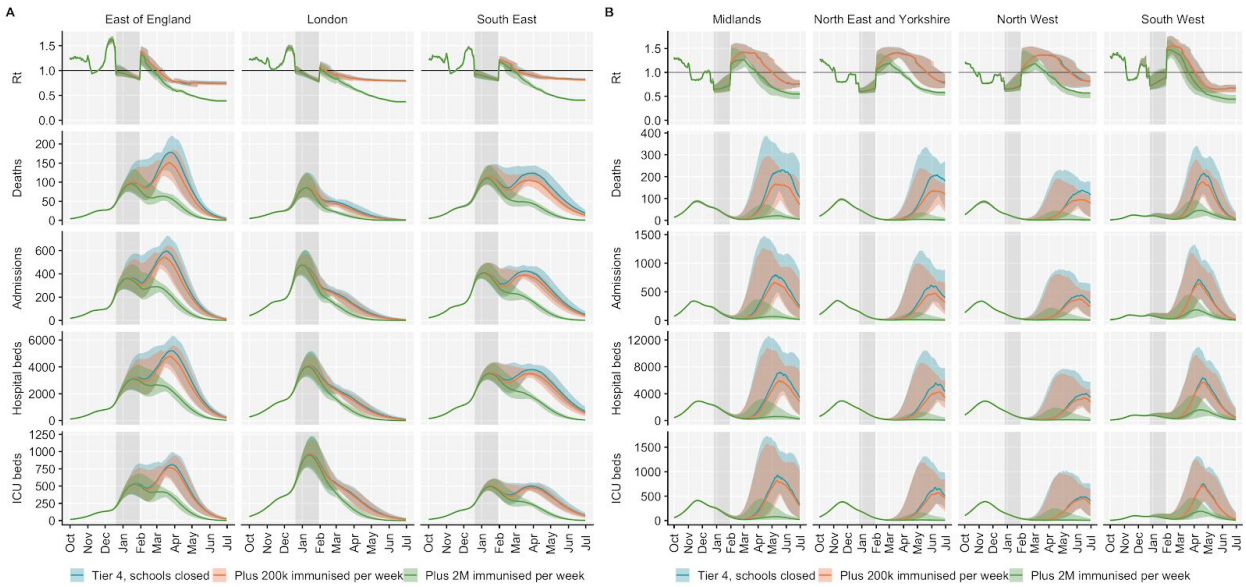

Fig. 5. Projections of epidemic dynamics under different vaccination scenarios. We compare a baseline scenario without vaccination and with Tier 4 restrictions including school closures across all regions of England until 31 January 2021 (as in **Fig. 4**) to two alternative vaccination scenarios: a roll-out of 200,000 people vaccinated per week, and a roll-out of 2 million people vaccinated per week. We assume that vaccination confers 95% vaccine efficacy against disease and 60% vaccine efficacy against infection, and that vaccination starts on 1 January 2021 with vaccine protection starting immediately upon receipt. This is intended to approximate the fact that vaccination started in early December, but that full protection occurs after a time lag and potentially after a second dose. Vaccines are given first to 70+ year olds until 85% coverage is reached in this age group, then to 60+ year olds until 85% coverage is reached in this age group, continuing into younger age groups in 10-year decrements. **(A)** Regions with data suggesting extensive spread of VOC 202012/01. **(B)** Regions without data suggesting extensive spread of VOC 202012/01. Grey shaded areas correspond to the time each region spends under Tier 4 restrictions.

Discussion

Combining multiple behavioural and epidemiological data sources with mathematical models, we estimated that the novel SARS-CoV-2 variant VOC 202012/01 is more transmissible than existing circulating SARS-CoV-2 viruses. As a result of this increased transmissibility, existing control measures are likely to be less effective, and countries may require stronger proactive interventions to achieve the same level of control. We found no evidence that the new variant is associated with higher disease severity, but without strengthened controls, there is a clear risk that future epidemic waves may be larger – and hence associated with greater burden – than previous waves.

Given the extensive Tier 4 measures introduced in London, East of England and South East England, there are limited options available to further reduce transmission in these regions. Educational settings are among the largest institutions linked to SARS-CoV-2 clusters that have remained open during November and December 2020 (*18*), which means school or university closures may be required to prevent a large epidemic in these affected regions in early 2021. We note that even Tier 4 measures together with closure of educational facilities are less stringent than the measures imposed in March 2020, and therefore it is possible that restrictions beyond Tier 4 may be required. If children are more susceptible to VOC 202012/01 than to preexisting variants, we expect that the impact of school closures would be larger; further work is urgently required to more completely characterise epidemiological differences between VOC 202012/01 and preexisting variants. Elsewhere in the UK, similarly strong measures are also likely to be required if the local prevalence of VOC 202012/01, and hence transmissibility, increases. We observe slight increases in hospital admissions in December 2020 in relation to the fitted single-variant model in regions where the spread of VOC 202012/01 is poorly characterised (e.g. Midlands, **Fig. S3**), which could be an indication of changes in underlying transmission. It remains to be seen whether this increase is due to VOC 202012/01.

The rise in transmission from VOC 202012/01 has several potential implications for vaccination. First, it means prompt and efficient vaccine delivery and distribution is even more important to reduce the impact of the epidemic in the near future. Moreover, increased transmission resulting from VOC 202012/01 will raise the herd immunity threshold, meaning the potential future burden of SARS-CoV-2 is larger and it will require higher vaccination coverage to achieve herd immunity. Given the relatively high rate of travel between the UK and other countries, and the high sequencing capacity in the UK relative to other locations worldwide (*19*), the new variant may have already spread elsewhere undetected. Furthermore, although VOC 202012/01 was first identified in England, a rapidly spreading variant with similar phenotypic properties has also been detected in South Africa (*20*), where there has been a marked increase in

transmission in late 2020. This implies vaccination timelines will also be a crucial determinant of future burden in other countries where similar new variants are present. Second, there is a need to determine whether VOC 202012/01 – or any subsequent emerging lineages – could affect the efficacy of vaccines. Vaccine developers may therefore need to consider experimenting with variant sequences as a precautionary measure, and powering post-licensure studies to detect differences in efficacy between the preexisting and new variants. Licensing authorities may need to clarify abbreviated pathways to marketing for vaccines that involve altering strain formulation without any other changes to their composition.

We have examined the impact of a small number of intervention and vaccination scenarios, and the scenarios we project should not be regarded as the only available options for policymakers. Moreover, there are substantial uncertainties not fully captured by our model: for example, we do not explicitly model care home or hospital transmission of SARS-CoV-2, and we assume that there are no further changes in the infection fatality ratio (IFR) of SARS-CoV-2 in the future. The IFR for SARS-CoV-2 declined substantially in the UK over mid-2020 (*10*) and it may further decrease in 2021, or alternatively it may increase if there are substantial pressures on the health service. Finally, there are uncertainties in the choice of model used to generate these predictions, and the exact choice will yield differences in the measures needed to control the epidemic. We note that both the increased transmissibility and the increased child susceptibility hypotheses are consistent in their conclusion that the difficult societal decision of closing schools will be a key public health question in the months ahead.

There are some limitations to our analysis. We can only assess relative support in the data for the hypotheses proposed, but there may be other plausible mechanisms driving the resurgence of cases that we did not consider. Our conclusions about school closures were based on the assumption that children had reduced susceptibility and infectiousness compared to adults (*15*), whereas the precise values of these parameters and the impact of school closures (*21*) remains the subject of scientific debate (*21*). We based our assumptions about the efficacy of Tier 4 on the efficacy of the second national lockdown in England in November 2020, as the policies are very similar. The Tier 4 intervention has not been in place for long enough in order to reliably quantify its impact from epidemiological or behavioural data, and contact rates during the December holiday season may not be representative of other times. Finally, as the emergence of VOC 202012/01 has only recently been identified, our estimates may change substantially as more data become available.

Despite these limitations, we found strong evidence that VOC 202012/01 is spreading significantly faster within southeast England than preexisting non-VOC 202012/01 variants. Our

modelling analysis suggests this difference can be explained by an overall higher infectiousness of VOC 202012/01—with some evidence that the increase may be particularly marked in children—but not by a shorter latent period or immune escape alone. Further experimental work could provide insights into the biological mechanisms for our observations, but given our projections of a rapid rise in future incidence from VOC 202012/01 without additional control measures, there is an urgent need to consider what new approaches may be required to sufficiently reduce the ongoing transmission of SARS-CoV-2.

Author contributions and acknowledgements

Nicholas G. Davies, Rosanna C. Barnard, Christopher I. Jarvis, Adam J. Kucharski, James Munday, Carl A. B. Pearson, Timothy W. Russell, Damien C. Tully, Sam Abbott, Amy Gimma, William Waites, Kerry Wong, Kevin van Zandvoort, Rosalind M. Eggo, Sebastian Funk, Mark Jit, Katherine E. Atkins, and W. John Edmunds conceived the study, performed analyses, and wrote the manuscript. The CMMID COVID-19 Working Group provided discussion and comments. Stefan Flasche, Rein Houben, Yalda Jafari, Mihály Koltai, Fabienne Krauer, Yang Liu, Rachel Lowe, and Billy Quilty gave input during conception and manuscript drafting.

The CMMID COVID-19 Working Group is (randomized order): Sophie R Meakin, James D Munday, Amy Gimma, Rosanna C Barnard, Timothy W Russell, Billy J Quilty, Yang Liu, Stefan Flasche, Jiayao Lei, Adam J Kucharski, William Waites, Sebastian Funk, Fiona Yueqian Sun, Fabienne Krauer, Rachel Lowe, Nikos I Bosse , Damien C Tully, Emily S Nightingale, Katharine Sherratt, Rosalind M Eggo, Kaja Abbas, Kathleen O'Reilly, Hamish P Gibbs, C Julian Villabona-Arenas, Naomi R Waterlow, W John Edmunds, Graham Medley, Oliver Brady, Jack Williams, Alicia Rosello, Christopher I Jarvis, Petra Klepac, Mihaly Koltai, Nicholas G. Davies, Frank G Sandmann, Anna M Foss, Sam Abbott, Yalda Jafari, Kiesha Prem, Yung-Wai Desmond Chan, Katherine E. Atkins, Carl A B Pearson, Joel Hellewell, Kevin van Zandvoort, Simon R Procter, Thibaut Jombart, Gwenan M Knight, Akira Endo, Matthew Quaife, Mark Jit, Alicia Showering, Samuel Clifford.

Funding declarations for the CMMID COVID-19 Working Group are as follows. SRM: Wellcome Trust (grant: 210758/Z/18/Z). JDM: Wellcome Trust (grant: 210758/Z/18/Z). AG: European Commission (EpiPose 101003688). RCB: European Commission (EpiPose 101003688). TWR: Wellcome Trust (grant: 206250/Z/17/Z). BJQ: This research was partly funded by the National Institute for Health Research (NIHR) (16/137/109 & 16/136/46) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care. BJQ is supported in part by a grant from the Bill and Melinda Gates Foundation (OPP1139859). YL: Bill & Melinda Gates Foundation (INV-003174), NIHR (16/137/109), European Commission (101003688). SFlasche: Wellcome Trust (grant: 208812/Z/17/Z). JYL: Bill & Melinda Gates Foundation (INV-003174). AJK: Wellcome Trust (grant: 206250/Z/17/Z), NIHR (NIHR200908). WW: UK Medical Research Council (MRC) (grant MR/V027956/1).

SFunk: Wellcome Trust (grant: 210758/Z/18/Z), NIHR (NIHR200908). FYS: NIHR EPIC grant (16/137/109). FK: Innovation Fund of the Joint Federal Committee (Grant number 01VSF18015), Wellcome Trust (UNS110424). RL: Royal Society Dorothy Hodgkin Fellowship. NIB: Health Protection Research Unit (grant code NIHR200908). DCT: No funding declared. ESN: Bill & Melinda Gates Foundation (OPP1183986). KS: Wellcome Trust (grant: 210758/Z/18/Z). RME: HDR UK (grant: MR/S003975/1), MRC (grant: MC_PC 19065), NIHR (grant: NIHR200908). KA: Bill & Melinda Gates Foundation (OPP1157270, INV-016832). KO'R: Bill and Melinda Gates Foundation (OPP1191821). HPG: This research was produced by CSIGN which is part of the EDCTP2 programme supported by the European Union (grant number RIA2020EF-2983-CSIGN). The views and opinions of authors expressed herein do not necessarily state or reflect those of EDCTP. This research is funded by the Department of Health and Social Care using UK Aid funding and is managed by the NIHR. The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health and Social Care (PR-OD-1017-20001). CJVA: European Research Council Starting Grant (Action number 757688). NRW: Medical Research Council (grant number MR/N013638/1). WJE: European Commission (EpiPose 101003688), NIHR (NIHR200908). GFM: NTD Modelling Consortium by the Bill and Melinda Gates Foundation (OPP1184344). OJB: Wellcome Trust (grant: 206471/Z/17/Z). JW: NIHR Health Protection Research Unit and NIHR HTA. AR: NIHR (grant: PR-OD-1017-20002). CIJ: Global Challenges Research Fund (GCRF) project 'RECAP' managed through RCUK and ESRC (ES/P010873/1). PK: This research was partly funded by the Royal Society under award RP\EA\180004, European Commission (101003688), Bill & Melinda Gates Foundation (INV-003174). MK: Foreign, Commonwealth and Development Office / Wellcome Trust. NGD: UKRI Research England; NIHR Health Protection Research Unit in Immunisation (NIHR200929); UK MRC (MC_PC_19065). FGS: NIHR Health Protection Research Unit in Modelling & Health Economics, and in Immunisation. AMF: No funding declared. SA: Wellcome Trust (grant: 210758/Z/18/Z). YJ: LSHTM, DHSC/UKRI COVID-19 Rapid Response Initiative. KP: Bill & Melinda Gates Foundation (INV-003174), European Commission (101003688). YWDC: No funding declared. KEA: European Research Council Starting Grant (Action number 757688). CABP: CABP is supported by the Bill & Melinda Gates Foundation (OPP1184344) and the UK Foreign, Commonwealth and Development Office (FCDO)/Wellcome Trust Epidemic Preparedness Coronavirus research programme (ref. 221303/Z/20/Z). JH: Wellcome Trust (grant: 210758/Z/18/Z). KvZ: KvZ is supported by the UK Foreign, Commonwealth and Development Office (FCDO)/Wellcome Trust Epidemic Preparedness Coronavirus research programme (ref. 221303/Z/20/Z), and Elrha's Research for Health in Humanitarian Crises (R2HC) Programme, which aims to improve health outcomes by strengthening the evidence base for public health interventions in humanitarian crises. The R2HC programme is funded by the UK Government (FCDO), the Wellcome Trust, and the UK National Institute for Health Research (NIHR). SRP: Bill and Melinda Gates Foundation (INV-016832). TJ: RCUK/ESRC (grant: ES/P010873/1); UK PH RST; NIHR HPRU Modelling & Health Economics (NIHR200908). GMK: UK Medical Research Council (grant: MR/P014658/1). AE: The Nakajima Foundation. MQ: European Research Council Starting Grant (Action Number #757699); Bill and Melinda Gates Foundation (INV-001754). MJ: Bill & Melinda Gates Foundation (INV-003174, INV-016832), NIHR (16/137/109, NIHR200929, NIHR200908), European Commission (EpiPose 101003688). AS: No funding declared. SC: Wellcome Trust (grant: 208812/Z/17/Z).

Methods

Summary of second wave control measures in England

A second national 'lockdown' was announced in England on the 31st of October 2020, lasting for four weeks from the 5th of November to the 3rd of December 2020. Restrictions included a stay at home order with a number of exemptions including for exercise, essential shopping, obtaining or providing medical care, education and work for those unable to work from home. Non-essential shops, retail and leisure venues were required to close. Pubs, bars and restaurants were allowed to offer takeaway services only. Following the end of this second national lockdown, regions in England were assigned tiered local restrictions according to medium, high and very high alert levels (Tiers 1, 2 and 3). On the 19th of December 2020, the UK government announced that a number of regions in the South East of England would be placed into a new higher 'Tier 4', corresponding to a Stay at Home alert level. Regional Tier 4 restrictions were broadly similar to the national lockdown restrictions, except that Tier 4 allowed places of worship to continue providing services and did not have a fixed four-week duration.

Proportion of variant over time

We used publicly-available data from the COG-UK Consortium (*12*, *22*) to calculate the proportion of VOC 202012/01 over time from 1 October to 1 December. Values were aggregated by NHS region and by day, and pointwise binomial confidence intervals were estimated for each day.

Estimate of R_t

We calculated the weekly proportion of positive tests that were S-gene negative during the week beginning 4 December out of all positive tests that tested for the S-gene by English local authority. We used reproduction number estimates using the method described in (*23*) and (*24*) and implemented in the EpiNow2 R package (*25*), downloaded from https://github.com/epiforecasts/covid-rt-estimates/blob/master/subnational/united-kingdom-local/cases/summary/rt.csv.

Estimation of change in mobility over time

We used anonymised mobility data collected from smartphone users by Google Community Mobility (*8*). Percentage change in mobility per day was calculated for each lower-tier local authority in England and a generalised additive model with a spline for time was fitted to these observations to provide a smoothed effect of the change in mobility over time.

Estimation of social contacts over time

We used data on reported social contacts from the CoMix survey (*9*), which is a weekly survey of face-to-face contact patterns, taken from a sample of approximately 2500 individuals broadly representative of the UK population with respect to age and geographical location. We calculated the distribution of contacts using 1000 bootstrap samples with replacement from the raw data. Bootstrap samples were calculated at the participant level, then all observations for those participants are included in a sample to respect the correlation structure of the data. We collect data in two panels which alternate weekly, therefore we calculated the mean smoothed over the 2 week intervals to give a larger number of participants per estimate and account for panel effects. We calculated the mean number of contacts (face to face conversational contact or physical contact) in the settings "home", "work", "education" (including childcare, nurseries, schools and universities and colleges), and "other" settings. We calculate the mean contacts by age group and area of residence (those areas which were subsequently placed under Tier 4 restrictions on December 20th as they were experiencing high and rapidly increasing incidence and those areas of England that were not placed under these restrictions). The mean number of contacts is influenced by a few individuals who report very high numbers of contacts (often in a work context). The means shown here are calculated based on truncating the maximum number of contacts recorded at 200 per individual per day.

Transmission dynamic model

We extended a previously developed modelling framework structured by age and spatial region (*10*, *11*) to include two variants of SARS-CoV-2 (VOC 202012/01 and non-VOC 202012/01) (**Fig. S1**). The model is a discrete-time deterministic compartmental model which allows for arbitrary delay distributions for transitions between compartments. We fitted this model to multiple regionally-stratified data sources across the 7 NHS England regions as previously: deaths, hospital admissions, hospital bed occupancy, ICU bed occupancy, daily incidence of new infections, PCR prevalence of active infection, and seroprevalence. In addition, for the East of England, South East and London regions, we fit to the daily frequency of VOC 202012/01 across each of the regions as reported in COG-UK data. Tiered restrictions were parameterised to match the stringency of Tiers 1–3 in the period of 14 October – 4 November (*10*), with Tier 4 parameterised to match the stringency of the second national lockdown in England (5th November – 1st December 2020) as measured by changes to Google Mobility indices for these periods (*10*). We assumed that Tiers 1–3 in the period after the end of the second national lockdown (December 2 onward) had the same stringency as Tiers 1–3 before the national lockdown, but that the threshold for a region entering Tier 3 was 200 cases per 100,000 individuals per week rather than 300 cases per 100,000 individuals per week, which was the assumption we adopted for previous work examining the impact of tiered restrictions (*10*). We also examined keeping schools (primary and secondary as well as universities) open versus

keeping schools closed during these lockdowns, and examined two alternative vaccination programmes (200,000 vaccinations per week versus 2,000,000 vaccinations per week). To model school closure, we removed all school contacts from our contact matrix based upon POLYMOD data and varying over time according to Google Mobility indices (as described previously (*10*)) and also assumed that transmissibility in all age groups would be reduced by a further multiplicative factor estimated previously from an observed increase in transmission in England when schools reopened in September 2020 (*10*). See Supporting Information for details of Bayesian inference including likelihood functions and prior distributions.

Apparent growth of VOC 202012/01 not a result of testing artefacts

The apparent frequency of VOC 202012/01 could be inflated relative to reality if this variant leads to increased test-seeking behaviour (e.g. if it leads to a higher rate of symptoms than preexisting variants). However, this would not explain the growth in the relative frequency of VOC 202012/01 over time. Mathematically, if variant 1 has growth rate r_1 and variant 2 has growth rate r_2, the relative frequency over time is $\exp(r_2 t) / (\exp(r_1 t) + \exp(r_2 t))$. However, if variant 1 has probability x of being reported and variant 2 has probability y, and both have growth rate r, the relative frequency over time is $y \exp(rt) / (x \exp(rt) + y \exp(rt))$, which is constant.

Code and data availability

Analysis code and data are available at https://www.github.com/nicholasdavies/newcovid.

Supporting Information

Details of Bayesian inference

To fit the model to data on deaths, hospital admissions, hospital bed and ICU bed occupancy, PCR positivity, and seroprevalence for each of the 7 NHS England regions, we performed Bayesian inference using Markov chain Monte Carlo, employing the Differential Evolution MCMC algorithm (*26*). For each posterior sample, we simulated epidemics from 1 January to 15 December 2020, using data that were current as of 20 December 2020. We used Google Community Mobility data up to 15 December 2020 to capture how interpersonal contact rates changed over the course of the epidemic; from 16 December 2020, we assumed that mobility indicators were "frozen" at their mean values for each region as measured over the week of 9–15 December, with further changes dictated by the introduction of tiered restrictions and lockdowns.

As previously described (*10*), when fitting deaths, hospital admissions, hospital bed occupancy and ICU bed occupancy, we used a negative binomial likelihood with size parameter fixed at 20 for each daily data point. For seroprevalence and PCR prevalence, we used a skew-normal likelihood for each data point fitted to produce the same mean and 95% confidence interval as was reported for the data, and took the expected value of the model prediction over the date range during which the prevalence was measured. For fitting to VOC 202012/01 relative frequency over time in the three heavily affected NHS England regions, we used a binomial likelihood with the daily proportion of detected samples that were VOC 202012/01.

As part of model estimation, we separately fit for each region: the start time of community transmission; the basic reproduction number R_0 prior to any changes in mobility or closure of schools; the delay from infection to hospital admission, to ICU admission, and to death; a region-specific relative probability of hospital admission and of ICU admission given infection; the relative infection fatality ratio at the start and at the end of the simulation period, as fatality due to COVID-19 has dropped substantially over time in the UK; a decreasing rate of effective contact between individuals over time, representing better practices of self-isolation and precautions against infection taken by individuals over the course of the year; coefficients determining the relative mobility of younger people, around age 20, relative to the rest of the population, for the months of July, August, and September onwards; and the magnitude and timing of a boost to R_0 around the end of summer, which we hypothesise is related to the opening of schools, but which was not fully captured in our model by the resumption of school-specific contacts on 1 September. Full details of all fitted parameters, along with prior distributions assumed for each parameter, are in Table S1.

We use two parametric functions extensively in parameterising the model. The first,

$$logistic(x) = exp(x) / (1 + exp(x))$$

is the standard logistic curve. The second,

$$asc(x, y_0, y_1, s_0, s_1) = y_0 + (y_1 - y_0)\,[logistic(s_o + x(s_1 - s_0)) - logistic(s_0)] / [logistic(s_1) - logistic(s_0)]$$

is a logistic-shaped curve parameterised to be a smooth S-shaped function of x from 0 to 1, which goes from y_0 at x = 0 to y_1 at x = 1, with an inflection point at $x = -s_0/(-s_0 + s_1)$ if $s_0 < 0$ and $s_1 > 0$.

Basic epidemiological parameters were broadly informed from the literature and previously reported (*10*). All parameters that we adopted as assumptions are given in Table S2.

Table S1. Details of fitted parameters.

Parameter	Description	Prior distribution	Notes
`tS`	Start date of epidemic in days after 1 January 2020	~uniform(0, 60)	Determines date at which seeding begins in region; starting on this date, one random individual per day contracts SARS-CoV-2 for 28 days
`u`	Basic susceptibility to infection	~normal(0.07, 0.01)	Determines basic reproduction number R_0
`death_mean`	Mean delay in days from start of infectious period to death	~normal(15, 2)	Prior informed by analysis of CO-CIN data (*27*)
`death_shape`	Shape parameter of gamma distribution for delay from start of infectious period to death	~normal(1.9, 0.2)	Prior informed by analysis of CO-CIN data (*27*)
`admission`	Mean delay in days from start of infectious period to hospital admission	~normal(7.5, 1)	Delay is assumed to follow a gamma distribution with shape parameter 0.71. Prior and shape of distribution informed by analysis of CO-CIN data (*27*).
`icu_admission`	Mean delay in days from start of infectious period to ICU admission	~normal(11.1, 1)	Delay is assumed to follow a gamma distribution with shape parameter 1.91. Prior and shape of distribution informed by analysis of CO-CIN data (*27*).
`hosp_rlo`	Log-odds of hospital admission, relative to age-specific probabilities of hospital admission given infection derived from Salje et al. (*28*).	~normal(0, 0.1)	Based on Salje et al. (*28*), we assumed that the basic shape of the age-specific probability of hospitalisation given infection was *logistic(7.37 + 0.068a)*, where *a* is the individual's age in years. This overall relationship is then adjusted according to the `hosp_rlo` parameter.
`icu_rlo`	Log-odds of ICU admission, relative to age-specific probabilities of ICU admission given hospital admission derived from CO-CIN data.	~normal(0, 0.1)	We fit a spline to CO-CIN data on hospital admission and ICU admission by age to derive the basic age-specific probability of ICU admission, which was then adjusted based on the `icu_rlo` parameter.
`cfr_rel`	Relative fatality rate of COVID-19 at beginning of 2020	normal(1, 0.05)	Based on Levin et al. (*29*), we assumed the basic shape of the age-specific infection fatality ratio of SARS-CoV-2 was *logistic(–7.56 + 0.121a*) (see entry for `hosp_rlo`). This is multiplied by `cfr_rel` to adjust the fatality rate for each region.
`cfr_rel2`	Relative fatality rate of COVID-19 at end of 2020	~normal(0.45, 0.1)	Based on CO-CIN data [27], we estimated that the mortality rate of COVID-19 decreased by approximately 55% by September 2020 relative to the beginning of the year. The product of `cfr_rel` and `cfr_rel2` gives the mortality rate by September. Specifically, the IFR is multiplied by a factor *asc(t / 366, cfr_rel, cfr_rel * cfr_rel2, –2.9, 7.8)* where *t* is the time in days since 1 January 2020.

`contact_final`	Relative rate of effective contact at end of 2020	~normal(1, 0.1) ≤ 1	To capture continued low incidence of SARS-CoV-2 infection in spite of rising contact rates as shown by mobility data and social contact surveys, we assume that the effective contact rate over time is multiplied by a factor $asc(t/366, 1, contact_{final}, -contact_{s0}, contact_{s1})$, where t is time in days since 1 January 2020.
`contact_s0`	Parameter for curve specified by `contact_final`	~exponential(0.1)	
`contact_s1`	Parameter for curve specified by `contact_final`	~exponential(0.1)	
`concentration1`	Increased contact among young people in July	~normal(2, 0.3) ≥ 2	Because initial increases in SARS-CoV-2 prevalence from July in England were especially apparent in young people, we allow increases in mobility to be more emphasized in young people starting from July. We model a relative contact-rate multiplier for individuals of age a as $beta(a/100 \mid a = 0.2(k-2)+1, b = 0.8(k-2)+1)$, where k is the concentration parameter and *beta* is the beta distribution probability density function. This gives flat contact rates across age groups when $k = 2$, and relatively higher contact rates in individuals around age 20 when $k > 2$.
`concentration2`	Increased contact among young people in August	~normal(2, 0.2) ≥ 2	
`concentration3`	Increased contact among young people from September	~normal(2, 0.1) ≥ 2	
`sep_boost`	Increase in transmission around 1 September 2020	~normal(1, 0.05)	After the date specified by `sep_when`, transmission is multiplied by the factor `sep_boost`. This is to capture a sudden increase in transmission rates observed around 1 September in England.
`sep_when`	Date of increase in transmission	~uniform(224, 264) (i.e. 12 Aug–21 Sep)	

Parameters for VOC 202012/01 strain

Parameter	Description	Prior distribution	Notes
`v2_when`	Introduction date of VOC 202012/01 in days after 1 January 2020	~uniform(144, 365)	On this date, ten random individuals contract VOC 202012/01
`v2_hosp_rlo`	Relative log-odds of hospitalisation for VOC 202012/01, compared to preexisting variants	~normal(0, 0.1)	Vague prior
`v2_cfr_rel`	Relative rate of death for VOC 202012/01, compared to preexisting variants	~normal(0, 0.1)	Vague prior
`v2_relu`	Relative transmission rate of VOC 202012/01, compared to preexisting variants, for model A	~lognormal(0, 0.2)	Vague prior
`v2_immesc`	Relative transmission rate of VOC 202012/01, for model B	~beta(2, 1)	Vague prior

`v2_susc`	Relative transmission rate of VOC 202012/01, compared to preexisting variants, for model C	~uniform(0, 1)	Vague prior; *v2_susc* = 0 corresponds to children having reduced susceptibility relative to adults as in ref. (*15*), while *v2_susc* = 1 corresponds to children having the same susceptibility as adults.
`v2_latent`	Relative length of latent period of VOC 202012/01, compared to preexisting variants, for model D	~normal(1, 0.2)	Vague prior

Table S2. Model parameters not subject to fitting.

Parameter	Description	Value	Notes
d_E	Latent period (E to I_P, E to I_S, L to I_S; days)	~gamma(μ = 2.5, k = 2.5)	Set to 2.5 so that incubation period (latent period plus period of preclinical infectiousness) is 5 days(*30*)
d_P	Duration of preclinical infectiousness (I_P to I_C; days)	~gamma(μ = 2.5, k = 4)	Assumed to be half the duration of total infectiousness in clinically-infected individuals (*31*)
d_C	Duration of clinical infectiousness (I_C to R; days)	~gamma(μ = 2.5, k = 4)	Infectious period set to 5 days, to result in a serial interval of approximately 6 days(*32*–*34*)
d_S	Duration of subclinical infectiousness (I_S to R; days)	~gamma(μ = 5.0, k = 4)	Assumed to be the same duration as total infectious period for clinical cases, including preclinical transmission
y_i	Probability of clinical symptoms given infection for age group i	Estimated from case distributions across 6 countries	(*15*)
f	Relative infectiousness of subclinical cases	50%	Assumed(*11*, *15*)
$c_{i,j}$	Number of age-j individuals contacted by an age-i individual per day, prior to changes in mobility	UK-specific contact matrix	(*35*)
N_i	Number of age-i individuals	From demographic data	(*36*)
Δt	Time step for discrete-time simulation	0.25 days	
$P(\text{ICU})_i$	Proportion of hospitalised cases that require critical care for age group i	Estimated from CO-CIN data	(*27*)
w_s	Waning rate of seropositivity	224 days^{-1}	Estimated from serology data
los_{hosp}	Length of stay in hospital	~lognormal(μ_{log} = 11.08, sd_{log} = 1.20)	Estimated from CO-CIN data (*27*)
los_{icu}	Length of stay in ICU	~lognormal(μ_{log} = 13.33, sd_{log} = 1.25)	Estimated from CO-CIN data (*27*)

$detect_0$, $detect_1$, $detect_{s0}$, $detect_{s1}$	Delay from hospital admission to SARS-CoV-2 test	$detect_0$ = 14 $detect_1$ = 1 $detect_{s0}$ = 5.86 $detect_{s1}$ = 33.4	To capture substantial delays in testing at the beginning of the epidemic in the UK, we assume that the delay from hospital admission to confirmed SARS-CoV-2 infection is $asc(t/366, detect_0, detect_1, detect_{s0}, detect_{s1})$, where t is time in days since 1 January 2020. Estimated from a previous round of model fitting.

Supplementary Figures

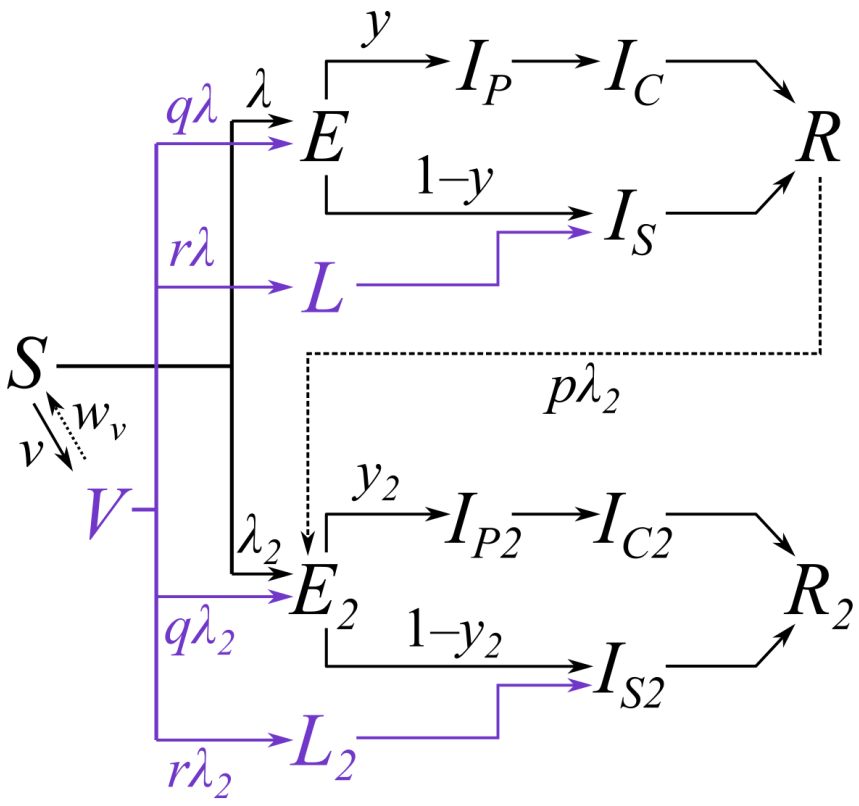

Fig. S1. Diagram of the two-strain model with vaccination. Subscripts for age group and region are omitted from this diagram and only certain key parameters are shown. Compartments and processes in purple apply to the vaccine model only. S, susceptible; E, exposed; L, latent (see below); I_P, preclinically infectious; I_C, clinically infectious; I_S, subclinically infectious; R, recovered; V, vaccinated. Subscript 2 represents compartments and parameters for VOC 202012/01. Above, λ and λ_2 are the force of infection for preexisting variants versus VOC 202012/01; y and y_2 are the fraction of cases that develop clinical symptoms for preexisting variants versus VOC 202012/01; v is the rate of vaccination; w_v is the waning rate of vaccination (assumed to be zero for this manuscript); p captures cross-protection against VOC 202012/01 conferred by immunity to preexisting variants; q captures vaccine protection against disease; and r captures vaccine protection against infection. L and L_2 are additional compartments for a latent period prior to subclinical infection only (i.e. with zero probability of clinical infection). For a vaccine with efficacy against disease e_d (e.g. $e_d = 0.95$ for this manuscript) and efficacy against infection e_i (e.g. $e_i = 0.6$ for this manuscript), we assume $r = (1 - e_i) * e_d$ and $q = (1 - e_i) * (1 - e_d)$.

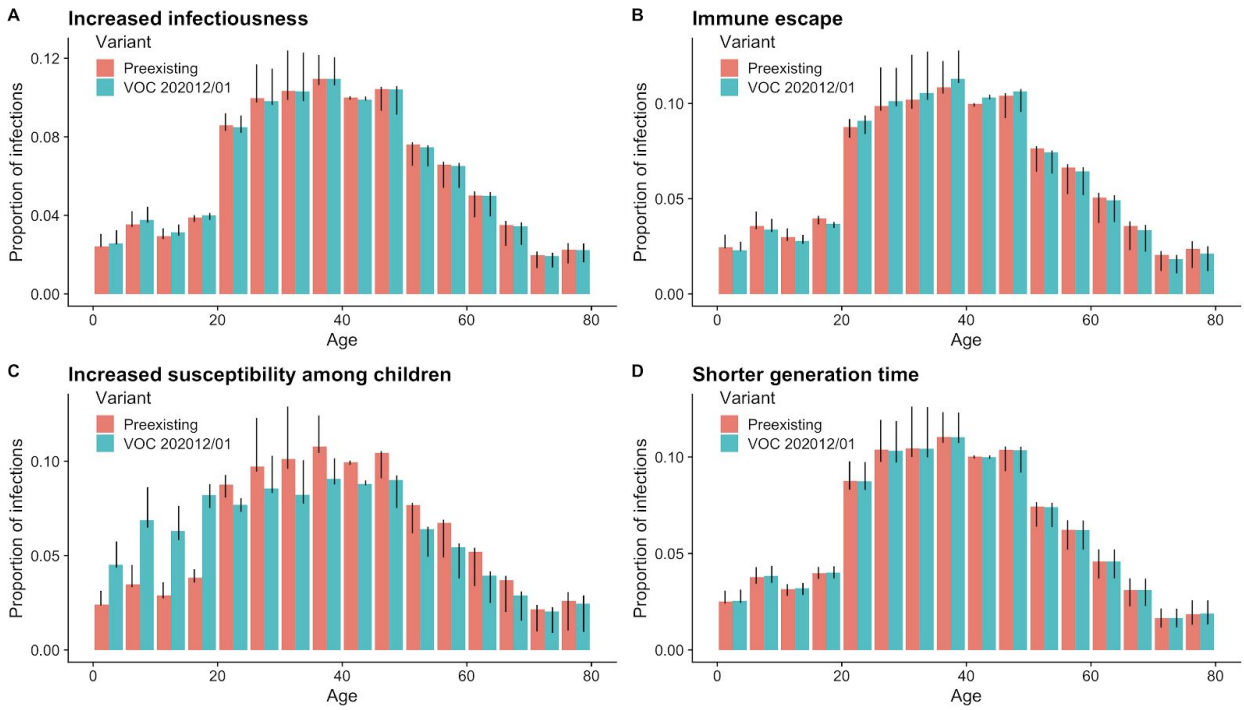

Fig. S2. Comparison of age distribution of infections in the two-strain model depending upon biological mechanism of VOC 202012/01 growth rate. Each panel shows a different assumed mechanism: **(A)** increased transmissibility; **(B)** immune escape; **(C)** increased susceptibility among children; **(D)** shorter generation time. Measured from infections in the fitted model between 1 October and 30 November, 2020, in the London, South East, and East of England NHS regions. Age distribution of infections are similar for both variants, except in hypothesis C (increased susceptibility among children), where VOC 202012/01 is relatively overrepresented in children.

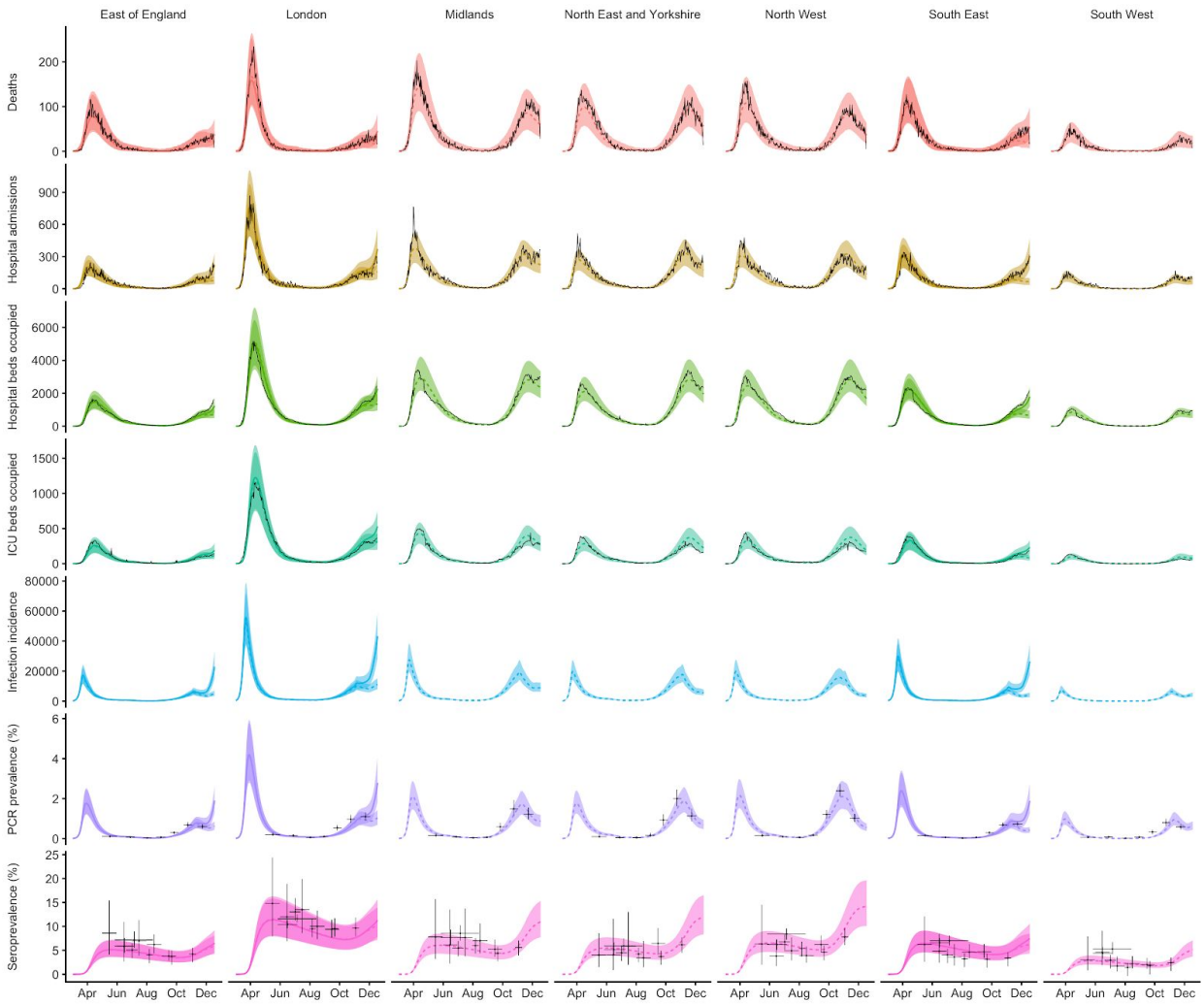

Fig. S3. Comparison of fitted model and data across 7 NHS England regions. Black lines show observed data, while coloured lines and shaded regions show median and 95% credible intervals from the fitted model. For London, South East and East of England, solid lines show fits for the model with two variants. For all NHS England regions, dashed lines show model fits assuming a single variant.

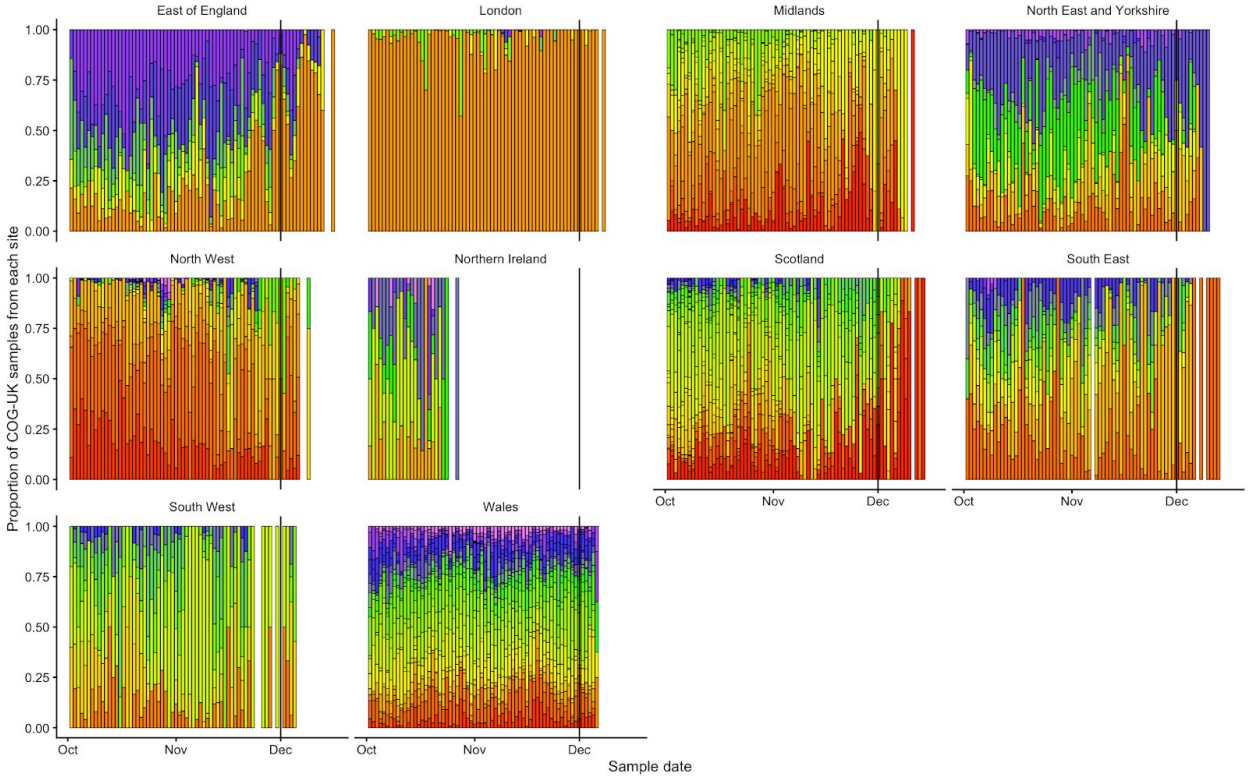

Fig S4. Proportion of COG-UK samples from each sample site in the UK by NHS Region / devolved administration. This plot shows the spatial heterogeneity in samples available for analysis from COG-UK over time. Each colour represents a different amalgamated longitude and latitude for the residence of the individual who provided the sample. There is an apparent lack of diversity in London sampling sites because the public COG-UK data (*22*) provide rounded longitude and latitude, and London has a much smaller area than other NHS regions. After 1 December 2020 (black vertical line), the diversity of sampling sites decreases. Accordingly, we cut off our analysis of COG-UK samples after this point.

References

1. Public Health England, Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01 (2020), (available at https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201).

2. H. Gu, Q. Chen, G. Yang, L. He, H. Fan, Y.-Q. Deng, Y. Wang, Y. Teng, Z. Zhao, Y. Cui, Y. Li, X.-F. Li, J. Li, N.-N. Zhang, X. Yang, S. Chen, Y. Guo, G. Zhao, X. Wang, D.-Y. Luo, H. Wang, X. Yang, Y. Li, G. Han, Y. He, X. Zhou, S. Geng, X. Sheng, S. Jiang, S. Sun, C.-F. Qin, Y. Zhou, Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science*. **369**, 1603–1607 (2020).

3. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*. **182**, 1295–1310.e20 (2020).

4. M. Hoffmann, H. Kleine-Weber, S. Pöhlmann, A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell*. **78**, 779–784.e5 (2020).

5. T. P. Peacock, D. H. Goldhill, J. Zhou, L. Baillon, R. Frise, O. C. Swann, R. Kugathasan, R. Penn, J. C. Brown, R. Y. Sanchez-David, L. Braga, M. K. Williamson, J. A. Hassard, E. Staller, B. Hanley, M. Osborn, M. Giacca, A. D. Davidson, D. A. Matthews, W. S. Barclay, The furin cleavage site of SARS-CoV-2 spike protein is a key determinant for transmission due to enhanced replication in airway cells, , doi:10.1101/2020.09.30.318311.

6. S. A. Kemp, D. A. Collier, R. Datir, S. Gayed, A. Jahun, M. Hosmillo, I. A. Ferreira, C. Rees-Spear, P. Mlcochova, I. U. Lumb, D. Roberts, A. Chandra, N. Temperton, K. Sharrocks, E. Blane, J. A. Briggs, The COVID-19 Genomics UK (COG-UK) Consortium, The CITIID-NIHR BioResource COVID-19 Collaboration, K. G. Smith, J. R. Bradley, C. Smith, R. Goldstein, I. G. Goodfellow, A. Smielewska, J. P. Skittrall, T. Gouliouris, E. Gkrania-Klotsas, C. J. R. Illingworth, L. E. McCoy, R. K. Gupta, Neutralising antibodies drive Spike mediated SARS-CoV-2 evasion. *medRxiv*, 2020.12.05.20241927 (2020).

7. S. Kemp, W. Harvey, R. Datir, D. Collier, I. Ferreira, A. Carabelii, D. L. Robertson, R. K. Gupta, Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion ΔH69/V70. *Cold Spring Harbor Laboratory* (2020), p. 2020.12.14.422555.

8. Google, COVID-19 Community Mobility Reports, (available at https://www.google.com/covid19/mobility/).

9. C. I. Jarvis, K. Van Zandvoort, A. Gimma, K. Prem, CMMID COVID-19 working group, P. Klepac, G. J. Rubin, W. J. Edmunds, Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med.* **18**, 124 (2020).

10. Assessment of tiered restrictions and a second lockdown on COVID-19 deaths and hospitalisations in England: a modelling study (2020), (available at https://cmmid.github.io/topics/covid19/uk-tiers-2nd-lockdown.html).

11. N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, W. J. Edmunds, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group, Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital

services in the UK: a modelling study. *Lancet Public Health*. **5**, e375–e385 (2020).

12. Data - COG-UK Consortium, (available at https://www.cogconsortium.uk/data/).

13. Real-Time Assessment of Community Transmission: Findings, (available at https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/real-time-assessment-of-community-transmission-findings/).

14. NERVTAG, "NERVTAG meeting on SARS-CoV-2 variant under investigation VUI-202012/01," (available at https://khub.net/documents/135939561/338928724/SARS-CoV-2+variant+under+investigation%2C+meeting+minutes.pdf/962e866b-161f-2fd5-1030-32b6ab467896).

15. N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, CMMID COVID-19 working group, R. M. Eggo, Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211 (2020).

16. R. M. Viner, O. T. Mytton, C. Bonell, G. J. Melendez-Torres, J. Ward, L. Hudson, C. Waddington, J. Thomas, S. Russell, F. van der Klis, A. Koirala, S. Ladhani, J. Panovska-Griffiths, N. G. Davies, R. Booy, R. M. Eggo, Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. *JAMA Pediatr.* (2020), doi:10.1001/jamapediatrics.2020.4573.

17. K. W. Ng, N. Faulkner, G. H. Cornish, A. Rosa, R. Harvey, S. Hussain, R. Ulferts, C. Earl, A. G. Wrobel, D. J. Benton, C. Roustan, W. Bolland, R. Thompson, A. Agua-Doce, P. Hobson, J. Heaney, H. Rickman, S. Paraskevopoulou, C. F. Houlihan, K. Thomson, E. Sanchez, G. Y. Shin, M. J. Spyer, D. Joshi, N. O'Reilly, P. A. Walker, S. Kjaer, A. Riddell, C. Moore, B. R. Jebson, M. Wilkinson, L. R. Marshall, E. C. Rosser, A. Radziszewska, H. Peckham, C. Ciurtin, L. R. Wedderburn, R. Beale, C. Swanton, S. Gandhi, B. Stockinger, J. McCauley, S. J. Gamblin, L. E. McCoy, P. Cherepanov, E. Nastouli, G. Kassiotis, Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science*. **370**, 1339–1343 (2020).

18. Public Health England, National flu and COVID-19 surveillance reports (2020), (available at https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports).

19. E. C. Thomson, L. E. Rosen, J. G. Shepherd, R. Spreafico, A. da Silva Filipe, J. A. Wojcechowskyj, C. Davis, L. Piccoli, D. J. Pascall, J. Dillen, S. Lytras, N. Czudnochowski, R. Shah, M. Meury, N. Jesudason, A. De Marco, K. Li, J. Bassi, A. O'Toole, D. Pinto, R. M. Colquhoun, K. Culap, B. Jackson, F. Zatta, A. Rambaut, S. Jaconi, V. B. Sreenu, J. Nix, R. F. Jarrett, M. Beltramello, K. Nomikou, M. Pizzuto, L. Tong, E. Cameroni, N. Johnson, A. Wickenhagen, A. Ceschi, D. Mair, P. Ferrari, K. Smollett, F. Sallusto, S. Carmichael, C. Garzoni, J. Nichols, M. Galli, J. Hughes, A. Riva, A. Ho, M. G. Semple, P. J. M. Openshaw, J. Kenneth Baillie, The ISARIC4C Investigators, the COVID-19 Genomics UK (COG-UK) consortium, S. J. Rihn, S. J. Lycett, H. W. Virgin, A. Telenti, D. Corti, D. L. Robertson, G. Snell, The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. *Cold Spring Harbor Laboratory* (2020), p. 2020.11.04.355842.

20. H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E. J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A. J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, M. Hsiao, S. Korsman, M.-A. Davies, L. Tyers, I. Mudau, D.

York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A. Alisoltani-Dehkordi, A. Godzik, C. K. Wibmer, B. T. Sewell, J. Lourenco, L. C. J. Alcantara, S. L. Kosakovsky Pond, S. Weaver, D. Martin, R. J. Lessells, J. N. Bhiman, C. Williamson, T. de de Oliveira, Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*, 2020.12.21.20248640 (2020).

21. The impact of school holidays on the social mixing patterns of school children. *Epidemics*. **3**, 103–108 (2011).

22. Imperial College London, Microreact - Open data visualization and sharing for genomic epidemiology, (available at https://beta.microreact.org/project/kcdhiZuzrAzPb52qmMB9bj-cog-uk-2020-12-18-sars-cov-2-in-the-uk/).

23. S. Abbott, J. Hellewell, R. N. Thompson, K. Sherratt, H. P. Gibbs, N. I. Bosse, J. D. Munday, S. Meakin, E. L. Doughty, J. Y. Chun, Y.-W. D. Chan, F. Finger, P. Campbell, A. Endo, C. A. B. Pearson, A. Gimma, T. Russell, S. Flasche, A. J. Kucharski, R. M. Eggo, S. Funk, CMMID COVID modelling group, Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*. **5** (2020), p. 112.

24. K. Sherratt, S. Abbott, S. R. Meakin, J. Hellewell, J. D. Munday, N. Bosse, CMMID Covid-19 working group, M. Jit, S. Funk, Evaluating the use of the reproduction number as an epidemiological tool, using spatio-temporal trends of the Covid-19 outbreak in England. *medRxiv*, 2020.10.18.20214585 (2020).

25. S. Abbott, EpiNow, (available at https://zenodo.org/record/3957490#.X-JX-en7SHE).

26. C. J. F. T. Braak, C. J. F. Ter Braak, A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*. **16** (2006), pp. 239–249.

27. A. B. Docherty, E. M. Harrison, C. A. Green, H. E. Hardwick, R. Pius, L. Norman, K. A. Holden, J. M. Read, F. Dondelinger, G. Carson, L. Merson, J. Lee, D. Plotkin, L. Sigfrid, S. Halpin, C. Jackson, C. Gamble, P. W. Horby, J. S. Nguyen-Van-Tam, A. Ho, C. D. Russell, J. Dunning, P. J. Openshaw, J. K. Baillie, M. G. Semple, ISARIC4C investigators, Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ*. **369**, m1985 (2020).

28. H. Salje, C. Tran Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, S. Cauchemez, Estimating the burden of SARS-CoV-2 in France. *Science*. **369**, 208–211 (2020).

29. A. T. Levin, W. P. Hanage, N. Owusu-Boaitey, K. B. Cochran, S. P. Walsh, G. Meyerowitz-Katz, Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur. J. Epidemiol.* (2020), doi:10.1007/s10654-020-00698-1.

30. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G.

Reich, J. Lessler, The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Intern. Med.* **172**, 577–582 (2020).

31. L. Ferretti, A. Ledda, C. Wymant, L. Zhao, V. Ledda, L. A.- Dorner, M. Kendall, A. Nurtay, H.-Y. Cheng, T.-C. Ng, H.-H. Lin, R. Hinch, J. Masel, A. Marm Kilpatrick, C. Fraser, The timing of COVID-19 transmission. *medRxiv*, 2020.09.04.20188516 (2020).

32. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).

33. H. Nishiura, N. M. Linton, A. R. Akhmetzhanov, Serial interval of novel coronavirus (COVID-19) infections, , doi:10.1101/2020.02.03.20019497.

34. Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, T. Feng, Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*. **20** (2020), pp. 911–919.

35. J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, W. J. Edmunds, Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).

36. N. Park, Population estimates for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics (2019), (available at https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2018).