Article

Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7

Nicholas G. Davies[1]✉, Christopher I. Jarvis[1], CMMID COVID-19 Working Group*, W. John Edmunds[1], Nicholas P. Jewell[2,3], Karla Diaz-Ordaz[2,3,5] & Ruth H. Keogh[2,3,5]

SARS-CoV-2 lineage B.1.1.7, a variant that was first detected in the UK in September 2020[1], has spread to multiple countries worldwide. Several studies have established that B.1.1.7 is more transmissible than pre-existing variants, but have not identified whether it leads to any change in disease severity[2]. Here we analyse a dataset that links 2,245,263 positive SARS-CoV-2 community tests and 17,452 deaths associated with COVID-19 in England from 1 November 2020 to 14 February 2021. For 1,146,534 (51%) of these tests, the presence or absence of B.1.1.7 can be identified because mutations in this lineage prevent PCR amplification of the spike (*S*) gene target (known as *S* gene target failure (SGTF)[1]). On the basis of 4,945 deaths with known SGTF status, we estimate that the hazard of death associated with SGTF is 55% (95% confidence interval, 39–72%) higher than in cases without SGTF after adjustment for age, sex, ethnicity, deprivation, residence in a care home, the local authority of residence and test date. This corresponds to the absolute risk of death for a 55–69-year-old man increasing from 0.6% to 0.9% (95% confidence interval, 0.8–1.0%) within 28 days of a positive test in the community. Correcting for misclassification of SGTF and missingness in SGTF status, we estimate that the hazard of death associated with B.1.1.7 is 61% (42–82%) higher than with pre-existing variants. Our analysis suggests that B.1.1.7 is not only more transmissible than pre-existing SARS-CoV-2 variants, but may also cause more severe illness.

Most community SARS-CoV-2 PCR tests in England are processed by one of six national 'Lighthouse' laboratories. Among the mutations carried by the B.1.1.7 lineage—also known as variant of concern (VOC) 202012/01—is a six-nucleotide deletion that prevents the amplification of the *S* gene target by the commercial PCR assay that is currently used in three of the Lighthouse laboratories[1]. By linking individual records of positive community tests with and without SGTF to a comprehensive line list of deaths associated with COVID-19 in England, we estimate the relative hazard of death associated with infection with the B.1.1.7 variant. We define confirmed SGTF as a compatible PCR result with cycle threshold (C_t) < 30 for *orf1ab*, C_t < 30 for the nucleocapsid (*N*) gene and no detectable *S* (C_t > 40); confirmed non-SGTF as any compatible PCR result with C_t < 30 for each of the *orf1ab*, *N* and *S* genes; and an inconclusive (missing) result as any other positive test, including tests processed by a laboratory that is incapable of assessing SGTF.

Characteristics of the study population

The study sample (Extended Data Table 1) comprises 2,245,263 individuals who had a positive community ('Pillar 2') test between 1 November 2020 and 14 February 2021. Just over half of those tested (1,146,534; 51.1%) had a conclusive SGTF reading and, of these, 58.8% had SGTF. Female individuals comprised 53.6% of the total sample; 44.3% of

individuals were aged 1–34 years, 34.4% aged 35–54 years, 15.1% aged 55–69 years, 4.3% aged 70–84 years and 1.9% aged 85 years or older. The majority of individuals (93.7%) lived in residential accommodation (defined as a house, flat, sheltered accommodation or house in multiple occupancy) and 3.1% lived in a care or nursing home. On the basis of self-identified ethnicity, 74.0% were white, 13.6% were Asian, 4.6% were Black and 7.8% were of other, mixed or unknown ethnicity. All seven NHS England regions are represented, with the London region contributing 22.5% of tests and the South West 5.9%. The first three weeks of the study period (1–21 November) contributed 15.5% of the total tests, and the final three weeks (24 January–14 February) 12.8%. The period between 3 and 23 January contributed 31.6% of tests.

For those samples for which SGTF status was measured, SGTF prevalence was similar in male and female individuals, but lower in the older age groups: 59.0% in 1–34-year-old individuals compared with 55.4% in those aged 85 years and older. In keeping with these age patterns, SGTF prevalence was lower in individuals living in a care or nursing home (54.3%) than those in residential accommodation (58.8%). SGTF prevalence by self-identified ethnicity was 58.0% in the white group, 57.6% in the Asian group, 69.6% in the Black group and 64.8% in the other, mixed or unknown ethnicity group. SGTF prevalence was the lowest in the most-deprived quintile of the index of multiple deprivation[3] (IMD) (53.9%). The highest prevalences of SGTF over the study period

[1]Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK. [2]Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. [3]Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, London, UK. [5]These authors contributed equally: Karla Diaz-Ordaz, Ruth H. Keogh. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: nicholas.davies@lshtm.ac.uk
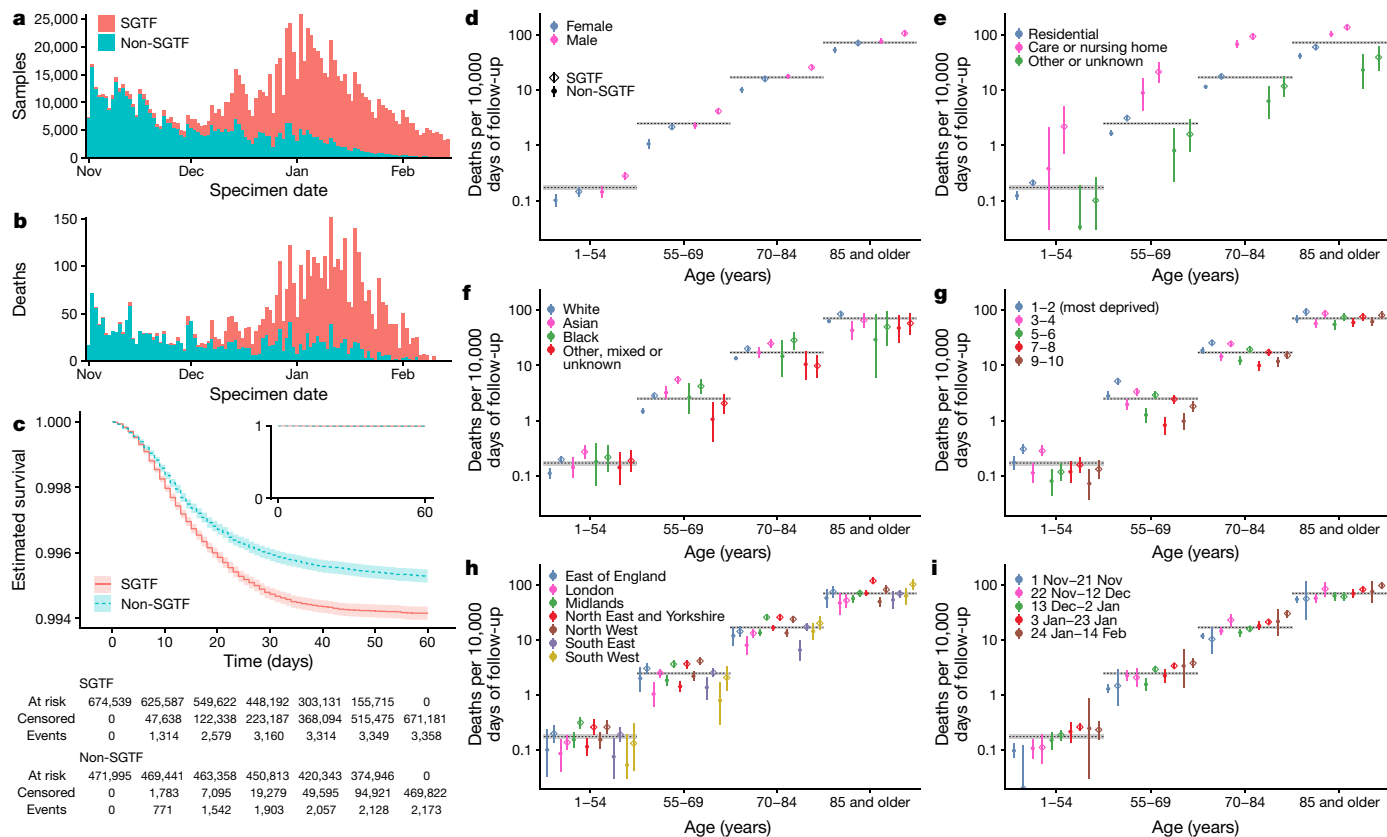
Fig. 1 | Descriptive analyses. a, The number of samples with and without SGTF by day from 1 November 2020 to 14 February 2021, the period covered by our main analysis. **b**, Number of deaths within 28 days of a positive test by specimen date for all data included in the analysis. **c**, Kaplan–Meier plot showing survival (point estimates and 95% confidence intervals) among individuals tested in the community in England with and without SGTF, in the subset for whom SGTF was measured. The inset shows the full y-axis range. **d**–**i**, Crude death rates (point estimates and 95% confidence intervals) among SGTF versus non-SGTF cases (in the subset for whom SGTF was measured; $n = 1,146,534$) for deaths within 28 days of a positive test stratified by broad age groups and sex (**d**), residence type (**e**), ethnicity (**f**), IMD decile (**g**), region of NHS England (**h**) and specimen date (**i**). Horizontal bars show the overall crude death rates (point estimates and 95% confidence intervals) by age group irrespective of SGTF status.

were observed in the East of England (77.5%), South East (77.3%) and London (75.4%) regions, and the prevalence of SGTF was the lowest in the North East and Yorkshire region (41.2%). The prevalence of SGTF increased steeply over time (Fig. 1a), from 5.8% during 1–21 November 2020 to 94.3% during 24 January–14 February 2021.

Missing SGTF status was strongly associated with age and residence type. The proportion with SGTF status missing was similar in age groups 1–34 (48.3%), 35–54 (47.8%) and 55–69 (48.2%), and then increased to 54.4% in the 70–84 age group and to 77.7% in the 85 and older age group. SGTF status was missing in 87.9% of tests for individuals living in a care or nursing home, compared with 47.4% of tests among individuals in residential accommodation. This is due in part to more extensive use of lateral flow immunoassay tests in care homes, which do not yield an SGTF reading. Missingness in SGTF status also differed substantially between regions of NHS England, ranging from 21.2% in the North West to 71.1% in the South West, which is largely explained by proximity to a Lighthouse laboratory that is capable of producing an SGTF reading (Extended Data Fig. 1). Missingness also depended on the date of the specimen, with the percentage missing being lower for the earlier specimen dates and highest (54.4%) in the 21-day period that contributed the most tests (3–23 January). There were also minor differences in missingness depending on ethnicity and IMD. Of the 48.9% of tests with missing SGTF status, 5.1% were inconclusive owing to high C_t values and the remaining 43.8% were not assessed for SGTF.

In total, 19,615 people in the study sample are known to have died (0.87% of 2,245,263). Crude death rates were substantially higher in the elderly and in those living in a care or nursing home (Supplementary

Table 1). The standard definition of a death associated with COVID-19 in England is any death that occurred within 28 days of the first positive SARS-CoV-2 test of an individual; 17,452 of the observed deaths (89.0%) met this criterion (Fig. 1b). Among those with known SGTF status, the crude death rate associated with COVID-19 was 1.86 deaths per 10,000 person-days of follow-up in the SGTF group, versus 1.42 deaths per 10,000 person-days in the non-SGTF group (Fig. 1c and Extended Data Table 2). Stratifying by broad age groups and by sex, residence type, ethnicity, IMD, region and specimen date, death rates within 28 days of a positive SARS-CoV-2 test were higher among SGTF than non-SGTF cases in 98 of the 104 strata assessed (94%; Fig. 1d–i and Supplementary Table 2).

Cox regression analyses

To estimate the association between SGTF and mortality while controlling for observed confounding (Extended Data Fig. 2), we fitted a series of Cox proportional hazards models[4] to the data. We stratified the baseline mortality hazard by lower-tier local authority (LTLA) and specimen date to control for geographical and temporal differences in the hazard—for example, due to changes in hospital pressure during the study period—and used spline terms for age and IMD and fixed effects for sex, ethnicity and residence type in the hazard model. All models were fitted twice, once using complete cases only, that is, by simply excluding individuals with missing SGTF status, and once using inverse probability weighting (IPW), that is, accounting for missingness by upweighting individuals whose characteristics—age, sex,
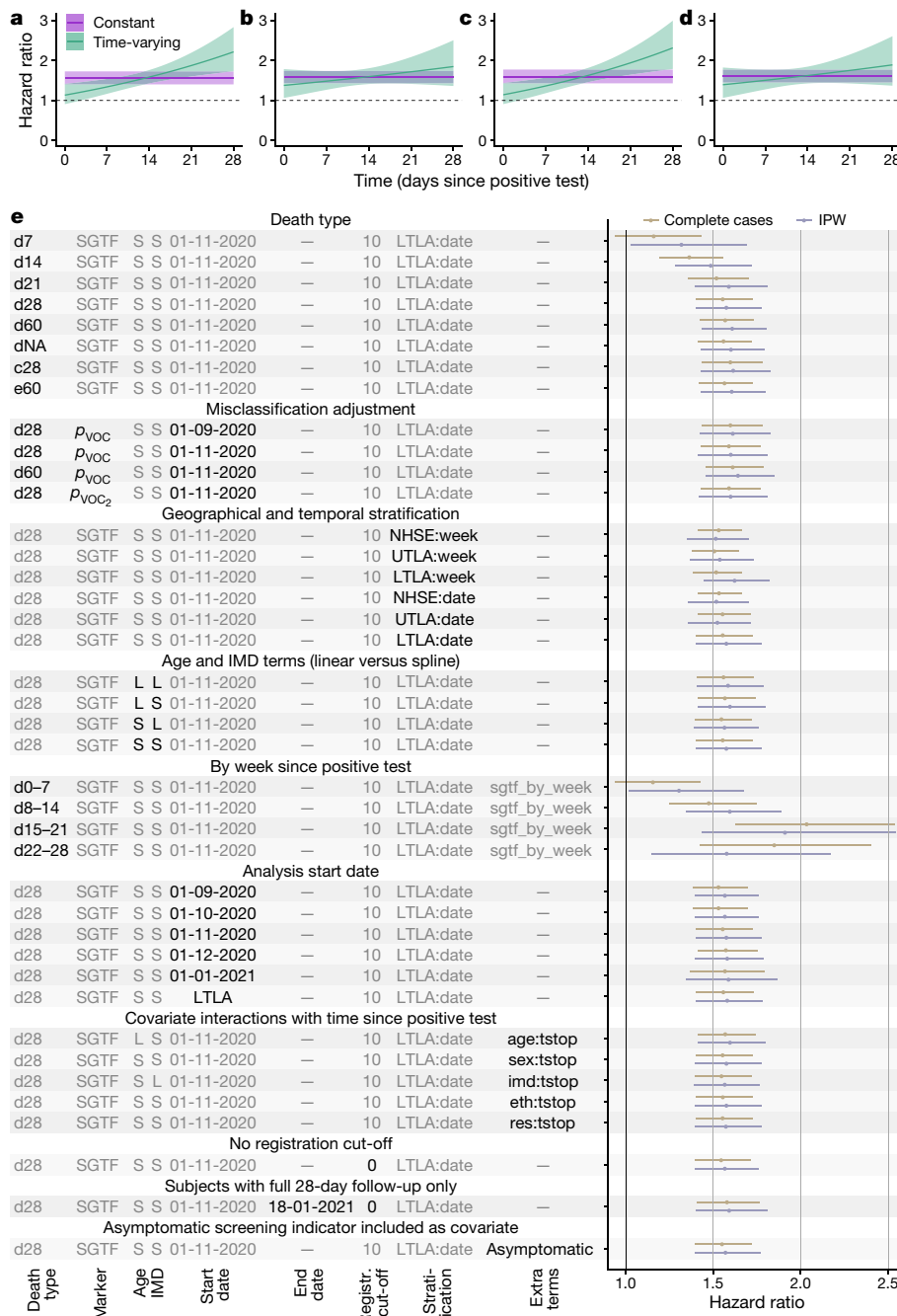
Fig. 2 | Survival analyses. a–d, Estimated hazard ratio of death (point estimate and 95% confidence intervals) within 28 days of a positive test for the SGTF analysis for complete cases (**a**), SGTF analysis with IPW (**b**), p_{VOC} analysis for complete cases (**c**) and p_{VOC}, analysis with IPW (**d**) in a model stratified by LTLA and specimen date and adjusted for the other covariates. **e**, Estimated hazard ratio of death (point estimates and 95% confidence intervals) across each model investigated. Death types are coded as follows: d*X*, all deaths within *X* days of a positive test; dNA, all deaths with no restriction on follow-up time; c28, death-certificate-confirmed deaths associated with COVID-19 within 28 days; e60, all deaths within 60 days plus all death-certificate-confirmed deaths associated with COVID-19 within any time period. S, spline term (for age or IMD); L, linear term (for age or IMD); NHSE, NHS England region ($n = 7$); UTLA, upper-tier local authority ($n = 150$); LTLA, lower-tier local authority ($n = 316$). LTLA start date signifies a start date chosen separately for each LTLA; *Y*:tstop signifies an interaction term between covariate *Y* and time since positive test (eth: ethnicity, res: residence type); p_{VOC_2} signifies sequence-based misclassification adjustment (see Methods).

Death type	Marker	Age	IMD	Start date	End date	Registr. cut-off	Stratification	Extra terms
Death type								
d7	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
d14	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
d21	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
d28	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
d60	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
dNA	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
c28	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
e60	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
Misclassification adjustment								
d28	p_{VOC}	S	S	01-09-2020	—	10	LTLA:date	—
d28	p_{VOC}	S	S	01-11-2020	—	10	LTLA:date	—
d60	p_{VOC}	S	S	01-11-2020	—	10	LTLA:date	—
d28	p_{VOC_2}	S	S	01-11-2020	—	10	LTLA:date	—
Geographical and temporal stratification								
d28	SGTF	S	S	01-11-2020	—	10	NHSE:week	—
d28	SGTF	S	S	01-11-2020	—	10	UTLA:week	—
d28	SGTF	S	S	01-11-2020	—	10	LTLA:week	—
d28	SGTF	S	S	01-11-2020	—	10	NHSE:date	—
d28	SGTF	S	S	01-11-2020	—	10	UTLA:date	—
d28	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
Age and IMD terms (linear versus spline)								
d28	SGTF	L	L	01-11-2020	—	10	LTLA:date	—
d28	SGTF	L	S	01-11-2020	—	10	LTLA:date	—
d28	SGTF	S	L	01-11-2020	—	10	LTLA:date	—
d28	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
By week since positive test								
d0–7	SGTF	S	S	01-11-2020	—	10	LTLA:date	sgtf_by_week
d8–14	SGTF	S	S	01-11-2020	—	10	LTLA:date	sgtf_by_week
d15–21	SGTF	S	S	01-11-2020	—	10	LTLA:date	sgtf_by_week
d22–28	SGTF	S	S	01-11-2020	—	10	LTLA:date	sgtf_by_week
Analysis start date								
d28	SGTF	S	S	01-09-2020	—	10	LTLA:date	—
d28	SGTF	S	S	01-10-2020	—	10	LTLA:date	—
d28	SGTF	S	S	01-11-2020	—	10	LTLA:date	—
d28	SGTF	S	S	01-12-2020	—	10	LTLA:date	—
d28	SGTF	S	S	01-01-2021	—	10	LTLA:date	—
d28	SGTF	S	S	LTLA	—	10	LTLA:date	—
Covariate interactions with time since positive test								
d28	SGTF	L	S	01-11-2020	—	10	LTLA:date	age:tstop
d28	SGTF	S	S	01-11-2020	—	10	LTLA:date	sex:tstop
d28	SGTF	S	L	01-11-2020	—	10	LTLA:date	imd:tstop
d28	SGTF	S	S	01-11-2020	—	10	LTLA:date	eth:tstop
d28	SGTF	S	S	01-11-2020	—	10	LTLA:date	res:tstop
No registration cut-off								
d28	SGTF	S	S	01-11-2020	—	0	LTLA:date	—
Subjects with full 28-day follow-up only								
d28	SGTF	S	S	01-11-2020	18-01-2021	0	LTLA:date	—
Asymptomatic screening indicator included as covariate								
d28	SGTF	S	S	01-11-2020	—	10	LTLA:date	Asymptomatic

Complete cases · IPW

IMD, ethnicity, residence type, NHS England region of residence and sampling week—are underrepresented among complete cases. The analysis of the complete cases assumes that whether an individual dies is independent of whether their SGTF status is observed or missing, given the individual's other characteristics included in the survival model, whereas the IPW analysis assumes that whether an individual has SGTF is independent of whether their SGTF status is observed or missing, given the individual's other characteristics included in the model used to derive weights for IPW[5].

For the analysis of the complete cases, the estimated hazard ratio for SGTF was 1.55 (95% confidence interval, 1.39–1.72), indicating that the hazard of death in the 28 days after a positive test is 55% (39–72%) higher for SGTF than for non-SGTF cases.

To assess the model assumption of proportional hazards, we added an interaction term between SGTF and time since a positive test. There was strong evidence of non-proportionality of hazards (likelihood ratio test $P(\chi_1^2 = 11) = 0.009$) (Fig. 2a and Extended Data Fig. 3), with the estimated time-varying hazard ratio increasing over time: 1.14 (0.92–1.40) on day 1 after the positive test, 1.58 (1.42–1.75) on day 14 and 2.24 (1.75–2.87) on day 28. Adding higher-order functions of time into the interaction terms did not significantly improve the fit of the model (likelihood ratio test $P(\chi_1^2 = 3.3) = 0.07$). We found no evidence that the effect of SGTF varied depending on age group (likelihood ratio test $P(\chi_4^2 = 5.8) = 0.22$), sex ($P(\chi_1^2 = 0.057) = 0.81$), IMD ($P(\chi_9^2 = 11) = 0.31$), ethnicity ($P(\chi_3^2 = 1.2) = 0.75$) or residence type ($P(\chi_2^2 = 0.33) = 0.85$). We note, however, that the relatively small number of deaths among 1–34-year-old individuals during the study period (44 deaths) does not permit robust assessment of the effect of SGTF in this age group. Other time-covariate interactions suggested that the time from positive test to death was slightly shorter among women, care home residents and elderly individuals; see Supplementary Note 1 for more details on models with interaction terms.

Table 1 | Absolute 28-day mortality risk for B.1.1.7

Sex	Age (years)	Baseline mortality	Mortality risk for B.1.1.7	
			SGTF for complete cases	p_{VOC} IPW
Female	1–34	0.00069%	0.0011% (0.00096–0.0012%)	0.0011% (0.00097–0.0012%)
	35–54	0.033%	0.050% (0.045–0.056%)	0.052% (0.046–0.059%)
	55–69	0.18%	0.28% (0.25–0.31%)	0.29% (0.26–0.33%)
	70–84	2.9%	4.4% (4.0–4.9%)	4.6% (4.0–5.1%)
	85 and older	13%	19% (17–21%)	20% (18–22%)
Male	1–34	0.0031%	0.0047% (0.0042–0.0052%)	0.0049% (0.0043–0.0055%)
	35–54	0.064%	0.099% (0.089–0.11%)	0.10% (0.090–0.12%)
	55–69	0.56%	0.86% (0.77–0.95%)	0.89% (0.78–1.0%)
	70–84	4.7%	7.2% (6.4–7.9%)	7.4% (6.6–8.3%)
	85 and older	17%	25% (23–27%)	26% (23–29%)

The absolute mortality risk is expressed as the case fatality ratio (%) among individuals testing positive in the community. The baseline risk (that is, for pre-existing SARS-CoV-2 variants) is derived using linked deaths within 28 days for all individuals testing positive in the community from 1 August to 31 October 2020. Adjusted risks are presented for the SGTF analysis for complete cases and for the misclassification-adjusted (p_{VOC}) IPW analysis, which yielded the lowest and highest mortality estimates, respectively, of the main models assessed (Fig. 2a–d).

For IPW analysis, a model to predict missingness is required. We evaluated a series of such models, including a cauchit model, which is a robust alternative to logistic regression that is suitable for IPW[5]. We selected the cauchit model as it fit well and resulted in less extreme weights than other models (Extended Data Fig. 4). The IPW analysis gave similar results to the analysis of the complete cases, yielding a hazard ratio of 1.58 (1.40–1.78). Similar to the analysis of the complete cases, the IPW analysis recovered an increasing hazard ratio with time since a positive test, but the increase was less marked (Fig. 2b) and did not significantly differ from zero (Wald test $P(\chi^2_1 = 1.4) = 0.23$).

Misclassification analysis

Before the emergence of B.1.1.7, a number of minor circulating SARS-CoV-2 lineages with mutations in the S gene could also cause SGTF[1]. Our main analyses are restricted to specimens from 1 November 2020 onwards to avoid diluting the measured effect of B.1.1.7 on mortality due to non-B.1.1.7 lineages that cause SGTF. As an alternative approach, we undertook a misclassification analysis[6], modelling the relative frequency of SGTF over time for each NHS England region as a low, time-invariant frequency of non-B.1.1.7 samples with SGTF plus a logistically growing[2] frequency of B.1.1.7 samples. This allowed us to estimate the probability, p_{VOC}, that a given SGTF sample was B.1.1.7 based on its specimen date and NHS England region (Extended Data Fig. 5). Again restricting the analysis to specimens from 1 November 2020 onward, we find a hazard ratio associated with p_{VOC} of 1.58 (1.42–1.76) for the analysis of the complete cases and 1.61 (1.42–1.82) for the IPW analysis (Fig. 2c, d).

Absolute risks

To put these results into context, we estimated absolute mortality risks by applying hazard ratios for SGTF to the baseline risk of death among individuals tested in the community between August and October 2020 (assumed to be illustrative of the case fatality ratio associated with pre-existing variants of SARS-CoV-2) (Table 1). For the analysis of the complete cases, in women aged 70–84 years, the estimated risk of

death within 28 days of a positive SARS-CoV-2 test increases from 2.9% without SGTF to 4.4% with SGTF (95% confidence interval, 4.0–4.9%) and for women 85 years or older, the risk increases from 13% to 19% (17–21%). For men aged 70–84 years, the risk of death within 28 days increases from 4.7% to 7.2% (6.4–7.9%) and for men 85 years or older, the risk increases from 17% to 25% (23–27%). Estimates based on the IPW analysis corrected for misclassification were marginally higher. These estimates reflect a substantial increase in absolute risk among older age groups, but the risk of death associated with COVID-19 after a positive test in the community remains below 1% in most individuals who are younger than 70 years old. Note that these estimates capture the fatality ratio among people tested in the community, and are thus likely to be higher than the infection fatality ratio, as many individuals with a SARS-CoV-2 infection are never tested.

Further investigations

We conducted a number of sensitivity analyses to verify the robustness of our results. Our main results were largely insensitive to: restriction of the analysis to deaths caused by COVID-19 confirmed on the death certificate; any follow-up time of 21 days or longer; coarseness of geographical and temporal stratification; use of linear versus spline terms for age and IMD; analysis start date; follow-up time–covariate interactions; removal of the 10-day death registration cut-off; and restriction of the analysis to individuals with a full 28-day follow-up period (Fig. 2e). Generally, the IPW analysis yielded marginally higher hazard ratios, with greater uncertainty. As a further sensitivity analysis, we adjusted for an indicator in community testing data for whether the individual was tested because of symptoms or owing to asymptomatic screening. Although we caution that symptomatic screening status may lie on the causal pathway between SGTF status and death, we found that this adjustment had no effect on the relative hazard of SGTF (1.54 (1.39–1.71); analysis of complete cases).

Discussion

We previously found that B.1.1.7 is substantially more transmissible than pre-existing SARS-CoV-2 variants, but could not robustly identify any associated change in disease severity using population-level analysis of early data[2]. This analysis of individual-level data, which controls for factors that could confound the association between B.1.1.7 infection and death, reveals an increase in COVID-19 mortality associated with the B.1.1.7 lineage. We stratify our analyses by test time and geographical location—mimicking matching on these variables—to account for changes in testing rates and changing pressures on hospital services over time and by region. Our findings are consistent with earlier reports by ourselves and other groups[7] and with contemporaneous studies[8–11] assessing the risk of severe outcomes among individuals with B.1.1.7 infection. Notably, our study is limited to individuals tested in the community. Indicators for infection with the B.1.1.7 variant are not currently available for most people who die from COVID-19 in England, as they are tested in the hospital rather than in the community and hospitals do not routinely collect genotypic data. However, this restricted focus allows us to capture the combined effect of an altered risk of hospitalization given a positive test and an altered risk of death given hospitalization, while only the latter would be measurable in a study of hospitalized patients only. Unfortunately, we were unable to account for vaccination status in this analysis.

We do not identify any mechanism for the increased mortality here. Infections with the B.1.1.7 variant are associated with higher viral concentrations in nasopharyngeal swabs, as measured by C_t values using PCR testing (Extended Data Fig. 6). Higher viral load could therefore be partly responsible for the observed increase in mortality; this could be assessed using a mediation analysis. Alternatively, changes in test-seeking behaviour could, in principle, explain our results. If

Article

B.1.1.7-associated infections were less likely to cause symptoms, but symptomatic cases of B.1.1.7 were more severe, then our study could overestimate changes in the infection fatality rate. However, we find no clear difference in SGTF frequency among community tests relative to a random sample of SARS-CoV-2 infections in the population (Extended Data Fig. 7), which suggests that variant-associated changes in test-seeking propensity do not explain our findings.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03426-1.

1. Chand, M. et al. *Investigation of Novel SARS-COV-2 Variant: Variant of Concern 202012/01*. Technical Briefing 1 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf (Public Health England, 2020).
2. Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* https://doi.org/10.1126/science.abg3055 (2021).
3. McLennan, D. et al. *English Indices of Deprivation 2019*. https://www.gov.uk/government/publications/english-indices-of-deprivation-2019-technical-report (Ministry of Housing, Communities & Local Government 2019).
4. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc. B* **34**, 187–202 (1972).
5. Seaman, S. R. & White, I. R. Review of inverse probability weighting for dealing with missing data. *Stat. Methods Med. Res.* **22**, 278–295 (2013).
6. Keogh, R. H. et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: part 1—basic theory and simple methods of adjustment. *Stat. Med.* **39**, 2197–2231 (2020).
7. Scientific Advisory Group for Emergencies. *NERVTAG: Update Note on B.1.1.7 Severity, 11 February 2021*. https://www.gov.uk/government/publications/nervtag-update-note-on-b117-severity-11-february-2021 (2021).
8. Challen, R. et al. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ* **372**, n579 (2021).
9. Grint, D. J. et al. Case fatality risk of the SARS-CoV-2 variant of concern B.1.1.7 in England. *Euro Surveill.* **26**, 21002 (2021).
10. Patone, M. et al. Analysis of severe outcomes associated with the SARS-CoV-2 variant of concern 202012/01 in England using ICNARC case mix programme and QResearch databases. Preprint at https://doi.org/10.1101/2021.03.11.21253364 (2021).
11. Bager, P. et al. Increased risk of hospitalisation associated with infection with SARS-CoV-2 lineage B.1.1.7 in Denmark. Preprint at https://doi.org/10.2139/ssrn.3792894 (2021).

CMMID COVID-19 Working Group

Kevin van Zandvoort[1], Samuel Clifford[1], Fiona Yueqian Sun[1], Sebastian Funk[1], Graham Medley[1], Yalda Jafari[1], Sophie R. Meakin[1], Rachel Lowe[1], Matthew Quaife[1], Naomi R. Waterlow[1], Rosalind M. Eggo[1], Jiayao Lei[1], Mihaly Koltai[1], Fabienne Krauer[1], Damien C. Tully[1], James D. Munday[1], Alicia Showering[1], Anna M. Foss[1], Kiesha Prem[1], Stefan Flasche[1], Adam J. Kucharski[1], Sam Abbott[1], Billy J. Quilty[1], Thibaut Jombart[1], Alicia Rosello[1], Gwenan M. Knight[1], Mark Jit[1], Yang Liu[1], Jack Williams[1], Joel Hellewell[1], Kathleen O'Reilly[1], Yung-Wai Desmond Chan[1], Timothy W. Russell[1], Simon R. Procter[1], Akira Endo[1], Emily S. Nightingale[1], Nikos I. Bosse[1], C. Julian Villabona-Arenas[1], Frank G. Sandmann[1], Amy Gimma[1], Kaja Abbas[1], William Waites[1], Katherine E. Atkins[1,4], Rosanna C. Barnard[1], Petra Klepac[1], Hamish P. Gibbs[1], Carl A. B. Pearson[1] & Oliver Brady[1]

[4]Centre for Global Health, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK.

Methods

Ethical approval
Approved by the Observational/Interventions Research Ethics Committee at the London School of Hygiene and Tropical Medicine (reference number 24020). Participant consent is not required for national infectious disease notification datasets in England.

Data sources
We linked three datasets provided by Public Health England: a line list of all positive tests in Pillar 2 (community) testing for SARS-CoV-2 for England, containing specimen date and demographic information on the participants; a line list of cycle threshold (C_t) values for the *orf1ab*, *N* (nucleocapsid), and *S* (spike) genes for positive tests that were processed in one of the three national laboratories (Alderley Park, Glasgow or Milton Keynes) using the Thermo Fisher TaqPath COVID-19 assay; and a line list of all deaths associated with COVID-19 in England, which combines and deduplicates deaths reported by hospitals in England, by the Office for National Statistics, through direct reporting from Public Health England Health Protection Teams, and through Demographic Batch Service tracing of laboratory-confirmed cases[12]. We link these datasets using a numeric identifier for Pillar 2 tests ('FINALID') common to all three datasets. We define SGTF as any test with $C_t < 30$ for *orf1ab* and *N* targets but no detectable *S* gene, and non-SGTF as any test with $C_t < 30$ for *orf1ab*, *N* and *S* targets. A small proportion (10.4%) of SGTF tests are inconclusive. The study population of interest is defined as all individuals who received a positive Pillar 2 test between 1 November 2020 and 14 February 2021. For our main analysis, we included only tests from after 1 November 2020 to avoid including an excess of tests with SGTF not resulting from infection by lineage B.1.1.7. In sensitivity analyses, we also consider extending the population to include tests performed between 1 September and 31 October 2020.

Our analysis does not include individuals who first tested positive in hospital—that is, those patients who presented to the hospital after the onset of symptoms without first being tested in the community. This is because the cycle threshold values used to ascertain SGTF status are not available for individuals who were not tested in the community. Of the 57,750 deaths associated with COVID-19 in England during the study period, 17,642 deaths (30.5%) can be linked to a positive Pillar 2 test; among these, 4,945 have non-missing SGTF status. So, although our study includes 1,098,729 Pillar 2 tests with non-missing SGTF status, which represents 51.1% of the 2,245,263 Pillar 2 tests over this period and 40.2% of the 2,736,806 combined Pillar 1 (hospital) and Pillar 2 (community) SARS-CoV-2 tests over this period, we can only assess SGTF status for 8.6% (4,945/57,750) of the individuals who died from COVID-19 over the study period. This is explained by differing mortality rates among individuals who first test positive in a hospital compared to those who are tested in the community, as the former group are much more likely to have a severe illness, as well as by missingness in the SGTF data.

There was a small amount of missing data for sex ($n = 14$, <0.01%), age ($n = 171$, <0.01%), and IMD and regional covariates ($n = 3,817$, 0.16%). There were no missing specimen dates. Individuals with missing age, sex or geographical location were excluded. We also excluded individuals from the dataset whose age was recorded as zero, as there were 17,913 age-0 individuals compared to 10,132 age-1 individuals in the dataset, suggesting that many of these age-0 individuals may have been miscoded. There were some missing data on ethnicity ($n = 47,491$, 2.1%) and we created a category that combines missing values with 'other' and 'mixed'. Missing values for residence type ($n = 63,905$, 2.8%) were also combined with an 'other' category. The full dataset used for the main analysis comprises 2,245,263 individuals, with SGTF status missing or inconclusive for 1,098,729 (48.9%). Missing data on the exposure is addressed in the analysis, described below.

We grouped residence types into three categories: residential, which included the 'residential dwelling (including houses, flats and sheltered accommodation)' and 'house in multiple occupancy' groups; care or nursing home; and other or unknown, which included the 'medical facilities (including hospitals and hospices, and mental health)', 'no fixed abode', 'other property classifications', 'overseas address', 'prisons, detention centres and secure units', 'residential institution (including residential education)' and 'undetermined' groups, as well as an unspecified residence type. We grouped ethnicities into four categories according to the broad categories used in the 2011 UK Census: Asian, which included the 'Bangladeshi (Asian or Asian British)', 'Chinese (other ethnic group)', 'Indian (Asian or Asian British)', 'Pakistani (Asian or Asian British)' and 'any other Asian background' groups; Black, which included the 'African (Black or Black British)', 'Caribbean (Black or Black British)' and 'any other Black background' groups; white, which included the 'British (white)', 'Irish (white)' and 'any other white background' groups; and 'other, mixed or unknown', which included the 'any other ethnic group', 'white and Asian (mixed)', 'white and Black African (mixed)', 'white and Black Caribbean (mixed)', 'any other mixed background' and 'Unknown' groups.

Statistical methods
There are several factors that we expect are associated with both SGTF and with risk of death, thus confounding the association between SGTF and risk of death in those individuals who were tested. Area of residence and specimen date were expected to be potentially strong confounding factors. Area of residence is expected to be strongly associated with SGTF status due to different virus variants circulating in different areas, and specimen date because the prevalence of SGTF is known to have greatly increased over time. Area of residence and specimen date are also expected to be associated with risk of death after a positive test, including due to differences associated with differential pressure on hospital resources by area and time. The following variables were also identified as potential confounding factors: sex, age, residence type (residential, care or nursing home, or other or unknown), ethnicity (white, Asian, Black or other, mixed or unknown) and IMD. The potential confounding factors are referred to collectively as the covariates. For descriptive analyses, age (in years) was categorized as 1–34, 35–54, 55–69, 70–84, or 85 and older.

Descriptive analyses were performed. We tabulated the distribution of the covariates in the whole study sample, the association between each covariate and SGTF status in the subset for whom SGTF was measured, and the association between each covariate and missing data in SGTF status (Extended Data Table 1). The subset for whom SGTF status was measured are referred to as the complete cases. The unadjusted association between SGTF and mortality in the complete cases was assessed using a Kaplan–Meier plot (Fig. 1c), and Kaplan–Meier plots and crude 28-day mortality rates are also presented separately according to the categories of the covariates (Extended Data Table 2 and Extended Data Fig. 2). Crude overall mortality rates (that is, not restricted to 28 days after a positive test) were obtained for the whole sample, by SGTF status in the complete cases, and in those with missing SGTF status, according to the categories of each covariate (Supplementary Table 1). We also obtained mortality rates by SGTF status (in the complete cases) for the categories of each covariate stratified by age group (Fig. 1d–i). Exact Poisson confidence intervals are used for mortality rates, assuming constant rates.

Approximately 49% of individuals in the study sample are missing data on SGTF status, due to their test not having been processed at one of the three laboratories using the Thermo Fisher TaqPath COVID-19 assay or the test being inconclusive. We performed analysis on the complete cases, restricted to the subset for whom SGTF status was measured and conclusive. This analysis of complete cases assumes that for each analysis, the missing data—in this case missing SGTF status—are independent of the outcome of interest given the variables included in the models. This is a specific type of 'missing not at random' assumption, as in particular it is allowed to depend on the underlying value of

Article

SGTF. We also performed an analysis of the complete cases using inverse probability weights[5] (IPW) to address the missing data on SGTF, under a 'missing at random' assumption. In the analysis, each individual with SGTF status measured is weighted by the inverse of their probability of having SGTF status measured based on their covariates. For the IPW, the missingness model estimated the probability of missingness using logistic regression with age (restricted cubic spline), sex, IMD decile (restricted cubic spline), ethnicity, residence type by asymptomatic screening indicator and NHS region by specimen week as predictors. We also considered a cauchit and a Gosset link for the missingness model, including the same predictors, as this was expected to provide better stability for the weights[5]. The fit of the missingness model was assessed using a Q–Q plot (Extended Data Fig. 4), and Hosmer–Lemeshow and Hinkley tests were used to choose the most appropriate model.

Cox regression[4] was used to estimate the association between SGTF and the hazard of mortality, conditioning on the potential confounding factors listed above. The analyses described here were applied to the complete cases and using IPW. For IPW analyses, the standard errors accounted for the weights, although the fact that the weights were estimated was not accounted for; this results in conservative standard errors. The baseline hazard in the Cox model was stratified by both specimen date and LTLA, therefore finely controlling for these variables. The stratification gives a large number of strata matched by specimen date and LTLA. Only those strata that contain individuals who died and individuals who survived contribute to the analysis. The analysis is therefore similar to the analysis that would be performed had we created a matched nested case–control sample. The remaining variables were included as covariates in the model (sex, age, residence type, ethnicity and IMD decile). Age was included as a restricted cubic spline with five knots, and IMD decile was included as a restricted cubic spline with three knots. The time origin for the analysis was specimen date and we considered deaths up to 28 days after the specimen date for the main analyses. Individuals who did not die within 28 days were censored at the earlier of 28 days after the specimen date and the administrative censoring date, which we chose as the date of the most recent death linkable to SGTF status minus 10 days (that is, 14 February 2021) to minimize any potential bias due to late reporting of deaths. We began by assuming the proportionality of the hazards for SGTF and the covariates included in the model. The assumption of proportional hazards was assessed by including in the model an interaction between each covariate and time, which was performed separately for SGTF and for each other covariate. Schoenfeld residual plots were also obtained for each covariate (Extended Data Fig. 3). We assessed whether the association between SGTF and the hazard was modified by age, sex, IMD, ethnicity and residence type. Models with and without interactions were compared using likelihood ratio tests for the analyses of the complete cases. For the analysis using IPW, we used Wald tests based on robust standard errors[13].

The analysis assumes that censoring is uninformative, which is plausible as all censoring is administrative.

Misclassification analysis

The exposure of SGTF is subject to misclassification, because a number of minor circulating SARS-CoV-2 lineages in addition to B.1.1.7 are also associated with failure to amplify the S gene target. Accordingly, a positive test with SGTF is not necessarily indicative of infection with B.1.1.7. A negative test of SGTF is assumed to be indicative of an absence of infection with B.1.1.7. Misclassification of an exposure can result in bias in its estimated association with the outcome. We fitted a logistic model to Pillar 2 SGTF frequencies by NHS region to estimate a 'background' rate of SGTF in the absence of B.1.1.7, assuming a beta-binomial prior. This model is then used to estimate the probability that an individual testing positive for SGTF is infected with B.1.1.7, separately for individuals in each NHS region. These probabilities can then be used in place of the indicator of SGTF exposure in the Cox models. This is the regression

calibration approach[6] to correcting for bias due to measurement error in an exposure.

We fitted models accounting for false-positive results (modelled as regionally varying background rates of SGTF associated with non-B.1.1.7 variants) to the SGTF data. Our logistic model for B.1.1.7 growth over time is as follows:

$$\text{logit}\,(f(t)) = \text{slope} \times (t - \text{intercept})$$

$$s(t) = f(t) + (1 - f(t)) \times \text{FP}$$

$$k_t \sim \text{betaBinomial}(n = n_t,$$
$$\alpha = s(t) \times (\text{conc} - 2) + 1,$$
$$\beta = (1 - s(t)) \times (\text{conc} - 2) + 1)$$

$$\text{slope} \sim \text{normal}(\mu = 0, \sigma = 1)$$

$$\text{intercept} \sim \text{normal}(\mu = 0, \sigma = 1,000)$$

$$\text{FP} \sim \text{beta}(\alpha = 1.5, \beta = 15)$$

$$\text{conc} \sim \text{normal}(\mu = 0, \sigma = 500) \geq 2$$

where $f(t)$ is the predicted frequency of B.1.1.7 among positive tests at time t (in days since 1 September 2020) based on the terms slope and intercept; $s(t)$ is the predicted frequency of SGTF at time t due to the combination of B.1.1.7 and a background false-positive rate (FP) among non-B.1.1.7 variants, conc is the 'concentration' parameter (conc = $\alpha + \beta$) of a beta distribution with mode $s(t)$; k_t is the number of SGTFs detected at time t; n_t is the total number of tests at time t; and the tilde (~) signifies 'distributed as'. All priors above are chosen to be vague, and the truncation of the concentration parameter to values greater than 2 ensures a unimodal distribution for the proportion of tests that are SGTF. The model above is fitted separately for each NHS England region. Then, p_{VOC} for a test with SGTF = 1 at time t is equal to $f(t)/s(t)$, and $p_{\text{VOC}} = 0$ for all tests with SGTF = 0. The model was fitted using Markov chain Monte Carlo with 10,000 iterations of burn-in and 5,000 iterations of sampling.

The model above was fitted using the same data source (that is, SGTF frequencies among Pillar 2 community tests for SARS-CoV-2) as our survival analysis. To verify the robustness of this model, we performed a sensitivity analysis using sequencing data from the COVID-19 UK Genomics Consortium[14] (https://www.cogconsortium.uk/) downloaded from the Microreact platform[15] (https://microreact.org/) on 11 January 2020 to estimate p_{VOC}. In this alternative analysis, we estimated p_{VOC} for each NHS England region and date as the number of samples that were VOC 202012/01 (that is, lineage B.1.1.7 with mutations Δ69/Δ70 and N501Y in spike) divided by the number of samples that were SGTF (that is, any lineage with Δ69/Δ70, the deletion that causes SGTF) for that NHS England region and date, setting $p_{\text{VOC}} = 1$ for all dates later than 31 December 2020 as there were no sequencing data available past this date, and filling any gaps in the data using linear interpolation. This yielded nearly identical results in our survival analysis compared with the analysis that uses the modelled p_{VOC} described above (Fig. 2e).

Absolute risks

Estimates from the final Cox models were used to obtain estimates of absolute risk of death within 28 days of a positive test for SGTF and p_{VOC}. Given the strong influence of age on risk of death, we present absolute risks by sex and age group (in years; 1–34, 35–54, 55–69, 70–84, 85 and older). Absolute risks of death (case fatality rate) within 28 days were estimated by age group and sex using data on individuals tested during August–October 2020; this is referred to as the baseline risk. The absolute risks of death for individuals with SGTF were then estimated as follows.

If the baseline absolute risk of death in a given age group is $1 - A$, then the estimated absolute risk of death with SGTF is $1 - A^{HR}$, where HR denotes the estimated hazard ratio obtained from the Cox model assuming proportional hazards. Standard errors are obtained via the delta method, and confidence intervals were based on normal approximations.

Sensitivity analyses

Several sensitivity analyses were performed. After establishing the final model using the process outlined above, we investigated the effect of using different variables for the stratification of the baseline hazard measuring region at a coarser level (the upper-tier local authority or NHS England region), as well as coarser test specimen time (week rather than exact date). Adjusting for these variables instead of using stratification was also explored. We also repeated the main analysis restricting data to specimens collected from September onwards, October onwards, November onwards or December onwards.

To assess the effect of imposing an administrative cut-off to follow-up time of 10 days before data extraction, we first reanalysed the data without this cut-off, as well as reanalysing the data restricting the analysis to individuals with a follow-up of at least 28 days.

Finally, we adjusted for symptomatic status associated with the test (asymptomatic versus symptomatic), which relates to whether the test was given for asymptomatic screening purposes or on the basis of a request by a (presumed symptomatic) individual, as only symptomatic individuals may request a community SARS-CoV-2 test in England.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

An anonymized dataset that enables the replication of the analysis is available at https://zenodo.org/record/4579857. Source data are provided with this paper.

Code availability

Analysis code deposited at time of publication is available at https://zenodo.org/record/4579857. The repository is maintained at https://github.com/nicholasdavies/cfrvoc.

12. Public Health England COVID-19 Epidemiology Cell. COVID-19 confirmed deaths in England: report. https://www.gov.uk/government/publications/covid-19-reported-sars-cov-2-deaths-in-england/covid-19-confirmed-deaths-in-england-report (accessed 20 February 2021).
13. Lumley, T. & Scott, A. Partial likelihood ratio tests for the Cox model under complex sampling. *Stat. Med.* **32**, 110–123 (2013).
14. The COVID-19 Genomics UK (COG-UK) Consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* **1**, E99–E100 (2020).
15. Argimón, S. et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microbial Genomics* **2**, 93 (2016).

Extended Data Fig. 1 | Missingness in SGTF status and proximity to SGTF-capable Lighthouse laboratories. The geographical location of the six Lighthouse laboratories in the UK; missingness is higher in the lower-tier local authorities (shaded regions) that are closer to a Lighthouse laboratory that is not capable of producing an SGTF reading. Map source: Office for National Statistics.
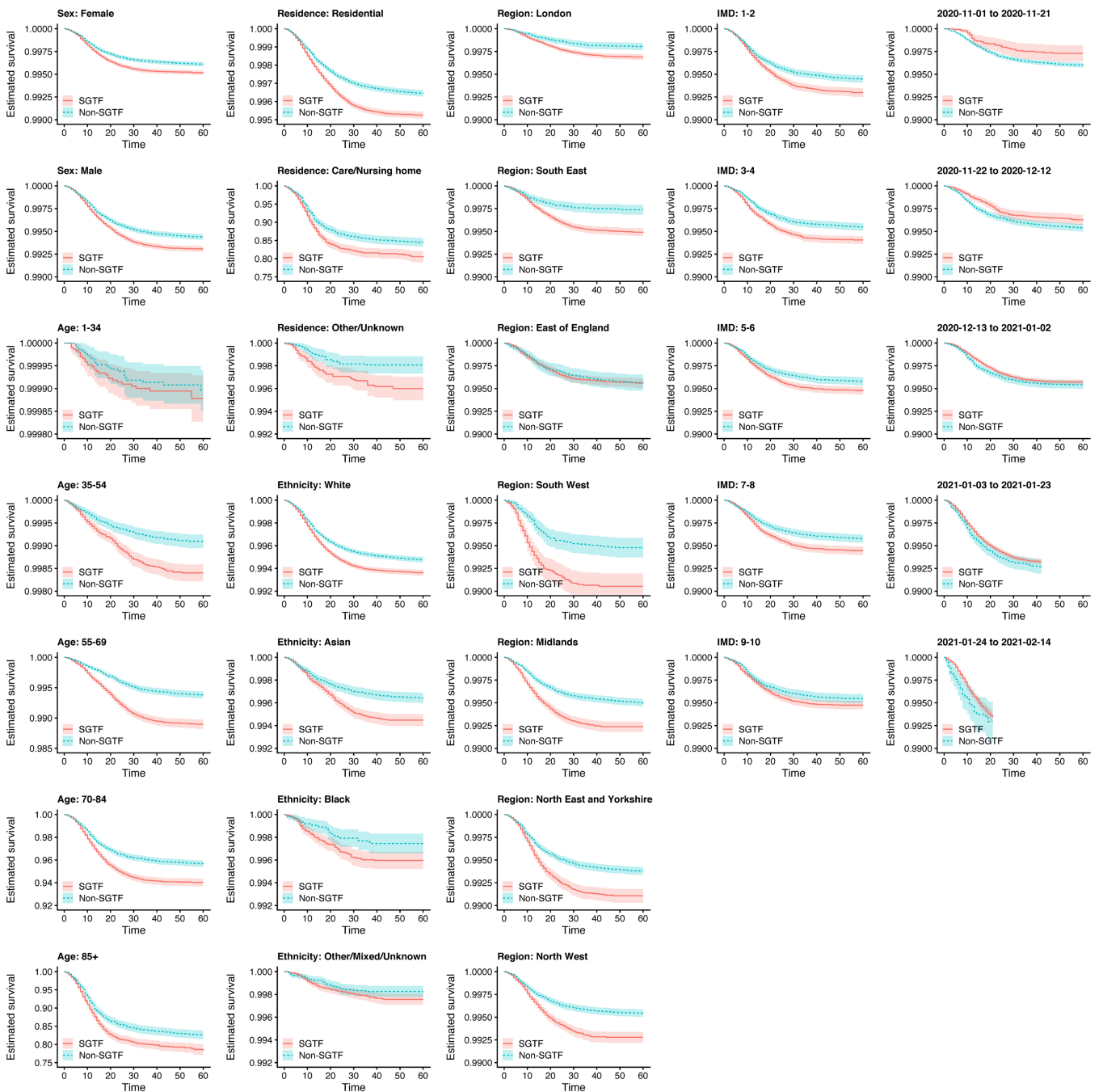
Extended Data Fig. 2 | Kaplan–Meier plots of survival within 60 days of a positive test for SGTF versus non-SGTF samples. Plots are stratified by sex, age group, place of residence, ethnicity, NHS England region, IMD decile (in five groups) and specimen date. Note that the *y*-axis ranges differ among panels.

These curves show the crude survival within each group (unadjusted for other covariates), and so do not necessarily signify differences in the effect of SGTF on survival for any specific group due to possible confounding factors. Shaded areas show 95% confidence intervals.
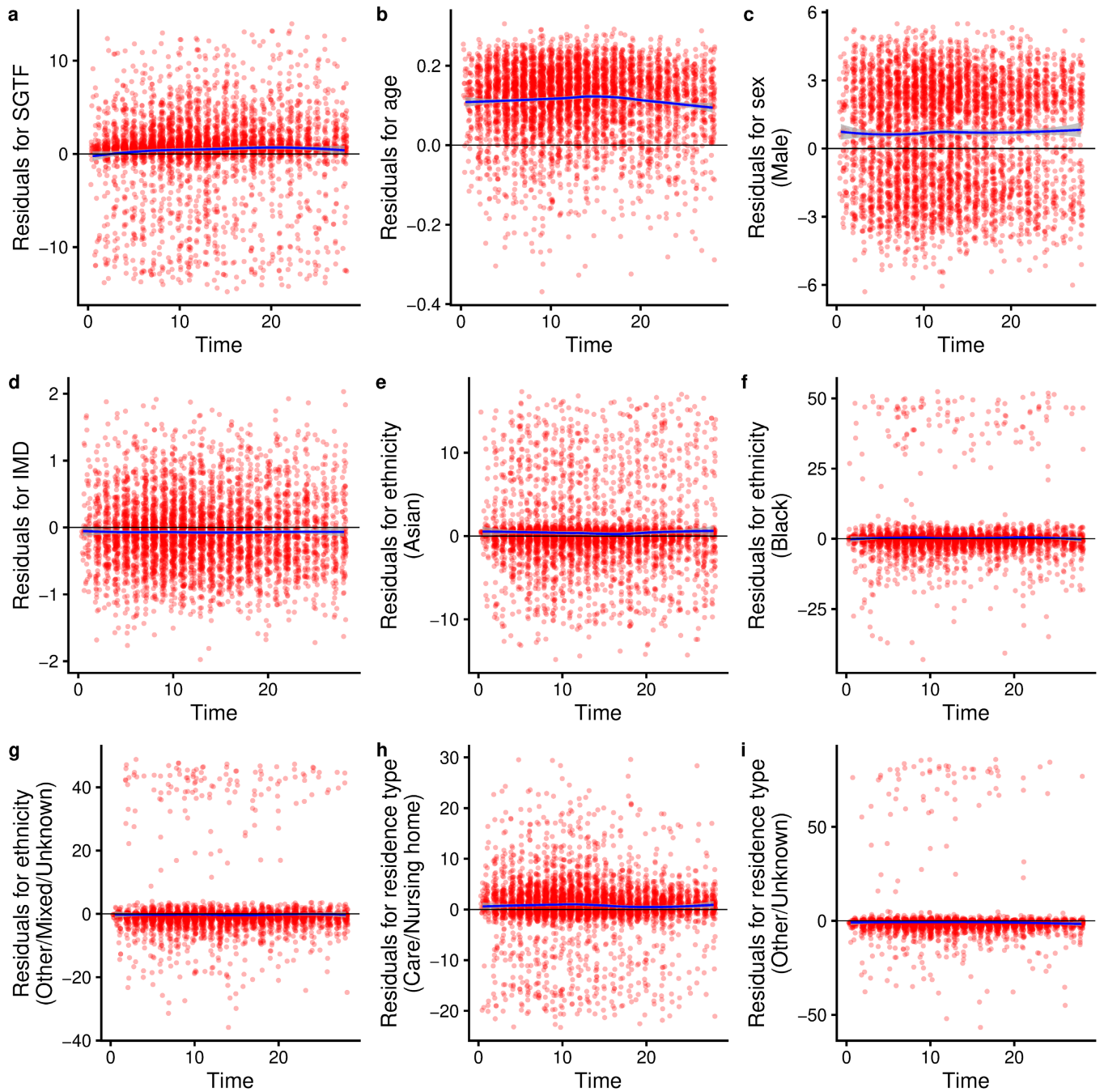
Extended Data Fig. 3 | Schoenfeld residuals for survival model by SGTF stratified by LTLA and specimen date. The model uses linear terms for age and IMD and a 28-day follow-up using complete cases. **a**–**i**, Residuals for SGTF (**a**), age (**b**), sex (**c**), IMD (**d**), ethnicity (**e**–**g**), and residence type (**h**, **i**). Two-sided Schoenfeld residual tests were performed. $P = 0.001$ for SGTF (**a**); $P = 0.039$ for age (**b**); $P = 0.101$ for sex (**c**); $P = 0.937$ for IMD decile (**d**); $P = 0.969$ for ethnicity (**e**–**g**); $P = 0.064$ for residence type (**h**–**i**); and $P = 0.027$ globally. The trend line shows the mean and 95% confidence intervals of a loess regression.
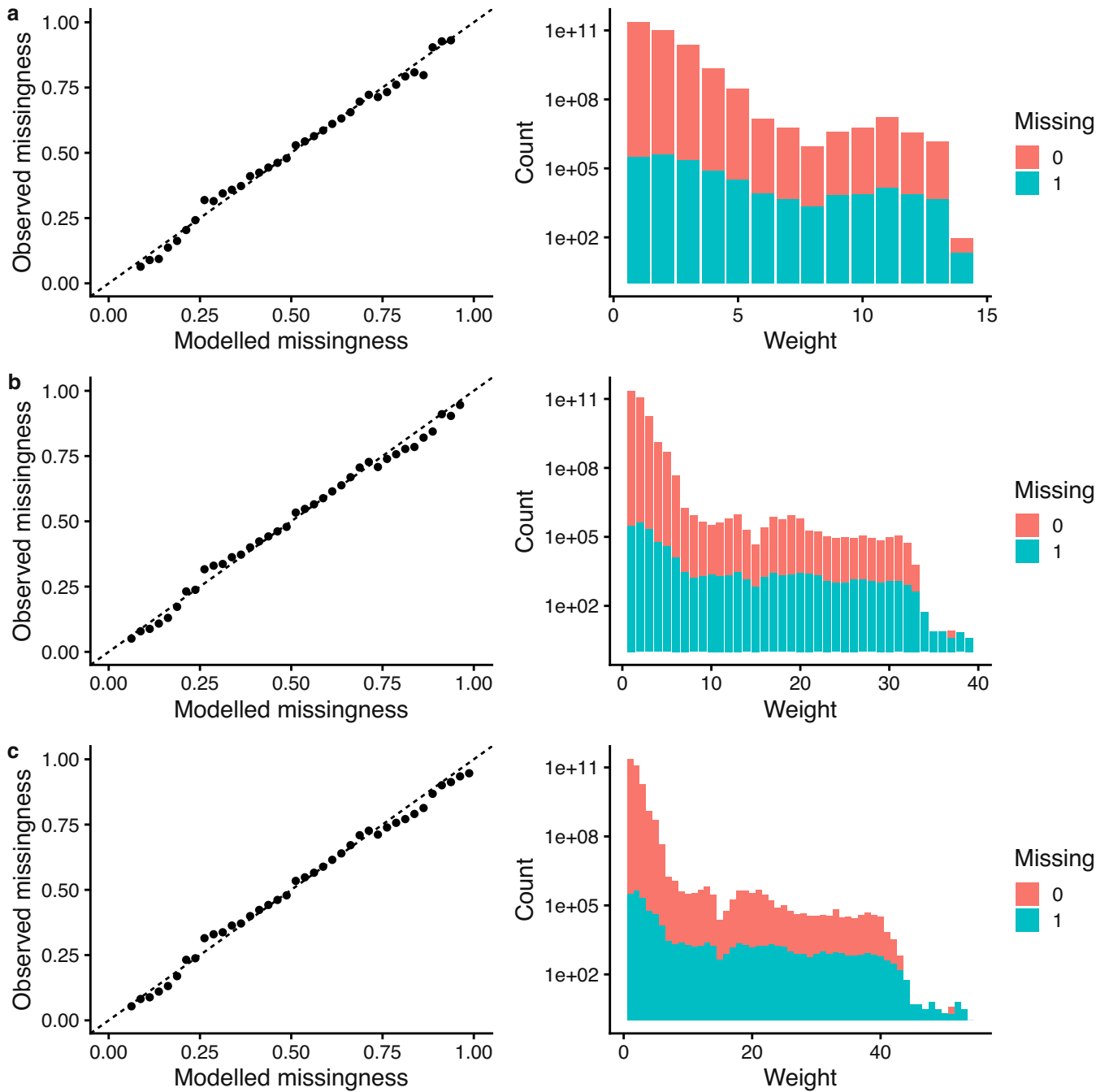
Extended Data Fig. 4 | Comparison of missingness models. a–c, Q–Q plot (left; mean and 95% confidence intervals) and distribution of weights (right) under different missingness models assessed for IPW with a cauchit link (**a**), a robit link (Student's *t*-distribution with d.f. = 4) (**b**) and a logit link (**c**).
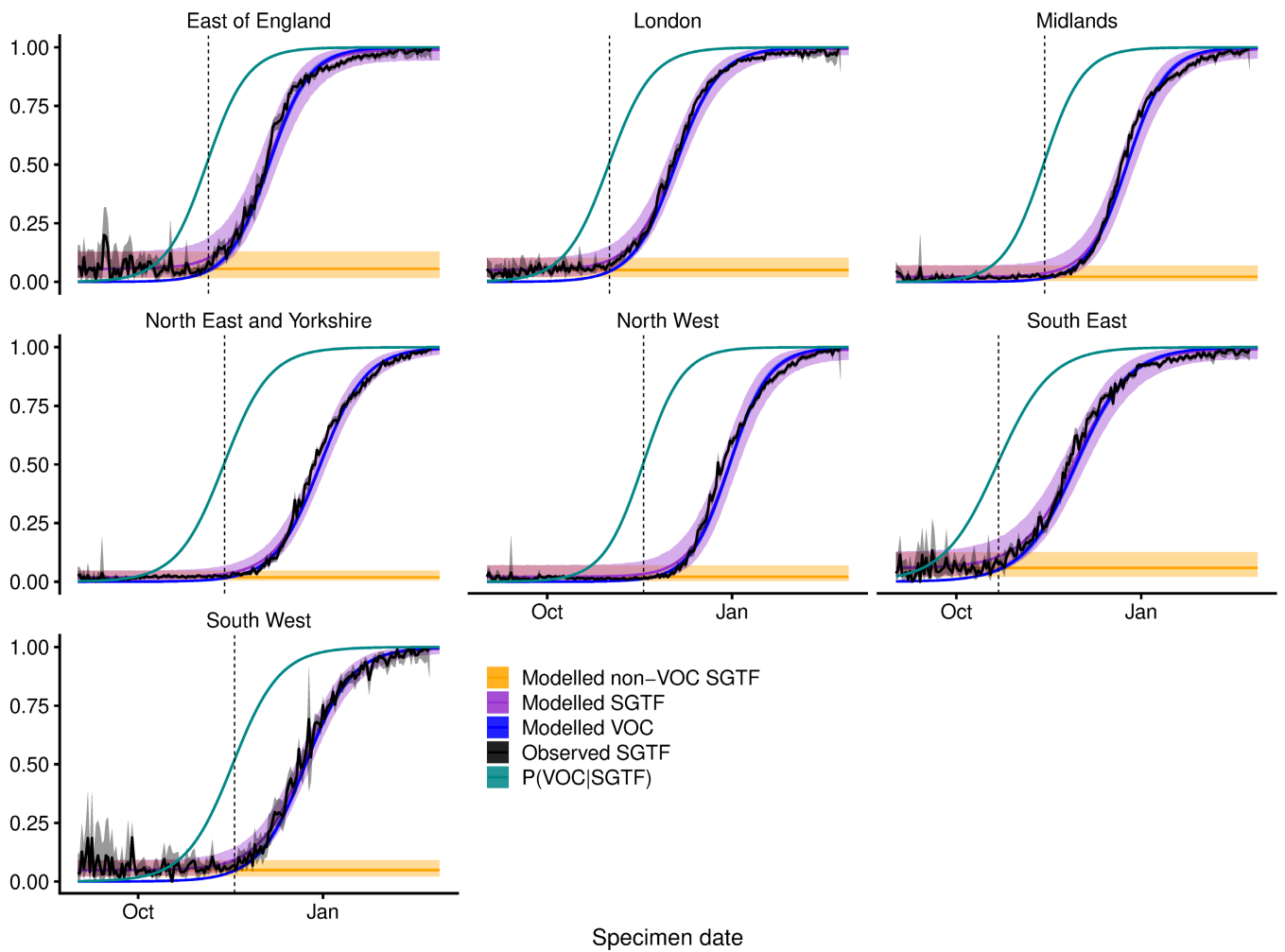
Extended Data Fig. 5 | Misclassification model. For each NHS England region, we fit a beta-binomial model (purple, modelled SGTF) to the observed SGTF frequencies among Pillar 2 tests (black, observed SGTF), which estimates a constant proportion of 'false-positive' SGTF samples among non-VOC 202012/01 (that is, non-B.1.1.7) specimens (orange, modelled non-VOC SGTF) and a logistically growing proportion of VOC 202012/01 (that is, B.1.1.7) specimens over time (blue, modelled VOC). This allows us to model the conditional probability that a specimen with SGTF represents VOC 202012/01 (teal, $P(\text{VOC}|\text{SGTF})$). For our misclassification survival analysis, $p_{\text{VOC}} = 0$ for non-SGTF specimens and $p_{\text{VOC}} = P(\text{VOC}|\text{SGTF})$ for SGTF specimens. Lines show medians and shaded areas show 95% credible intervals. Dashed vertical lines show the date on which $P(\text{VOC}|\text{SGTF})$ first exceeds 0.5.
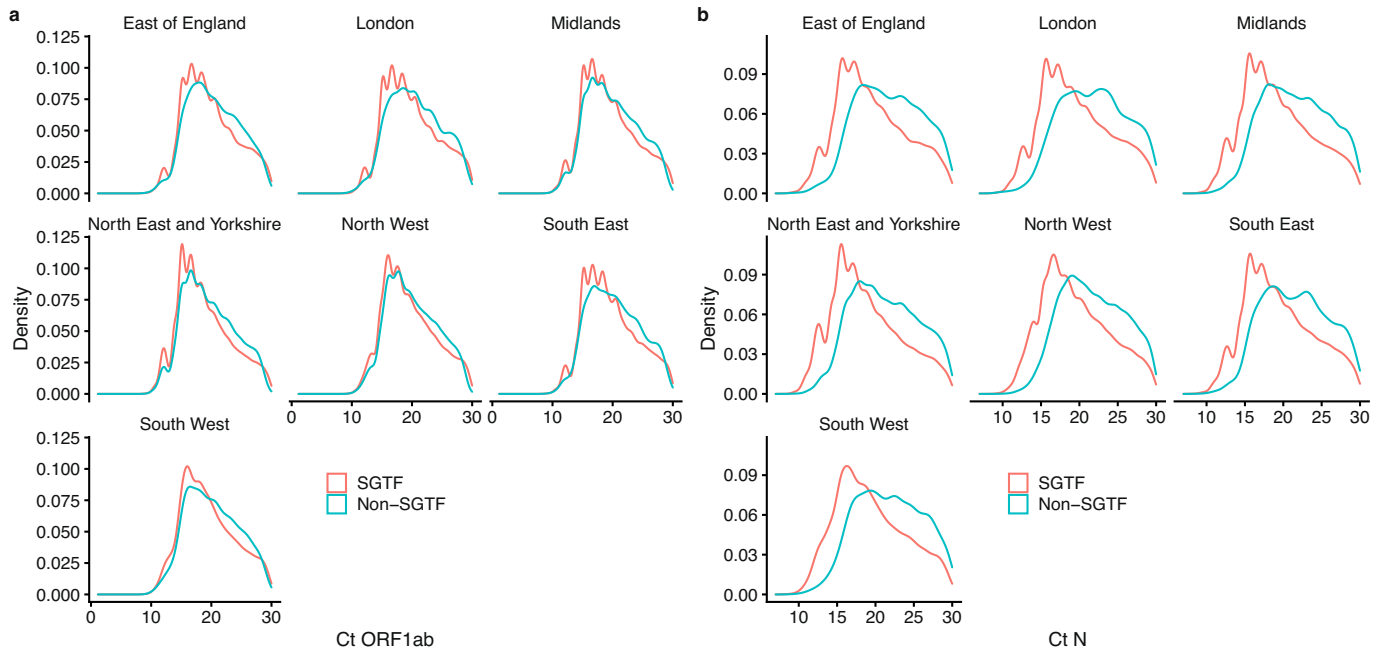
Extended Data Fig. 6 | C_t values for SGTF versus non-SGTF. a, b, The distribution of C_t values for *orf1ab* (**a**) and *N* (**b**) gene targets among specimens collected between 1 January and 14 February 2021.
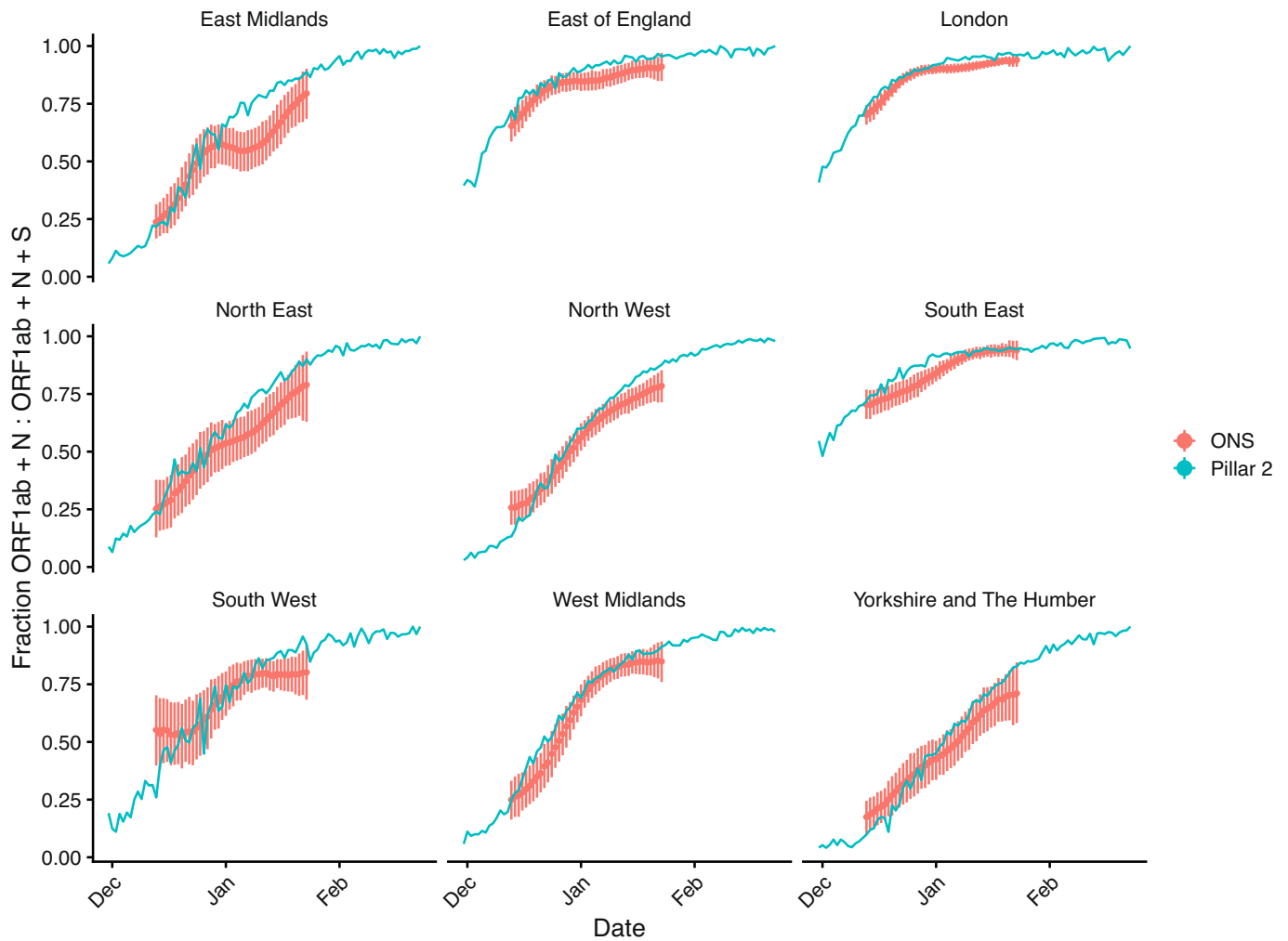
Extended Data Fig. 7 | S-gene dropout in community tests relative to a random sample of SARS-CoV-2 infections in the community. Comparison of the proportion of samples with *S*-gene dropout in the Pillar 2 (that is, community testing) sample used in this analysis compared to Office for National Statistics (ONS) random sampling of the community. This comparison suggests that *S*-gene dropout samples are not overrepresented in testing data relative to the prevalence of S-gene dropout in the community, suggesting that the increased hazard of death among positive community tests estimated in this study is not the result of a decrease in the average propensity for test-seeking among individuals infected with B.1.1.7. Point and ranges for ONS data show mean and 95% credible intervals.

Extended Data Table 1 | Characteristics of study participants, 1 November 2020–14 February 2021

	All	Missing	SGTF	Non-SGTF	SGTF prevalence	Missingness
	2,245,263 (100%)	1,098,729 (100%)	674,539 (100%)	471,995 (100%)	674,539 / 1,146,534 (58.8%)	1,098,729 / 2,245,263 (48.9%)
Sex						
Female	1,204,474 (53.6%)	604,118 (55%)	350,063 (51.9%)	250,293 (53%)	350,063 / 600,356 (58.3%)	604,118 / 1,204,474 (50.2%)
Male	1,040,789 (46.4%)	494,611 (45%)	324,476 (48.1%)	221,702 (47%)	324,476 / 546,178 (59.4%)	494,611 / 1,040,789 (47.5%)
Age						
1-34	994,161 (44.3%)	480,435 (43.7%)	303,163 (44.9%)	210,563 (44.6%)	303,163 / 513,726 (59%)	480,435 / 994,161 (48.3%)
35-54	772,080 (34.4%)	368,767 (33.6%)	240,406 (35.6%)	162,907 (34.5%)	240,406 / 403,313 (59.6%)	368,767 / 772,080 (47.8%)
55-69	339,862 (15.1%)	163,879 (14.9%)	102,141 (15.1%)	73,842 (15.6%)	102,141 / 175,983 (58%)	163,879 / 339,862 (48.2%)
70-84	96,653 (4.3%)	52,607 (4.8%)	23,587 (3.5%)	20,459 (4.3%)	23,587 / 44,046 (53.6%)	52,607 / 96,653 (54.4%)
85 and older	42,507 (1.9%)	33,041 (3%)	5,242 (0.8%)	4,224 (0.9%)	5,242 / 9,466 (55.4%)	33,041 / 42,507 (77.7%)
Place of residence						
Residential	2,103,186 (93.7%)	997,801 (90.8%)	650,403 (96.4%)	454,982 (96.4%)	650,403 / 1,105,385 (58.8%)	997,801 / 2,103,186 (47.4%)
Care/Nursing home	69,653 (3.1%)	61,238 (5.6%)	4,569 (0.7%)	3,846 (0.8%)	4,569 / 8,415 (54.3%)	61,238 / 69,653 (87.9%)
Other/Unknown	72,424 (3.2%)	39,690 (3.6%)	19,567 (2.9%)	13,167 (2.8%)	19,567 / 32,734 (59.8%)	39,690 / 72,424 (54.8%)
Ethnicity						
White	1,661,317 (74%)	801,261 (72.9%)	498,491 (73.9%)	361,565 (76.6%)	498,491 / 860,056 (58%)	801,261 / 1,661,317 (48.2%)
Asian	304,239 (13.6%)	141,753 (12.9%)	93,538 (13.9%)	68,948 (14.6%)	93,538 / 162,486 (57.6%)	141,753 / 304,239 (46.6%)
Black	104,222 (4.6%)	59,641 (5.4%)	31,018 (4.6%)	13,563 (2.9%)	31,018 / 44,581 (69.6%)	59,641 / 104,222 (57.2%)
Other/Mixed/Unknown	175,485 (7.8%)	96,074 (8.7%)	51,492 (7.6%)	27,919 (5.9%)	51,492 / 79,411 (64.8%)	96,074 / 175,485 (54.7%)
IMD decile						
1-2 (most deprived)	511,627 (22.8%)	220,159 (20%)	157,151 (23.3%)	134,317 (28.5%)	157,151 / 291,468 (53.9%)	220,159 / 511,627 (43%)
3-4	518,743 (23.1%)	262,880 (23.9%)	154,287 (22.9%)	101,576 (21.5%)	154,287 / 255,863 (60.3%)	262,880 / 518,743 (50.7%)
5-6	449,578 (20%)	232,880 (21.2%)	132,750 (19.7%)	83,948 (17.8%)	132,750 / 216,698 (61.3%)	232,880 / 449,578 (51.8%)
7-8	405,967 (18.1%)	203,710 (18.5%)	120,260 (17.8%)	81,997 (17.4%)	120,260 / 202,257 (59.5%)	203,710 / 405,967 (50.2%)
9-10	359,348 (16%)	179,100 (16.3%)	110,091 (16.3%)	70,157 (14.9%)	110,091 / 180,248 (61.1%)	179,100 / 359,348 (49.8%)
NHS England region						
East of England	278,518 (12.4%)	180,053 (16.4%)	76,352 (11.3%)	22,113 (4.7%)	76,352 / 98,465 (77.5%)	180,053 / 278,518 (64.6%)
London	506,242 (22.5%)	305,857 (27.8%)	151,081 (22.4%)	49,304 (10.4%)	151,081 / 200,385 (75.4%)	305,857 / 506,242 (60.4%)
Midlands	426,468 (19%)	203,450 (18.5%)	109,281 (16.2%)	113,737 (24.1%)	109,281 / 223,018 (49%)	203,450 / 426,468 (47.7%)
North East and Yorkshire	276,566 (12.3%)	58,991 (5.4%)	89,615 (13.3%)	127,960 (27.1%)	89,615 / 217,575 (41.2%)	58,991 / 276,566 (21.3%)
North West	268,856 (12%)	56,906 (5.2%)	106,257 (15.8%)	105,693 (22.4%)	106,257 / 211,950 (50.1%)	56,906 / 268,856 (21.2%)
South East	356,046 (15.9%)	199,190 (18.1%)	121,190 (18%)	35,666 (7.6%)	121,190 / 156,856 (77.3%)	199,190 / 356,046 (55.9%)
South West	132,567 (5.9%)	94,282 (8.6%)	20,763 (3.1%)	17,522 (3.7%)	20,763 / 38,285 (54.2%)	94,282 / 132,567 (71.1%)
Specimen date						
1 Nov-21 Nov	348,521 (15.5%)	119,271 (10.9%)	13,288 (2%)	215,962 (45.8%)	13,288 / 229,250 (5.8%)	119,271 / 348,521 (34.2%)
22 Nov-12 Dec	237,937 (10.6%)	86,881 (7.9%)	44,857 (6.7%)	106,199 (22.5%)	44,857 / 151,056 (29.7%)	86,881 / 237,937 (36.5%)
13 Dec-2 Jan	660,572 (29.4%)	355,297 (32.3%)	214,439 (31.8%)	90,836 (19.2%)	214,439 / 305,275 (70.2%)	355,297 / 660,572 (53.8%)
3 Jan-23 Jan	710,152 (31.6%)	386,329 (35.2%)	272,576 (40.4%)	51,247 (10.9%)	272,576 / 323,823 (84.2%)	386,329 / 710,152 (54.4%)
24 Jan-14 Feb	288,081 (12.8%)	150,951 (13.7%)	129,379 (19.2%)	7,751 (1.6%)	129,379 / 137,130 (94.3%)	150,951 / 288,081 (52.4%)

All, *n* (%); Missing, *n* (%); SGTF, *n* with SGTF (%) in subset of tests with non-missing SGTF status; Non-SGTF, *n* with non-SGTF (%) in subset of tests with non-missing SGTF status; SGTF prevalence, *n* with SGTF/total (%) in subset with known SGTF status; Missingness, *n* with missing SGTF status/total (%).

Article

Extended Data Table 2 | Rates of death within 28 days of a positive test among study participants

| | Death rates | | | | SGTF | |
	All	Missing	SGTF	Non-SGTF	prevalence (%)	Missingness (%)
	17,452 / 57,616,200 (3.03)	12,507 / 27,937,559 (4.48)	3,088 / 16,617,517 (1.86)	1,857 / 13,061,124 (1.42)	62%	72%
Sex						
Female	9,244 / 30,944,020 (2.99)	7,053 / 15,398,004 (4.58)	1,361 / 8,619,534 (1.58)	830 / 6,926,482 (1.2)	62%	76%
Male	8,208 / 26,672,181 (3.08)	5,454 / 12,539,554 (4.35)	1,727 / 7,997,984 (2.16)	1,027 / 6,134,643 (1.67)	63%	66%
Age						
1-34	71 / 25,680,074 (0.03)	29 / 12,326,846 (0.02)	25 / 7,508,250 (0.03)	17 / 5,844,978 (0.03)	60%	41%
35-54	712 / 19,864,316 (0.36)	341 / 9,403,924 (0.36)	261 / 5,943,218 (0.44)	110 / 4,517,174 (0.24)	70%	48%
55-69	2,244 / 8,701,146 (2.58)	1,117 / 4,165,412 (2.68)	787 / 2,495,452 (3.15)	340 / 2,040,282 (1.67)	70%	50%
70-84	5,839 / 2,417,392 (24.15)	3,939 / 1,300,303 (30.29)	1,144 / 563,478 (20.3)	756 / 553,610 (13.66)	60%	67%
85 and older	8,586 / 953,274 (90.07)	7,081 / 741,074 (95.55)	871 / 107,120 (81.31)	634 / 105,080 (60.33)	58%	82%
Place of residence						
Residential	6,790 / 54,144,918 (1.25)	3,114 / 25,493,353 (1.22)	2,363 / 16,050,245 (1.47)	1,313 / 12,601,320 (1.04)	64%	46%
Care/Nursing home	10,343 / 1,633,711 (63.31)	9,152 / 1,448,972 (63.16)	671 / 88,819 (75.55)	520 / 95,920 (54.21)	56%	88%
Other/Unknown	319 / 1,837,572 (1.74)	241 / 995,234 (2.42)	54 / 478,453 (1.13)	24 / 363,885 (0.66)	69%	76%
Ethnicity						
White	15,403 / 42,509,581 (3.62)	11,306 / 20,333,667 (5.56)	2,510 / 12,178,060 (2.06)	1,587 / 9,997,854 (1.59)	61%	73%
Asian	1,234 / 7,881,667 (1.57)	648 / 3,621,568 (1.79)	388 / 2,346,561 (1.65)	198 / 1,913,538 (1.03)	66%	53%
Black	334 / 2,704,512 (1.23)	205 / 1,532,482 (1.34)	101 / 796,623 (1.27)	28 / 375,406 (0.75)	78%	61%
Other/Mixed/Unknown	481 / 4,520,440 (1.06)	348 / 2,449,842 (1.42)	89 / 1,296,272 (0.69)	44 / 774,326 (0.57)	67%	72%
IMD decile						
1-2 (most deprived)	3,407 / 12,934,018 (2.63)	1,972 / 5,512,666 (3.58)	816 / 3,711,599 (2.2)	619 / 3,709,752 (1.67)	57%	58%
3-4	3,656 / 13,332,300 (2.74)	2,541 / 6,699,684 (3.79)	726 / 3,821,830 (1.9)	389 / 2,810,787 (1.38)	65%	70%
5-6	3,642 / 11,555,099 (3.15)	2,801 / 5,923,094 (4.73)	549 / 3,307,366 (1.66)	292 / 2,324,639 (1.26)	65%	77%
7-8	3,475 / 10,477,056 (3.32)	2,671 / 5,208,213 (5.13)	522 / 2,997,440 (1.74)	282 / 2,271,402 (1.24)	65%	77%
9-10	3,272 / 9,317,728 (3.51)	2,522 / 4,593,902 (5.49)	475 / 2,779,282 (1.71)	275 / 1,944,544 (1.41)	63%	77%
NHS England region						
East of England	2,513 / 7,214,404 (3.48)	2,163 / 4,623,842 (4.68)	272 / 1,976,732 (1.38)	78 / 613,831 (1.27)	78%	86%
London	1,875 / 13,363,072 (1.4)	1,439 / 7,990,202 (1.8)	359 / 3,999,013 (0.9)	77 / 1,373,858 (0.56)	82%	77%
Midlands	3,596 / 10,680,665 (3.37)	2,448 / 4,839,783 (5.06)	678 / 2,683,772 (2.53)	470 / 3,157,110 (1.49)	59%	68%
North East and Yorkshire	2,175 / 6,979,508 (3.12)	944 / 1,528,482 (6.18)	565 / 1,922,484 (2.94)	666 / 3,528,543 (1.89)	46%	43%
North West	1,869 / 6,745,667 (2.77)	910 / 1,439,165 (6.32)	554 / 2,393,090 (2.31)	405 / 2,913,412 (1.39)	58%	49%
South East	3,731 / 9,274,042 (4.02)	3,155 / 5,136,804 (6.14)	496 / 3,148,042 (1.58)	80 / 989,196 (0.81)	86%	85%
South West	1,693 / 3,358,841 (5.04)	1,448 / 2,379,282 (6.09)	164 / 494,384 (3.32)	81 / 485,175 (1.67)	67%	86%
Specimen date						
1 Nov-21 Nov	2,143 / 9,726,040 (2.2)	1,405 / 3,317,578 (4.24)	29 / 371,689 (0.78)	709 / 6,036,774 (1.17)	4%	66%
22 Nov-12 Dec	1,994 / 6,631,448 (3.01)	1,452 / 2,409,916 (6.03)	141 / 1,254,200 (1.12)	401 / 2,967,330 (1.35)	26%	73%
13 Dec-2 Jan	4,250 / 18,432,794 (2.31)	3,112 / 9,900,846 (3.14)	777 / 5,993,916 (1.3)	361 / 2,538,032 (1.42)	68%	73%
3 Jan-23 Jan	7,621 / 19,281,830 (3.95)	5,641 / 10,464,988 (5.39)	1,638 / 7,409,298 (2.21)	342 / 1,407,543 (2.43)	83%	74%
24 Jan-14 Feb	1,444 / 3,544,088 (4.07)	897 / 1,844,230 (4.86)	503 / 1,588,412 (3.17)	44 / 111,445 (3.95)	92%	62%

Total number of deaths, number of days of follow-up and deaths per 10,000 days of follow-up are reported for: all deaths, deaths with Missing SGTF status, deaths with known SGTF and deaths with known Non-SGTF. SGTF prevalence among deaths with non-missing SGTF status (%) and missingness in SGTF status among deaths are also reported.

Corresponding author(s): Nicholas G. Davies

Last updated by author(s): Mar 4, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used by investigators in collecting data, as this is a secondary data analysis using data that have already been collected.
Data analysis	R version 3.6.3 was used to analyse data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

An anonymised data set allowing replication of the analysis is available at https://zenodo.org/record/4579857.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was determined by the full data set available, as this was an observational study rather than a clinical trial. The full sample size used in the paper was 2,245,263 subjects.
Data exclusions	We excluded individuals with missing data for sex (n = 14), age (n = 171), and IMD and regional covariates (n = 3,817). We also excluded age-0 individuals as there were 17,913 age-0 individuals compared to 10,312 age-1 individuals in the dataset, suggesting that some age-0 individuals may have been miscoded.
Replication	As this was not an experimental study, it was not possible to replicate any experiment. However, we performed multiple sensitivity analyses considering different time period, covariates, and stratification methods and found consistent results for each replication. We reported all sensitivity analyses performed. We also analysed the data using both complete-cases and inverse probability weighting methods.
Randomization	Individuals were not randomised, as the exposure of interest was infection with one of two variants of SARS-CoV-2. Our statistical analysis controls for age, sex, ethnicity, type of residence (residential housing, care/nursing home, or other), index of multiple deprivation, date of virus test and local authority of residence.
Blinding	Individuals were not allocated to groups by the investigators. Investigators were not blinded to the exposure groups during analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Subjects were aged 1-110; male and female; White, Asian, Black, and other/mixed/unknown ethnicities; from all deciles of the Index of Multiple Deprivation in England, from all 7 NHS England regions, residing in residential housing, care homes, or other accommodation, and tested for SARS-CoV-2 between 1 September 2020 and 14 February 2021. Full details of the study population are presented in Extended Data Table 1.
Recruitment	This is a secondary data analysis study using linked data from all individuals who took a SARS-CoV-2 community test in England between 1 September 2020 and 14 February 2021.
Ethics oversight	Approved by the Observational / Interventions Research Ethics Committee at the London School of Hygiene and Tropical Medicine (reference number 24020).

Note that full information on the approval of the study protocol must also be provided in the manuscript.