PRELIMINARY – NOT PEER REVIEWED

Increased hazard of death in community-tested cases of SARS-CoV-2 Variant of Concern 202012/01

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*], Ruth H. Keogh[2,3*]

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.

[†]Corresponding author. E-mail: nicholas.davies@lshtm.ac.uk
*Equal contribution

VOC 202012/01, a SARS-CoV-2 variant first detected in the United Kingdom in September 2020, has spread to multiple countries worldwide. Several studies have established that this novel variant is more transmissible than preexisting variants, but have not identified whether it leads to any change in disease severity. We analyse a large database of SARS-CoV-2 community test results and COVID-19 deaths, representing 52% of all SARS-CoV-2 community tests in England from 1 September 2020 to 5 February 2021. This subset of SARS-CoV-2 tests can identify VOC 202012/01 because mutations in this lineage prevent PCR amplification of the spike gene target (S gene target failure, SGTF). We estimate that the hazard of death among SGTF cases is 58% (95% CI 40–79%) higher than among non-SGTF cases after adjustment for age, sex, ethnicity, deprivation level, care home residence, local authority of residence and test date. This corresponds to the absolute risk of death for a male aged 55–69 increasing from 0.6% to 0.9% (95% CI 0.8–1.0%) over the 28 days following a positive test in the community. Correcting for misclassification of SGTF and missingness in SGTF status, we estimate a 71% (48–97%) higher hazard of death associated with VOC 202012/01. Our analysis suggests that VOC 202012/01 is not only more transmissible than preexisting 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 Variant of Concern (VOC) 202012/01 is a 6-nucleotide deletion which prevents amplification of the S gene target by the commercial PCR assay used in three of the Lighthouse labs[1]. By linking individual records of positive community tests with and without S gene target failure (SGTF) to a comprehensive line list of COVID-19 deaths in England, we estimate the relative hazard of death associated with infection by VOC 202012/01. We define confirmed SGTF as a compatible PCR result with cycle threshold (Ct) < 30 for ORF1ab, Ct < 30 for N, and no detectable S (Ct > 40); confirmed non-SGTF as any compatible PCR result with Ct < 30 for each of ORF1ab, N, and S; and an inconclusive (missing) result as any other positive community test, including tests processed by a laboratory incapable of assessing SGTF. We address missing SGTF status in our analysis.

Characteristics of the study population

The study sample (**Table 1**) includes a total of 1,994,449individuals who had a positive community ("Pillar 2") test between 1 November 2020 and 25 January 2021. Just over half of those tested (1,028,296, 52%) had a conclusive SGTF reading and, of these, 48% had SGTF. Females comprised 53.7% of the total sample; 44.4% were aged 1–34 years, 34.3% aged 35–54, 15.1% aged 55–69, 4.3% aged 70–85 and 1.9% aged 85 or older. The majority of individuals (93.7%) lived in residential accommodation (defined as residing in a house, flat, sheltered accommodation, or house in multiple occupancy), with 3.1% living in a care or nursing home. Based on self-identified ethnicity, 73.8% were White, 13.7% Asian, 4.7% Black and 7.8% of other, mixed or unknown ethnicity. The data include tests performed in all 7 NHS England regions, with the London region contributing 23.4% of tests and the South West 5.8%. The first two weeks of the study period (1–14 Nov) contributed 12.6% of the total tests, and the final two weeks (10–25 Jan) 22.2%. The period between 27 Dec and 9 Jan contributed 30.5% of tests.

In those with SGTF status measured, SGTF prevalence was similar in males and females but lower in the older age groups: 54.9% in the 1-34 year olds compared with 48.6% in those aged 85 and older. In keeping with these age patterns, SGTF prevalence was lower in individuals living in a care or nursing home (45.2%, compared to 54.7% among those in residential accommodation). SGTF prevalence by self-identified ethnicity was 53.5% in the White group, 54.0% in the Asian group, 67.2% in the Black group, and 61.6% in the other, mixed, or unknown ethnicity group. SGTF prevalence was lowest in the most deprived index of multiple deprivation[1] (IMD) decile (43.9%) and highest in the least deprived decile (58.7%). The highest prevalences of SGTF over the study period were observed in the East of England (75.7%), South East (75.6%) and London (74.0%) NHS England regions, and prevalence of SGTF was lowest in the North East and Yorkshire region (32.5%). The prevalence of SGTF also increased steeply over time (**Fig. 1a**), ranging from 4.9% during 1–14 November 2020 to 87.4% during 10 –25 January 2021.

Having missing SGTF status was strongly associated with age and place of residence. The proportion with SGTF status missing was similar in age groups 1-34 (47.9%), 35-54 (47.1%) and 55-69 (47.7%), and then rose to 54.3% in the 70-84 age group and to 78.6% in the 85 and older age group. SGTF status was missing in 89.1% of tests for individuals living in a care or nursing home, compared to 46.9% of tests among individuals in residential accommodation. This is partly due 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 by NHS England region, ranging from 21.5% in the North West to 70.8% in the South West. Missingness also depended on specimen date, with the percentage missing being lower for the earlier specimen dates and highest (55.4%) in the 2 week period that contributed the most tests (27 December-9 January). There were also some more minor differences in the percentages of missingness of SGTF status by ethnicity and IMD. Of the 48% of tests with missing SGTF status, 9% were inconclusive due to high Ct values and the remaining 39% were not analysed in one of the three Lighthouse labs capable of producing an SGTF result.

The most commonly used definition of a COVID-19 death in England is any death occurring within 28 days of a positive SARS-CoV-2 test. **Table 2** presents crude death rates within 28 days of a positive test per 10,000 person-days of follow-up. Death rates for unlimited follow-up (i.e. not restricted to 28 days) are shown in Table S1; the maximum observed follow-up was 85 days. A total of 13,860 individuals out of the 1,994,449 in the study sample are known to have died (0.69%), 12,967 of whom (92.8%) died within 28 days of their first positive test (**Fig. 1b**). As expected, crude death rates were substantially higher in the elderly and in those living in a care or nursing home.

Crude survival assessed by Kaplan-Meier curves was lower in the SGTF group (**Fig. 1c**). Stratifying by broad age groups and looking at death rates by sex, place of residence, ethnicity, IMD, NHS England region, and specimen date, it can be seen that death rates within 28 days of a positive SARS-CoV-2 test are higher among SGTF than non-SGTF cases in 99 of the 108 strata assessed (92%; **Figs. 1d–i**).

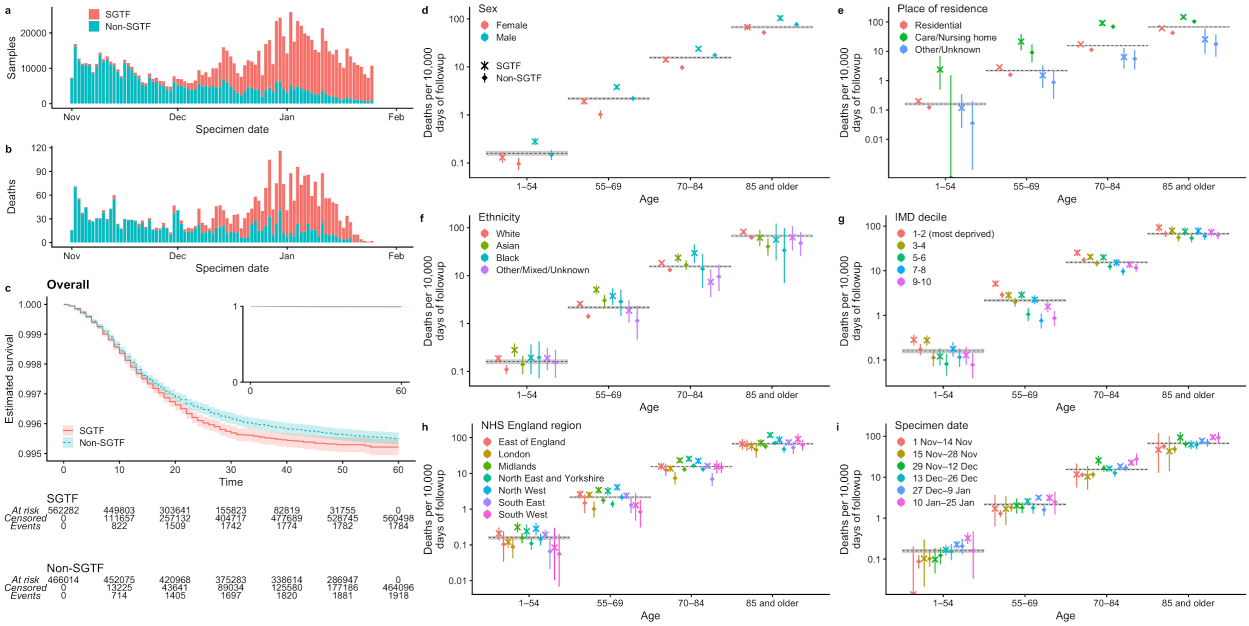

Fig. 1. Descriptive analyses. a The number of samples with and without SGTF by day from 1 November 2020 to 25 January 2021, the period covered by our main analysis. **b** Number of deaths within 28 days of positive test by specimen date included in the analysis. **c** Kaplan-Meier plot showing survival among individuals tested in the community in England with and without SGTF, in the subset with SGTF measured. Inset shows the full *y*–axis range. **d–i** Crude death rates (with 95% confidence intervals) in SGTF and non-SGTF tests (in the subset with SGTF measured) for deaths within 28 days of positive test stratified by broad age groups and (**d**) sex, (**e**) place of residence, (**f**) ethnicity, (**g**) index of multiple deprivation, (**h**) NHS England region, and (**i**) specimen date. Dotted lines show the overall crude death rates by age group irrespective of SGTF status, with the shaded area showing the 95% CIs.

Cox regression analyses

To estimate the effect of SGTF on mortality while controlling for observed confounding, we fitted a series of Cox proportional hazards models[2] to the data. We stratified the analysis by lower tier local authority (LTLA) and specimen date to control for geographical and temporal differences in the baseline 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. All models were fitted twice, once using complete cases only, i.e. by simply excluding individuals with missing SGTF status, and once using inverse probability weighting (IPW), i.e. accounting for missingness by upweighting individuals whose characteristics—age, sex, IMD, ethnicity, residence type, NHS England region of residence and sampling week—are underrepresented among complete cases.

For the complete-cases analysis, the estimated hazard ratio for SGTF was 1.58 (95% CI 1.40–1.79), indicating that the hazard of death within 28 days of a positive test is 58% (40–79%) higher in those with SGTF compared to non-SGTF (**Fig. 2a**). We included an interaction term between SGTF and time since positive test in the model to assess the proportional hazards assumption. There was strong evidence of non-proportionality of hazards (likelihood ratio test $P(\chi^2_1 = 7.1) = 0.008$; **Fig. 2a; Fig. S11**). The estimated time-varying hazard ratio increases over time: 1.19 (0.94–1.52) one day after the positive test, 1.66 (1.46–1.88) on day 14, and 2.36 (1.71–3.25) on day 28. There was no evidence that adding higher-order functions of time into the interaction terms improved model fit (likelihood ratio test $P(\chi^2_1 = 1.0) = 0.32$), and no evidence of a significant interaction between time and age ($P(\chi^2_1 = 0.03) = 0.87$), time and sex ($P(\chi^2_1 = 3.6) = 0.056$), time and IMD ($P(\chi^2_1 = 0.10) = 0.75$), time and ethnicity ($P(\chi^2_3 = 1.4) = 0.71$), or time and residence type ($P(\chi^2_2 = 1.5) = 0.47$).

We found no evidence of a significant interaction between SGTF and age group (likelihood ratio test $P(\chi^2_4 = 6.7) = 0.15$), sex ($P(\chi^2_1 = 0.44) = 0.51$), IMD ($P(\chi^2_9 = 5.0) = 0.84$), or ethnicity ($P(\chi^2_3 = 0.95) = 0.81$). There was some evidence of an interaction between SGTF and residence type ($P(\chi^2_2 = 6.8) = 0.034$), with the associated hazard ratio for SGTF being 1.53 (1.35–1.74) in standard residential accommodation , 2.43 (1.72–3.45) in care/nursing homes, and 1.64 (0.80–3.38) in “other” residence types (i.e. residential institutions including residential education, prisons and detention centres, medical facilities, no fixed abode and other/unknown).

In the investigation of a model for the probability of missingness in SGTF status, the cauchit model was found to provide a good fit and to result in less extreme weights than the logistic model. The IPW analysis was therefore performed using weights derived from the cauchit model. The IPW analysis yielded similar results to the complete-cases analysis, generally with marginally higher hazard ratios and wider CIs (**Fig. 2e**); the hazard ratio associated with SGTF for the IPW analysis was 1.67 (1.46–1.90). While the IPW analysis recovered a similarly time-varying hazard ratio to the complete-cases analysis, the increase was less marked (**Fig. 2c**) and the inclusion of a time-varying term did not significantly improve model fit (Wald test$P(\chi^2_1 = 1.0) = 0.33$).

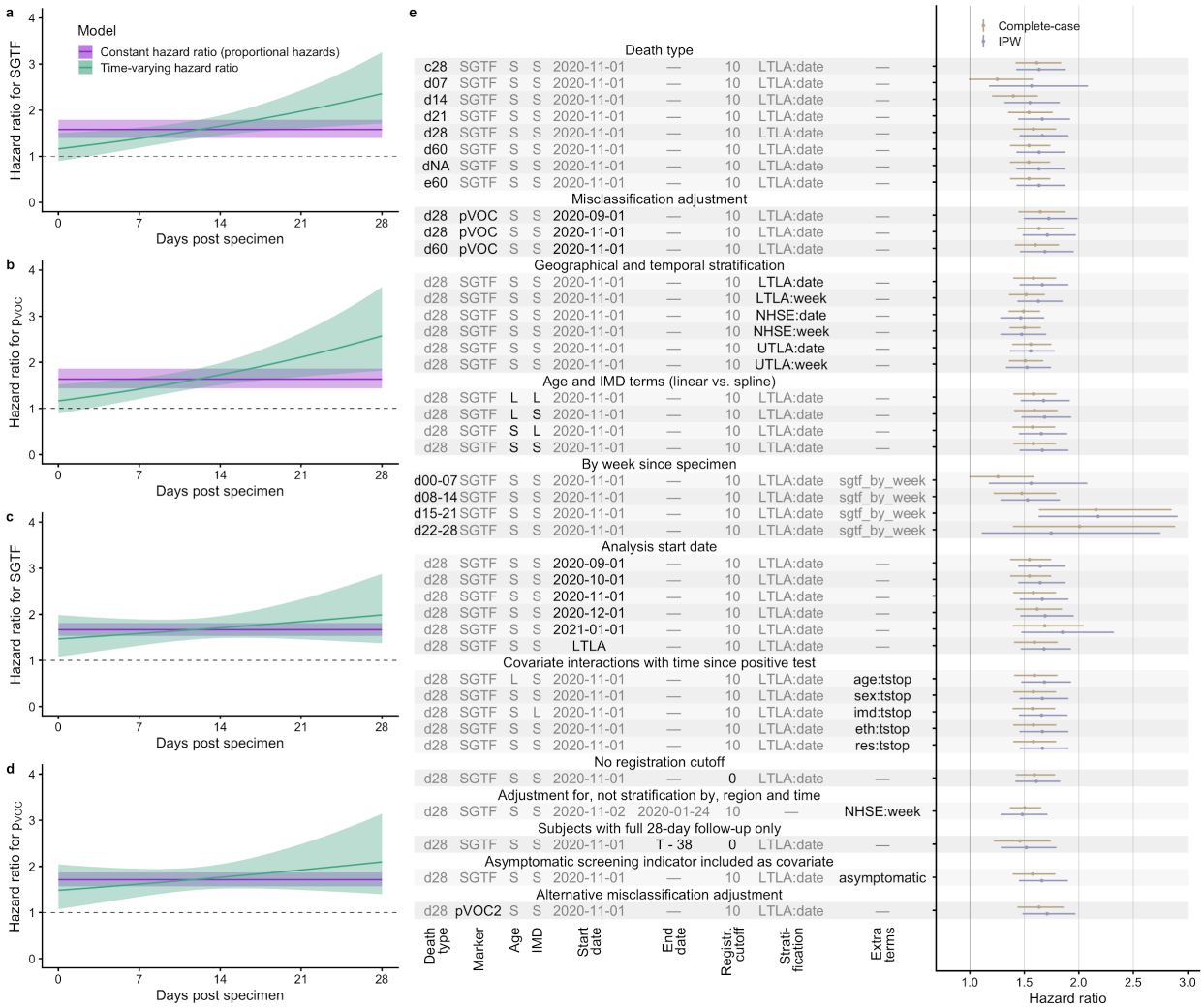

Fig. 2. Survival analyses. a–d Estimated hazard ratio of death within 28 days of positive test for (**a**) SGTF, complete-cases analysis; (**b**) p_{VOC}, complete-cases analysis; (**c**) SGTF, IPW analysis; and (**d**) p_{VOC}, IPW analysis, in model stratified by LTLA and specimen date and adjusted for the other covariates. **e** Estimated hazard ratio of death within 28 days of positive test across each model investigated. Death types are coded as follows: d*X*, all deaths within *X* days of a positive test; c28, death-certificate-confirmed COVID-19 deaths within 28 days; e60, all deaths within 60 days plus all death-certificate-confirmed COVID-19 deaths within any time period. S, spline term (for Age or IMD); L, linear term (for Age or IMD); LTLA, lower-tier local authority (n = 316); UTLA, upper-tier local authority (n = 150); NHSE, NHS England region (n = 7). LTLA start date signifies a start date chosen separately for each LTLA (see Methods). Point estimates and 95% confidence intervals shown.

170 ***Misclassification analysis***
171
172 Prior to the emergence of VOC 202012/01, a number of minor circulating SARS-CoV-2 lineages
173 with spike mutations could also cause SGTF[3]. Our main analyses are restricted to specimens
174 from 1 November 2020 onwards to minimise the number of these non-VOC 202012/01 lineages
175 among SGTF-positive samples. However, the appearance of non-VOC 202012/01 samples in
176 SGTF may dilute the estimated effect of VOC 202012/01 on the hazard of mortality. We
177 therefore undertook a misclassification analysis[4], modelling the relative frequency of SGTF over
178 time for each NHS England region as a combination of a low, time-invariant frequency of non-
179 VOC 202012/01 samples with SGTF plus a logistically growing[5] frequency of VOC 202012/01
180 samples with SGTF, which allows us to assign to each SGTF sample a probability p_{VOC} that the
181 sample is VOC 202012/01 based upon its specimen date and NHS England region (**Fig. S9**).
182 Again restricting the analysis to specimens from 1 November 2020 onward, we find a hazard
183 ratio associated with p_{VOC} of 1.63 (1.44–1.86) for the complete-cases analysis and 1.71 (1.48–
184 1.97) for the IPW analysis.
185
186 ***Absolute risks***
187
188 To put these results into context, we estimated how the absolute risk of death due to COVID-19
189 may differ had an individual been infected with VOC 202012/01 compared with had they been
190 infected with the original variant. We calculated absolute risks by applying 28--day hazard ratios
191 for SGTF to the baseline risk of death estimated among individuals tested in the community
192 between August–October 2020 (expected to be representative of the CFR associated with
193 preexisting variants of SARS-CoV-2; **Table 3**). The risk of death due to COVID-19 following a
194 positive test in the community remains below 1% in most individuals younger than 70 years old.
195 For the complete cases analysis, in females aged 70–84, the estimated risk of death within 28
196 days of a positive SARS-CoV-2 test with SGTF increases from 2.9% to 4.5% (95% CI 4.0–
197 5.1%) and for females 85 or older increases from 13% to 20% (17–22%). For males aged 70–84
198 the risk of death within 28 days increases from 4.7% to 7.3% (6.4–8.2%) and for males 85 or
199 older it increases from 17% to 26% (23–28%). Estimates based on the IPW analysis were
200 marginally higher. These estimates reflect a substantial increase in absolute risk amongst older
201 age groups. Note that these estimates do not reflect the infection fatality ratio, but the fatality
202 ratio among people tested in the community, and are thus likely to be higher than the infection
203 fatality rate as many infected individuals will not have been tested.
204
205 ***Further investigations***

206 We conducted a number of sensitivity analyses to verify the robustness of our results. Our main
207 results were largely insensitive to: restriction to death-certificate-confirmed COVID-19 deaths
208 only; any follow-up time of 21 days or longer; coarseness of geographical and temporal
209 stratification; use of linear versus spline terms for age and IMD; analysis start date; followup
210 time–covariate interactions; removal of the 10-day death registration cutoff; and restriction of the
211 analysis to individuals with a full 28-day follow-up period (**Fig. 2e**; **Table S2**). Pillar 2 testing
212 data include an indicator for whether the subject was tested because of symptoms or due to

asymptomatic screening. Although symptomatic status may lie on the causal pathway between SGTF status and death, we adjusted for symptomatic status as a further sensitivity analysis and found that it had no effect on the relative hazard of SGTF (1.58 [1.39–1.79], complete-cases analysis).

Discussion

Our analysis identifies an increased hazard of death associated with VOC 202012/01 infection relative to infection by preexisting SARS-CoV-2 variants. We controlled for several factors that we hypothesised could confound the association between VOC 202012/01 infection status and mortality. By controlling for test time and geographical location, via stratified analysis, mimicking matching on these variables, we aimed to account for the fact that VOC 202012/01 infection increased rapidly over time and differed substantially by region, and also that the hospitals in which some individuals will have required care were subject to pressure on health services that changed over time and by region.

We do not attempt to identify the mechanism for an increased mortality rate in this analysis. There is some evidence that infections with VOC 202012/01 may be associated with higher viral loads, as measured by Ct values detected during PCR testing of specimens (**Fig. S10**). Higher viral loads resulting from infection with VOC 202012/01 may be partly responsible for the observed increase in mortality, partly because they may reduce the efficacy of standard antiviral treatments for COVID-19. The impact of viral load on observed SGTF mortality could be assessed using a mediation analysis, which is outside the remit of this study.

Another potential explanation for an increased mortality rate among individuals testing positive for VOC 202012/01 may be that this variant leads to changes in testing behaviour. If individuals infected with this variant are less likely to show symptoms, then only relatively more severe cases may get tested, and consequently our study would overestimate the infection fatality rate. However, comparison to random population testing carried out by the Office for National Statistics suggests no clear difference in the proportion of SGTF among Pillar 2 tests relative to the population at large (**Fig. S12**).

We previously identified that the novel SARS-CoV-2 lineage VOC 202012/01 appears to have a substantially greater transmission rate than preexisting variants of SARS-CoV-2[5], but could not robustly estimate any increase or decrease in associated disease severity from ecological analysis. The individual-level linked community testing data analysed here suggest that the fatality rate among individuals infected with VOC 202012/01 is higher than that associated with infection by preexisting variants. Crucially, due to the nature of the data currently available, we were only able to assess mortality among individuals who received a positive test for SARS-CoV-2 in the community. Indicators for VOC 202012/01 are not currently available for the vast majority of individuals who die due to COVID-19, as they are first tested in hospital. Accordingly, the evidence we provide here must be contextualised with further study of a larger population sample, and in other settings. Nonetheless, by focusing on individuals tested in the community, our analysis captures any combined effect of an altered risk of hospitalisation given positive test

and an altered risk of death given hospitalisation, which would not be fully captured by a study focusing on hospitalised patients only.

Our findings are consistent with those identified by other groups using different methods to verify the increased risk of death among community-tested individuals with SGTF[6]. Estimates of increased mortality based upon Pillar 2 data will become more robust as test results and mortality outcomes continue to accumulate over time. However, our approach of comparing outcomes between individuals with and without SGTF who were tested in the same place and at the same time would no longer accrue additional information at the point when SGTF becomes effectively fixed in England, which may occur as soon as February 2021 if current trends continue[5].

Methods

Data sources — We linked three datasets provided by Public Health England: a line list of all positive tests in England's "Pillar 2" (community) testing for SARS-CoV-2, containing specimen date and demographic information on the test subject; a line list of cycle threshold (Ct) 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) utilising the Thermo Fisher TaqPath COVID-19 assay; and a line list of all deaths due to COVID-19 in England, which combines and deduplicates deaths reported by hospitals in England, by the Office for National Statistics, via direct reporting from Public Health England Health Protection Team, and via Demographic Batch Service tracing of laboratory-confirmed cases [7]. We link these datasets using a numeric identifier for Pillar 2 tests ('FINALID') common to all three datasets. We define S gene target failure (SGTF) as any test with Ct < 30 for ORF1ab and N targets but no detectable S gene, and non-SGTF as any test with Ct < 30 for ORF1ab, N, and S targets. A small proportion (9%) 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 25 January 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 VOC 202012/01. In sensitivity analyses, we also consider extending the population to include tests performed between 1 September and 31 October 2020..

The linked dataset available for analysis excludes individuals who first tested positive in hospital, that is, those who presented to hospital after symptom onset without first being tested in the community. This is because cycle threshold values used to ascertain SGTF status are not available for individuals who were not tested in the community. Our study sample comprises all community tests between 1 November 2020 and 25 January 2021, but only 7% of the total number of COVID-19 deaths were recorded within 28 days following a positive test in either the community or in hospital during this period. This is explained by differing mortality rates among individuals who first test positive in a hospital compared to those who first receive a community test.

There was a small amount of missing data for sex (n = 13, <0.01%), age (n = 151, <0.01%), and IMD and regional covariates (n = 3,428, 0.15%). 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 16,936 age-0 individuals compared to 8,867 age-1 individuals in the dataset, suggesting that many of these age-0 individuals may have been miscoded. There was some missing data on ethnicity (n = 43,032, 2%) and we created a category that combines missing values with "Other" and "Mixed". Missing values for residence type (n = 67,458, 3%) were also combined with an "Other" category. The data set used for the main analysis comprises 1,994,449 individuals, and SGTF status is missing for 966,153 (48%). In addition, the SGTF status of 97,461 individuals (9%) with an inconclusive SGTF test was set to missing. 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, sheltered accommodation)” and “House in multiple occupancy (HMO)” groups; Care/Nursing home; and Other/Unknown, which included the "Medical facilities (including hospitals and hospices, and mental health)", "No fixed abode", "Other property classifications", "Overseas address", "Prisons, detention centres, secure units", "Residential institution (including residential education)", and "Undetermined" groups, as well as 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 / 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 to be associated with both SGTF and with risk of death, thus confounding the association between SGTF and risk of death in those tested. Area of residence and specimen date were expected to be potentially strong confounders. 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 following a test, including due to differential pressure on hospital resources by area and time. The following variables were also identified as potential confounders: sex, age, place of residence (Residential, Care/Nursing home, or Other/Unknown), ethnicity (White, Asian, Black, or Other/Mixed/Unknown), index of multiple deprivation (IMD). The potential confounders are referred to collectively as the covariates. For descriptive analyses, age (in years) was categorised as 1-34, 35-54, 55-69, 70-84, 85 and older.

Descriptive analyses were performed. We tabulated the distribution of the covariates in the whole study sample, and the association between each covariate and SGTF status in the subset with SGTF measured (Table 1). We also summarised the association between each covariate and missing data in SGTF status (Table 1). The subset with SGTF status 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 mortality rates (**Table 2**) are also presented separately according to categories of the covariates (**Figs. S1–S7**). Crude overall mortality rates were obtained for the whole sample, by SGTF status in the complete cases, and and in those with missing SGTF status, according to categories of each covariate (**Table 2**). We also obtained mortality rates by SGTF status (in the complete cases) for categories of each covariate stratified by age group. Exact Poisson CIs are used for mortality rates, assuming constant rate.

Approximately 46% of individuals in the study sample are missing data on SGTF status, due to their test not being sent to one of the three laboratories utilising the Thermo Fisher TaqPath COVID-19 assay or the test being inconclusive. We performed complete cases analysis, restricted to the subset with SGTF status measured. This complete case analysis assumes that for each analysis, the missing data, in this case missing SGTF status, is independent from 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 SFTG. We also performed an analysis of the complete cases using inverse probability weights[8] (IPW) to address the missing data on SGTF, under a missing at random assumption (MAR). 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 , 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[9]. The fit of the missingness model was assessed using a Q-Q plot (**Fig. S11**), and Hosmer-Lemeshow and Hinkley tests were used to choose the most appropriate model.

Cox regression[2] was used to estimate the association between SGTF and the hazard for mortality, conditioning on the potential confounders listed above. The analyses described here were applied to the complete cases and using IPW, For IPW analyses, the standard errors (SEs) accounted for the weights, though the fact that the weights were estimated was not accounted for. This results in conservative SEs. 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 die and individuals who survive contribute to the analysis. The analysis is therefore similar to that which would be performed had we created a matched nested case-control sample. The remaining variables were included as covariates in the model (sex, age, place of residence, ethnicity, IMD decile). Age and IMD were included as restricted cubic splines with 3 knots. The time origin for the analysis was specimen date and we considered deaths up to 28 days after the specimen date. Individuals who did not die within 28 days were censored at the earlier of 28 days post 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 (i.e., 25 January 2021) in order to minimise any potential bias due to late reporting of deaths. We began by assuming proportionality of hazards for SGTF and the covariates included in the model. The proportional hazards assumption 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 (**Fig. S8**). We assessed whether the association between SGTF and the hazard was modified by age, sex, IMD, ethnicity, and place of residence. Models with and without interactions were compared using likelihood ratio tests for the complete cases analyses. For the analysis using IPW we used Wald tests based on robust standard errors[10].

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 variants of SARS-CoV-2 in addition to VOC 202012/01 are also associated with failure to amplify the spike gene target. Accordingly, a positive test with SGTF is not necessarily indicative of infection with VOC 202012/01. A negative test of SGTF is assumed to be indicative of absence of infection with VOC 202012/01. 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 VOC 202012/01, assuming a beta binomial prior. This model is then used to estimate the probability that an individual testing positive with SGTF is infected with VOC 202012/01, 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[4] to correcting for bias due to measurement error in an exposure. .

We fitted models accounting for false positives (modelled as regionally-varying background rates of SGTF associated with non-VOC 202012/01 variants) to the SGTF data. Our logistic model for VOC 202012/01 growth over time is as follows:

$$logit\ (f(t)) = (slope \times (t - intercept))$$
$$s(t) = f(t) + (1 - f(t)) \times falsepos$$
$$k_t \sim betaBinomial(n = n_t, \alpha = s(t) \times (conc - 2) + 1, \beta = (1 - s(t)) \times (conc - 2) + 1)$$
$$slope \sim normal(\mu = 0, \sigma = 1)$$
$$intercept \sim normal(\mu = 0, \sigma = 1000)$$
$$falsepos \sim beta(\alpha = 1.5, \beta = 15)$$
$$conc \sim normal(\mu = 0, \sigma = 500) \geq 2$$

Here, $f(t)$ is the predicted frequency of VOC 202012/01 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 S gene target failure at time t due to the combination of VOC 202012/01 and a background false positive rate *falsepos*, *conc* is the "concentration" parameter (= α + β) of a beta distribution with mode $s(t)$; k_t is the number of S gene target failures detected at time t; and n_t is the total number of tests at time t. All priors above are chosen to be vague, and the truncation of *conc* 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_{VOC} = 0 for all tests with SGTF = 0.

The model above was fitted using the same data source (i.e. 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[11] downloaded from the Microreact platform[12] on 11 January 2020 to estimate p_{VOC}. In this alternative analysis we estimated p_{VOC} for each NHS England region and

446 date as the number of samples that were VOC 202012/01 (i.e. lineage B.1.1.7 with mutations
447 Δ69/Δ70 and N501Y in Spike) divided by the number of samples that were SGTF (i.e. any
448 lineage with Δ69/Δ70, the deletion that causes SGTF) for that NHS England region and date,
449 setting p_{VOC} = 1 for all dates later than 31 December 2020 as there were no sequencing data
450 available past this date, and filling any gaps in the data using linear interpolation. This yielded
451 nearly identical results to our modelled probability of VOC (**Fig. 2e**).
452
453 *Absolute risks* — Estimates from the final Cox models were used to obtain estimates of absolute
454 risk of death for 28 and 60 days with SGTF and p_{VOC}. Given the strong influence of age on risk
455 of death, we present absolute risks by sex and age group (1-34, 35-54, 55-69, 70-84, 85+).
456 Absolute risks of death (case fatality rate) within 28 and 60 days were estimated by age group
457 and sex using data on individuals tested during September 2020; this is referred to as the
458 baseline risk. The absolute risks of death for individuals with SGTF were then estimated as
459 follows. If the baseline absolute risk of death in a given age group is $(1 - A)$, then the estimated
460 absolute risk of death with SGTF is $(1 - A^{HR})$, where HR denotes the estimated hazard ratio
461 obtained from the Cox model assuming proportional hazards. We applied the hazard ratio from
462 28 days to the baseline risk for 28 days, and the hazard ratio for 60 days to the baseline risk for
463 60 days, to estimate absolute risks of death for individuals with SGTF and uncertainty of these
464 estimates. Standard errors are obtained via the delta method, and CIs based on normal
465 approximations.
466
467 *Sensitivity analyses* — Several sensitivity analyses were performed. After establishing the final
468 model through using the process outlined above we investigated the impact of using different
469 variables for stratification of the baseline hazard measuring region at a coarser level (UTLA, or
470 NHS England region), as well as coarser test specimen time (week rather than exact date).
471 Adjusting for these variables instead of using stratification was also explored. We also repeated
472 the main analysis restricting data to specimens collected from September onwards, October
473 onwards, November onwards, or December onwards.

474 To assess the impact of imposing an administrative cutoff to follow-up time of 10 days prior to
475 data extraction, we first reanalysed the data without this cutoff, as well as reanalysing the data
476 restricting the analysis to individuals with at least 28 days' follow-up.

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

481 **Acknowledgements**
482 We gratefully acknowledge the assistance of Public Health England in providing the analysis
483 data and authorising release of an anonymised data set.
484
485 **Funding statement**

NGD: UK Research and Innovation (UKRI) Research England; NIHR Health Protection Research Unit in Immunisation (NIHR200929); UK Medical Research Council (MC_PC_19065). CIJ: Global Challenges Research Fund project 'RECAP' managed through Research Councils UK and the Economic and Social Research Council (ES/P010873/1). WJE: European Commission (EpiPose 101003688), National Institutes of Health Research (NIHR200908). NPJ: National Institutes of Health / National Institute of Allergy and Infectious Diseases (R01AI148127). KDO: Royal Society-Wellcome Trust Sir Henry Dale Fellowship 218554/Z/19/Z. RHK: UKRI Future Leaders Fellowship (MR/S017968/1).

Working group authors and acknowledgements

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

Funding statements for the CMMID COVID-19 working group are as follows. 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). SC: Wellcome Trust (grant: 208812/Z/17/Z). FYS: NIHR EPIC grant (16/137/109). SFunk: Wellcome Trust (grant: 210758/Z/18/Z), NIHR (NIHR200908). GFM: NTD Modelling Consortium by the Bill and Melinda Gates Foundation (OPP1184344). YJ: LSHTM, DHSC/UKRI COVID-19 Rapid Response Initiative. SRM: Wellcome Trust (grant: 210758/Z/18/Z). RL: Royal Society Dorothy Hodgkin Fellowship. WJE: European Commission (EpiPose 101003688), NIHR (NIHR200908). MQ: European Research Council Starting Grant (Action Number #757699); Bill and Melinda Gates Foundation (INV-001754). NRW: Medical Research Council (grant number MR/N013638/1). RME: HDR UK (grant: MR/S003975/1), MRC (grant: MC_PC 19065), NIHR (grant: NIHR200908). NGD: UKRI Research England; NIHR Health Protection Research Unit in Immunisation (NIHR200929); UK MRC (MC_PC_19065). JYL: Bill & Melinda Gates Foundation (INV-003174). MK: Foreign, Commonwealth and Development Office / Wellcome Trust. FK: Innovation Fund of the Joint Federal Committee (Grant number 01VSF18015), Wellcome Trust (UNS110424). DCT: No funding declared. JDM: Wellcome Trust (grant: 210758/Z/18/Z). AS: No funding declared. AMF: No funding declared. KP: Gates (INV-003174), European Commission (101003688). SFlasche: Wellcome Trust (grant: 208812/Z/17/Z). AJK: Wellcome Trust (grant:

206250/Z/17/Z), NIHR (NIHR200908). SA: Wellcome Trust (grant: 210758/Z/18/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). TJ: RCUK/ESRC (grant: ES/P010873/1); UK PH RST; NIHR HPRU Modelling & Health Economics (NIHR200908). AR: NIHR (grant: PR-OD-1017-20002). GMK: UK Medical Research Council (grant: MR/P014658/1). MJ: Gates (INV-003174, INV-016832), NIHR (16/137/109, NIHR200929, NIHR200908), European Commission (EpiPose 101003688). YL: Gates (INV-003174), NIHR (16/137/109), European Commission (101003688). JW: NIHR Health Protection Research Unit and NIHR HTA. JH: Wellcome Trust (grant: 210758/Z/18/Z). KO'R: Bill and Melinda Gates Foundation (OPP1191821). YWDC: No funding declared. TWR: Wellcome Trust (grant: 206250/Z/17/Z). CIJ: Global Challenges Research Fund (GCRF) project 'RECAP' managed through RCUK and ESRC (ES/P010873/1). SRP: Bill and Melinda Gates Foundation (INV-016832). AE: The Nakajima Foundation. ESN: Gates (OPP1183986). NIB: Health Protection Research Unit (grant code NIHR200908). CJVA: European Research Council Starting Grant (Action number 757688). FGS: NIHR Health Protection Research Unit in Modelling & Health Economics, and in Immunisation. AG: European Commission (EpiPose 101003688). KA: Bill & Melinda Gates Foundation (OPP1157270, INV-016832). WW: MRC (grant MR/V027956/1). KEA: European Research Council Starting Grant (Action number 757688). RCB: European Commission (EpiPose 101003688). PK: This research was partly funded by the Royal Society under award RP\EA\180004, European Commission (101003688), Bill & Melinda Gates Foundation (INV-003174). 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 SocialCare (PR-OD-1017-20001). 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). OJB: Wellcome Trust (grant: 206471/Z/17/Z).

Ethical approval

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

Code and data availability

Analysis code is available at https://github.com/nicholasdavies/cfrvoc. An anonymised data set allowing replication of the analysis is available at the same URL.

573 **Tables**

574

575 **Table 1. Characteristics of study subjects, 1 November 2020–25 January 2021.**

576

	All N(%)	Missing N(%)	SGTF N(%) In subset with SGTF status measured	Non-SGTF N(%) In subset with SGTF status measured	SGTF prevalence N with SGTF/Total (%), in subset with SGTF status measured	Missingness N with missing SGTF status/Total (%)
	1,994,449 (100%)	966,153 (100%)	562,282 (100%)	466,014 (100%)	562,282 / 1,028,296 (54.7%)	966,153 / 1,994,449 (48.4%)
Sex						
Female	1,071,783 (53.7%)	533,309 (55.2%)	291,479 (51.8%)	246,995 (53%)	291,479 / 538,474 (54.1%)	533,309 / 1,071,783 (49.8%)
Male	922,666 (46.3%)	432,844 (44.8%)	270,803 (48.2%)	219,019 (47%)	270,803 / 489,822 (55.3%)	432,844 / 922,666 (46.9%)
Age in years						
1–34	886,034 (44.4%)	424,034 (43.9%)	253,832 (45.1%)	208,168 (44.7%)	253,832 / 462,000 (54.9%)	424,034 / 886,034 (47.9%)
35–54	684,762 (34.3%)	322,780 (33.4%)	201,019 (35.8%)	160,963 (34.5%)	201,019 / 361,982 (55.5%)	322,780 / 684,762 (47.1%)
55–69	300,329 (15.1%)	143,368 (14.8%)	84,213 (15%)	72,748 (15.6%)	84,213 / 156,961 (53.7%)	143,368 / 300,329 (47.7%)
70–84	86,320 (4.3%)	46,885 (4.9%)	19,372 (3.4%)	20,063 (4.3%)	19,372 / 39,435 (49.1%)	46,885 / 86,320 (54.3%)
85 and older	37,004 (1.9%)	29,086 (3%)	3,846 (0.7%)	4,072 (0.9%)	3,846 / 7,918 (48.6%)	29,086 / 37,004 (78.6%)
Place of residence						
Residential	1,868,902 (93.7%)	876,831 (90.8%)	542,875 (96.5%)	449,196 (96.4%)	542,875 / 992,071 (54.7%)	876,831 / 1,868,902 (46.9%)
Care/Nursing home	61,380 (3.1%)	54,681 (5.7%)	3,027 (0.5%)	3,672 (0.8%)	3,027 / 6,699 (45.2%)	54,681 / 61,380 (89.1%)
Other/Unknown	64,167 (3.2%)	34,641 (3.6%)	16,380 (2.9%)	13,146 (2.8%)	16,380 / 29,526 (55.5%)	34,641 / 64,167 (54%)
Ethnicity						
White	1,471,201 (73.8%)	704,038 (72.9%)	410,447 (73%)	356,716 (76.5%)	410,447 / 767,163 (53.5%)	704,038 / 1,471,201 (47.9%)
Asian	273,184 (13.7%)	124,762 (12.9%)	80,133 (14.3%)	68,289 (14.7%)	80,133 / 148,422 (54%)	124,762 / 273,184 (45.7%)

Black	93,848 (4.7%)	52,970 (5.5%)	27,480 (4.9%)	13,398 (2.9%)	27,480 / 40,878 (67.2%)	52,970 / 93,848 (56.4%)
Other/Mixed/Unknown	156,216 (7.8%)	84,383 (8.7%)	44,222 (7.9%)	27,611 (5.9%)	44,222 / 71,833 (61.6%)	84,383 / 156,216 (54%)
Index of Multiple Deprivation decile						
1	205,863 (10.3%)	75,437 (7.8%)	57,280 (10.2%)	73,146 (15.7%)	57,280 / 130,426 (43.9%)	75,437 / 205,863 (36.6%)
2	238,287 (11.9%)	113,291 (11.7%)	65,777 (11.7%)	59,219 (12.7%)	65,777 / 124,996 (52.6%)	113,291 / 238,287 (47.5%)
3	240,562 (12.1%)	119,894 (12.4%)	67,143 (11.9%)	53,525 (11.5%)	67,143 / 120,668 (55.6%)	119,894 / 240,562 (49.8%)
4	221,232 (11.1%)	111,768 (11.6%)	62,722 (11.2%)	46,742 (10%)	62,722 / 109,464 (57.3%)	111,768 / 221,232 (50.5%)
5	206,211 (10.3%)	105,116 (10.9%)	58,259 (10.4%)	42,836 (9.2%)	58,259 / 101,095 (57.6%)	105,116 / 206,211 (51%)
6	194,183 (9.7%)	99,921 (10.3%)	54,158 (9.6%)	40,104 (8.6%)	54,158 / 94,262 (57.5%)	99,921 / 194,183 (51.5%)
7	184,217 (9.2%)	92,220 (9.5%)	51,572 (9.2%)	40,425 (8.7%)	51,572 / 91,997 (56.1%)	92,220 / 184,217 (50.1%)
8	179,610 (9%)	88,611 (9.2%)	50,386 (9%)	40,613 (8.7%)	50,386 / 90,999 (55.4%)	88,611 / 179,610 (49.3%)
9	172,325 (8.6%)	85,417 (8.8%)	49,511 (8.8%)	37,397 (8%)	49,511 / 86,908 (57%)	85,417 / 172,325 (49.6%)
10	151,959 (7.6%)	74,478 (7.7%)	45,474 (8.1%)	32,007 (6.9%)	45,474 / 77,481 (58.7%)	74,478 / 151,959 (49%)
NHS England region						
East of England	250,910 (12.6%)	160,554 (16.6%)	68,444 (12.2%)	21,912 (4.7%)	68,444 / 90,356 (75.7%)	160,554 / 250,910 (64%)
London	467,366 (23.4%)	278,708 (28.8%)	139,622 (24.8%)	49,036 (10.5%)	139,622 / 188,658 (74%)	278,708 / 467,366 (59.6%)
Midlands	364,764 (18.3%)	161,896 (16.8%)	90,121 (16%)	112,747 (24.2%)	90,121 / 202,868 (44.4%)	161,896 / 364,764 (44.4%)
North East and Yorkshire	240,130 (12%)	53,697 (5.6%)	60,552 (10.8%)	125,881 (27%)	60,552 / 186,433 (32.5%)	53,697 / 240,130 (22.4%)
North West	231,160 (11.6%)	49,691 (5.1%)	77,642 (13.8%)	103,827 (22.3%)	77,642 / 181,469 (42.8%)	49,691 / 231,160 (21.5%)
South East	324,067 (16.2%)	179,493 (18.6%)	109,297 (19.4%)	35,277 (7.6%)	109,297 / 144,574 (75.6%)	179,493 / 324,067 (55.4%)

South West	116,052 (5.8%)	82,114 (8.5%)	16,604 (3%)	17,334 (3.7%)	16,604 / 33,938 (48.9%)	82,114 / 116,052 (70.8%)
Specimen date						
1 Nov–14 Nov	251,389 (12.6%)	86,954 (9%)	8,027 (1.4%)	156,408 (33.6%)	8,027 / 164,435 (4.9%)	86,954 / 251,389 (34.6%)
15 Nov–28 Nov	168,861 (8.5%)	57,694 (6%)	12,236 (2.2%)	98,931 (21.2%)	12,236 / 111,167 (11%)	57,694 / 168,861 (34.2%)
29 Nov–12 Dec	166,423 (8.3%)	61,600 (6.4%)	37,890 (6.7%)	66,933 (14.4%)	37,890 / 104,823 (36.1%)	61,600 / 166,423 (37%)
13 Dec–26 Dec	356,259 (17.9%)	186,943 (19.3%)	111,237 (19.8%)	58,079 (12.5%)	111,237 / 169,316 (65.7%)	186,943 / 356,259 (52.5%)
27 Dec–9 Jan	607,884 (30.5%)	336,667 (34.8%)	211,717 (37.7%)	59,500 (12.8%)	211,717 / 271,217 (78.1%)	336,667 / 607,884 (55.4%)
10 Jan–25 Jan	443,633 (22.2%)	236,295 (24.5%)	181,175 (32.2%)	26,163 (5.6%)	181,175 / 207,338 (87.4%)	236,295 / 443,633 (53.3%)

Table 2. Rates of death within 28 days of positive test among study subjects. Total number of deaths, number of days of followup, and deaths per 10,000 days of followup reported.

	All	Missing SGTF status	SGTF	Non-SGTF
	12,790 / 43,774,085 (2.92)	9,408 / 20,572,452 (4.57)	1,722 / 10,961,652 (1.57)	1,660 / 12,239,982 (1.36)
Sex				
Female	6,733 / 23,564,628 (2.86)	5,293 / 11,396,647 (4.64)	714 / 5,682,724 (1.26)	726 / 6,485,256 (1.12)
Male	6,057 / 20,209,458 (3)	4,115 / 9,175,804 (4.48)	1,008 / 5,278,928 (1.91)	934 / 5,754,726 (1.62)
Age				
1–34	50 / 19,706,054 (0.03)	20 / 9,177,761 (0.02)	16 / 5,026,025 (0.03)	14 / 5,502,268 (0.03)
35–54	512 / 15,077,156 (0.34)	244 / 6,891,346 (0.35)	165 / 3,947,176 (0.42)	103 / 4,238,634 (0.24)
55–69	1,533 / 6,490,368 (2.36)	775 / 3,011,166 (2.57)	454 / 1,583,118 (2.87)	304 / 1,896,083 (1.6)
70–84	4,364 / 1,818,684 (24)	3,025 / 956,804 (31.62)	656 / 351,155 (18.68)	683 / 510,724 (13.37)
85 and older	6,331 / 681,824 (92.85)	5,344 / 535,374 (99.82)	431 / 54,177 (79.55)	556 / 92,272 (60.26)
Place of residence				
Residential	4,890 / 41,205,718 (1.19)	2,271 / 18,777,676 (1.21)	1,422 / 10,615,334 (1.34)	1,197 / 11,812,709 (1.01)
Care/Nursing home	7,664 / 1,202,997 (63.71)	6,941 / 1,081,248 (64.19)	279 / 39,006 (71.53)	444 / 82,744 (53.66)
Other/Unknown	236 / 1,365,370 (1.73)	196 / 713,528 (2.75)	21 / 307,312 (0.68)	19 / 344,530 (0.55)
Ethnicity				
White	11,340 / 32,415,402 (3.5)	8,557 / 15,118,520 (5.66)	1,370 / 7,959,838 (1.72)	1,413 / 9,337,044 (1.51)
Asian	887 / 6,020,388 (1.47)	472 / 2,630,158 (1.79)	236 / 1,569,368 (1.5)	179 / 1,820,862 (0.98)
Black	227 / 1,960,509 (1.16)	137 / 1,065,071 (1.29)	63 / 544,514 (1.16)	27 / 350,924 (0.77)
Other/Mixed/Unknown	336 / 3,377,786 (0.99)	242 / 1,758,702 (1.38)	53 / 887,932 (0.6)	41 / 731,152 (0.56)
Index of Multiple Deprivation decile				

1	1,211 / 4,415,279 (2.74)	666 / 1,572,214 (4.24)	223 / 949,311 (2.35)	322 / 1,893,754 (1.7)
2	1,284 / 5,142,166 (2.5)	848 / 2,362,174 (3.59)	213 / 1,229,998 (1.73)	223 / 1,549,994 (1.44)
3	1,266 / 5,223,651 (2.42)	886 / 2,522,071 (3.51)	177 / 1,295,302 (1.37)	203 / 1,406,278 (1.44)
4	1,381 / 4,793,728 (2.88)	1,034 / 2,348,872 (4.4)	197 / 1,217,608 (1.62)	150 / 1,227,248 (1.22)
5	1,341 / 4,515,820 (2.97)	1,060 / 2,227,180 (4.76)	160 / 1,159,819 (1.38)	121 / 1,128,821 (1.07)
6	1,346 / 4,257,826 (3.16)	1,025 / 2,122,558 (4.83)	182 / 1,079,483 (1.69)	139 / 1,055,786 (1.32)
7	1,256 / 4,088,424 (3.07)	995 / 1,988,609 (5)	141 / 1,033,880 (1.36)	120 / 1,065,936 (1.13)
8	1,284 / 4,006,295 (3.2)	991 / 1,921,732 (5.16)	161 / 1,014,684 (1.59)	132 / 1,069,878 (1.23)
9	1,308 / 3,878,992 (3.37)	1,045 / 1,864,794 (5.6)	132 / 1,023,407 (1.29)	131 / 990,790 (1.32)
10	1,113 / 3,451,904 (3.22)	858 / 1,642,247 (5.22)	136 / 958,161 (1.42)	119 / 851,496 (1.4)
NHS England region				
East of England	1,783 / 5,511,951 (3.23)	1,527 / 3,471,280 (4.4)	183 / 1,455,449 (1.26)	73 / 585,222 (1.25)
London	1,426 / 10,377,194 (1.37)	1,073 / 5,914,652 (1.81)	281 / 3,127,714 (0.9)	72 / 1,334,828 (0.54)
Midlands	2,615 / 7,840,529 (3.34)	1,868 / 3,284,336 (5.69)	326 / 1,575,152 (2.07)	421 / 2,981,042 (1.41)
North East and Yorkshire	1,729 / 5,542,494 (3.12)	858 / 1,228,923 (6.98)	271 / 1,000,771 (2.71)	600 / 3,312,800 (1.81)
North West	1,318 / 4,904,162 (2.69)	713 / 1,116,972 (6.38)	259 / 1,157,800 (2.24)	346 / 2,629,389 (1.32)
South East	2,801 / 7,142,670 (3.92)	2,391 / 3,859,346 (6.2)	336 / 2,341,873 (1.43)	74 / 941,450 (0.79)
South West	1,118 / 2,455,085 (4.55)	978 / 1,696,942 (5.76)	66 / 302,894 (2.18)	74 / 455,250 (1.63)
Specimen date				
1 Nov–14 Nov	1,449 / 7,017,250 (2.06)	907 / 2,420,610 (3.75)	20 / 224,495 (0.89)	522 / 4,372,145 (1.19)
15 Nov–28 Nov	1,257 / 4,708,381 (2.67)	922 / 1,600,898 (5.76)	25 / 342,278 (0.73)	310 / 2,765,204 (1.12)

29 Nov–12 Dec	1,402 / 4,638,356 (3.02)	1,013 / 1,708,928 (5.93)	124 / 1,059,364 (1.17)	265 / 1,870,064 (1.42)
13 Dec–26 Dec	2,078 / 9,944,324 (2.09)	1,514 / 5,211,339 (2.91)	349 / 3,110,010 (1.12)	215 / 1,622,976 (1.32)
27 Dec–9 Jan	4,706 / 13,706,868 (3.43)	3,636 / 7,635,678 (4.76)	814 / 4,709,516 (1.73)	256 / 1,361,674 (1.88)
10 Jan–25 Jan	1,898 / 3,758,906 (5.05)	1,416 / 1,994,996 (7.1)	390 / 1,515,990 (2.57)	92 / 247,919 (3.71)

583
584

585 **Table 3. Absolute 28-day mortality risk associated with SGTF, as expressed by case**
586 **fatality ratio (%) among individuals testing positive in the community.** The baseline (i.e.
587 original variant) absolute risk after 28 days post-test is derived using linked deaths for all
588 individuals testing positive in the community from 1 August – 31 October 2020. Results
589 presented for both complete cases and IPW analysis.
590

Sex	Age	Baseline CFR	Variant CFR (complete cases)	Variant CFR (IPW)
Female	0-34	0.00069%	0.0011% (0.00096-0.0012)	0.0012% (0.001-0.0013)
Female	35-54	0.033%	0.052% (0.045-0.058)	0.054% (0.047-0.062)
Female	55-69	0.18%	0.29% (0.25-0.32)	0.3% (0.26-0.34)
Female	70-84	2.9%	4.5% (4-5.1)	4.7% (4.1-5.4)
Female	85 and older	13%	20% (17-22)	20% (18-23)
Male	0-34	0.0031%	0.0048% (0.0042-0.0055)	0.0051% (0.0044-0.0058)
Male	35-54	0.063%	0.099% (0.087-0.11)	0.1% (0.09-0.12)
Male	55-69	0.56%	0.88% (0.77-0.99)	0.93% (0.8-1)
Male	70-84	4.7%	7.3% (6.4-8.2)	7.6% (6.7-8.6)
Male	85 and older	17%	26% (23-28)	27% (24-30)

591
592

Supplementary tables

Table S1. Rates of death within any time period following positive test among study subjects, including missing SGTF status. Total number of deaths, number of days of followup, and deaths per 10,000 days of followup reported.

	All	Missing	SGTF	Non-SGTF
	13,860 / 69,160,118 (2)	10,127 / 29,806,446 (3.4)	1,785 / 13,504,310 (1.32)	1,948 / 25,849,361 (0.75)
Sex				
Female	7,314 / 37,253,988 (1.96)	5,729 / 16,563,712 (3.46)	734 / 7,008,795 (1.05)	851 / 13,681,482 (0.62)
Male	6,546 / 31,906,130 (2.05)	4,398 / 13,242,734 (3.32)	1,051 / 6,495,516 (1.62)	1,097 / 12,167,880 (0.9)
Age				
1–34	55 / 31,260,216 (0.02)	20 / 13,348,169 (0.01)	16 / 6,266,754 (0.03)	19 / 11,645,294 (0.02)
35–54	570 / 23,686,455 (0.24)	262 / 9,887,338 (0.26)	176 / 4,861,304 (0.36)	132 / 8,937,812 (0.15)
55–69	1,712 / 10,271,464 (1.67)	838 / 4,359,417 (1.92)	480 / 1,898,842 (2.53)	394 / 4,013,205 (0.98)
70–84	4,724 / 2,899,684 (16.29)	3,269 / 1,410,603 (23.17)	673 / 415,546 (16.2)	782 / 1,073,534 (7.28)
85 and older	6,799 / 1,042,298 (65.23)	5,738 / 800,919 (71.64)	440 / 61,864 (71.12)	621 / 179,516 (34.59)
Place of residence				
Residential	5,344 / 64,971,901 (0.82)	2,429 / 26,956,788 (0.9)	1,475 / 13,075,518 (1.13)	1,440 / 24,939,596 (0.58)
Care/Nursing home	8,271 / 2,010,503 (41.14)	7,495 / 1,800,672 (41.62)	287 / 46,928 (61.16)	489 / 162,904 (30.02)
Other/Unknown	245 / 2,177,714 (1.13)	203 / 1,048,988 (1.94)	23 / 381,865 (0.6)	19 / 746,862 (0.25)
Ethnicity				
White	12,289 / 51,755,410 (2.37)	9,218 / 22,149,052 (4.16)	1,416 / 9,840,646 (1.44)	1,655 / 19,765,712 (0.84)
Asian	964 / 9,523,947 (1.01)	500 / 3,775,242 (1.32)	249 / 1,903,515 (1.31)	215 / 3,845,190 (0.56)
Black	242 / 2,813,798 (0.86)	143 / 1,438,408 (0.99)	66 / 662,912 (1)	33 / 712,478 (0.46)
Other/Mixed/Unknown	365 / 5,066,964 (0.72)	266 / 2,443,744 (1.09)	54 / 1,097,238 (0.49)	45 / 1,525,982 (0.29)

Index of Multiple Deprivation decile				
1	1,346 / 7,642,162 (1.76)	742 / 2,532,840 (2.93)	229 / 1,104,186 (2.07)	375 / 4,005,136 (0.94)
2	1,412 / 8,218,058 (1.72)	923 / 3,439,324 (2.68)	224 / 1,492,162 (1.5)	265 / 3,286,572 (0.81)
3	1,362 / 8,143,723 (1.67)	948 / 3,584,962 (2.64)	182 / 1,593,884 (1.14)	232 / 2,964,876 (0.78)
4	1,484 / 7,437,915 (2)	1,108 / 3,352,900 (3.3)	202 / 1,498,640 (1.35)	174 / 2,586,375 (0.67)
5	1,453 / 6,998,806 (2.08)	1,134 / 3,179,536 (3.57)	170 / 1,447,943 (1.17)	149 / 2,371,327 (0.63)
6	1,442 / 6,598,334 (2.19)	1,095 / 3,035,862 (3.61)	185 / 1,343,834 (1.38)	162 / 2,218,638 (0.73)
7	1,383 / 6,420,691 (2.15)	1,083 / 2,874,795 (3.77)	145 / 1,289,344 (1.12)	155 / 2,256,552 (0.69)
8	1,366 / 6,290,169 (2.17)	1,047 / 2,768,153 (3.78)	168 / 1,260,494 (1.33)	151 / 2,261,522 (0.67)
9	1,410 / 6,051,438 (2.33)	1,122 / 2,679,682 (4.19)	138 / 1,277,721 (1.08)	150 / 2,094,036 (0.72)
10	1,202 / 5,358,822 (2.24)	925 / 2,358,393 (3.92)	142 / 1,196,100 (1.19)	135 / 1,804,328 (0.75)
NHS England region				
East of England	1,867 / 7,696,944 (2.43)	1,586 / 4,694,240 (3.38)	191 / 1,818,968 (1.05)	90 / 1,183,736 (0.76)
London	1,518 / 14,484,721 (1.05)	1,127 / 7,683,234 (1.47)	303 / 4,019,058 (0.75)	88 / 2,782,428 (0.32)
Midlands	2,900 / 13,471,944 (2.15)	2,063 / 5,496,554 (3.75)	331 / 1,772,414 (1.87)	506 / 6,202,975 (0.82)
North East and Yorkshire	1,954 / 10,587,604 (1.85)	981 / 2,121,664 (4.62)	276 / 1,159,505 (2.38)	697 / 7,306,436 (0.95)
North West	1,475 / 8,495,294 (1.74)	811 / 1,790,338 (4.53)	265 / 1,241,257 (2.13)	399 / 5,463,698 (0.73)
South East	2,955 / 10,426,000 (2.83)	2,517 / 5,320,110 (4.73)	353 / 3,131,366 (1.13)	85 / 1,974,524 (0.43)
South West	1,191 / 3,997,611 (2.98)	1,042 / 2,700,304 (3.86)	66 / 361,742 (1.82)	83 / 935,566 (0.89)
Specimen date				
1 Nov–14 Nov	1,846 / 19,613,076 (0.94)	1,159 / 6,742,156 (1.72)	24 / 621,195 (0.39)	663 / 12,249,725 (0.54)

15 Nov–28 Nov	1,533 / 10,949,347 (1.4)	1,118 / 3,709,086 (3.01)	32 / 780,457 (0.41)	383 / 6,459,804 (0.59)
29 Nov–12 Dec	1,625 / 8,237,496 (1.97)	1,166 / 3,025,100 (3.85)	141 / 1,843,366 (0.76)	318 / 3,369,030 (0.94)
13 Dec–26 Dec	2,251 / 12,857,434 (1.75)	1,631 / 6,680,294 (2.44)	384 / 4,020,394 (0.96)	236 / 2,156,746 (1.09)
27 Dec–9 Jan	4,707 / 13,743,859 (3.42)	3,637 / 7,654,812 (4.75)	814 / 4,722,910 (1.72)	256 / 1,366,137 (1.87)
10 Jan–25 Jan	1,898 / 3,758,906 (5.05)	1,416 / 1,994,996 (7.1)	390 / 1,515,990 (2.57)	92 / 247,919 (3.71)

599

Table S2. Hazard ratios for SGTF / VOC across models.

	parameter	HR	95% LCL	95% UCL	Death type	Marker	Age term	IMD term	start date	end date	reg. cutoff (days)	strata	xvars	weighting
Death type														
	sgtf	1.61	1.42	1.84	c28	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.64	1.42	1.88	c28	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.25	0.99	1.58	d07	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.57	1.18	2.08	d07	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.40	1.20	1.62	d14	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.55	1.32	1.83	d14	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.54	1.35	1.76	d21	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.66	1.44	1.92	d21	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.58	1.40	1.79	d28	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.67	1.46	1.90	d28	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.54	1.37	1.74	d60	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.64	1.43	1.87	d60	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.54	1.37	1.74	dNA	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.63	1.43	1.87	dNA	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.54	1.37	1.74	e60	SGTF	S	S	2020-11-01		10	LTLA:date		cc

	sgtf	1.64	1.43	1.87	e60	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
Misclassification adjustment														
	p_voc	1.65	1.44	1.88	d28	pVOC	S	S	2020-09-01		10	LTLA:date		cc
	p_voc	1.72	1.49	1.99	d28	pVOC	S	S	2020-09-01		10	LTLA:date		ipw
	p_voc	1.63	1.44	1.86	d28	pVOC	S	S	2020-11-01		10	LTLA:date		cc
	p_voc	1.71	1.48	1.97	d28	pVOC	S	S	2020-11-01		10	LTLA:date		ipw
	p_voc	1.60	1.41	1.82	d60	pVOC	S	S	2020-11-01		10	LTLA:date		cc
	p_voc	1.69	1.46	1.95	d60	pVOC	S	S	2020-11-01		10	LTLA:date		ipw
Geographical and temporal stratification														
	sgtf	1.58	1.40	1.79	d28	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.67	1.46	1.90	d28	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.52	1.36	1.69	d28	SGTF	S	S	2020-11-01		10	LTLA:week		cc
	sgtf	1.63	1.44	1.85	d28	SGTF	S	S	2020-11-01		10	LTLA:week		ipw
	sgtf	1.49	1.36	1.64	d28	SGTF	S	S	2020-11-01		10	NHSE:date		cc
	sgtf	1.47	1.28	1.68	d28	SGTF	S	S	2020-11-01		10	NHSE:date		ipw
	sgtf	1.50	1.37	1.65	d28	SGTF	S	S	2020-11-01		10	NHSE:week		cc
	sgtf	1.48	1.28	1.70	d28	SGTF	S	S	2020-11-01		10	NHSE:week		ipw
	sgtf	1.56	1.39	1.75	d28	SGTF	S	S	2020-11-01		10	UTLA:date		cc

	sgtf	1.56	1.37	1.78	d28	SGTF	S	S	2020-11-01		10	UTLA:date		ipw
	sgtf	1.51	1.36	1.67	d28	SGTF	S	S	2020-11-01		10	UTLA:week		cc
	sgtf	1.52	1.33	1.74	d28	SGTF	S	S	2020-11-01		10	UTLA:week		ipw
Age and IMD terms (linear vs. spline)														
	sgtf	1.58	1.40	1.79	d28	SGTF	L	L	2020-11-01		10	LTLA:date		cc
	sgtf	1.68	1.47	1.92	d28	SGTF	L	L	2020-11-01		10	LTLA:date		ipw
	sgtf	1.59	1.41	1.80	d28	SGTF	L	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.69	1.47	1.93	d28	SGTF	L	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.57	1.39	1.78	d28	SGTF	S	L	2020-11-01		10	LTLA:date		cc
	sgtf	1.66	1.45	1.89	d28	SGTF	S	L	2020-11-01		10	LTLA:date		ipw
	sgtf	1.58	1.40	1.79	d28	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.67	1.46	1.90	d28	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
By week since specimen														
	sgtf	1.26	1.00	1.59	d00-07	SGTF	S	S	2020-11-01		10	LTLA:date	sgtf_by_week	cc
	sgtf	1.56	1.17	2.08	d00-07	SGTF	S	S	2020-11-01		10	LTLA:date	sgtf_by_week	ipw
	sgtf	1.48	1.22	1.79	d08-14	SGTF	S	S	2020-11-01		10	LTLA:date	sgtf_by_week	cc
	sgtf	1.53	1.28	1.83	d08-14	SGTF	S	S	2020-11-01		10	LTLA:date	sgtf_by_week	ipw
	sgtf	2.16	1.63	2.85	d15-21	SGTF	S	S	2020-11-01		10	LTLA:date	sgtf_by_week	cc

	sgtf	2.18	1.63	2.91	d15-21	SGTF	S	S	2020-11-01		10	LTLA:date	sgtf_by_week	ipw
	sgtf	2.01	1.40	2.88	d22-28	SGTF	S	S	2020-11-01		10	LTLA:date	sgtf_by_week	cc
	sgtf	1.75	1.11	2.75	d22-28	SGTF	S	S	2020-11-01		10	LTLA:date	sgtf_by_week	ipw
Analysis start date														
	sgtf	1.55	1.37	1.75	d28	SGTF	S	S	2020-09-01		10	LTLA:date		cc
	sgtf	1.65	1.44	1.88	d28	SGTF	S	S	2020-09-01		10	LTLA:date		ipw
	sgtf	1.55	1.37	1.75	d28	SGTF	S	S	2020-10-01		10	LTLA:date		cc
	sgtf	1.65	1.44	1.88	d28	SGTF	S	S	2020-10-01		10	LTLA:date		ipw
	sgtf	1.58	1.40	1.79	d28	SGTF	S	S	2020-11-01		10	LTLA:date		cc
	sgtf	1.67	1.46	1.90	d28	SGTF	S	S	2020-11-01		10	LTLA:date		ipw
	sgtf	1.62	1.42	1.85	d28	SGTF	S	S	2020-12-01		10	LTLA:date		cc
	sgtf	1.69	1.46	1.95	d28	SGTF	S	S	2020-12-01		10	LTLA:date		ipw
	sgtf	1.69	1.39	2.04	d28	SGTF	S	S	2021-01-01		10	LTLA:date		cc
	sgtf	1.85	1.47	2.32	d28	SGTF	S	S	2021-01-01		10	LTLA:date		ipw
	sgtf	1.59	1.41	1.81	d28	SGTF	S	S	LTLA		10	LTLA:date		cc
	sgtf	1.68	1.47	1.93	d28	SGTF	S	S	LTLA		10	LTLA:date		ipw
Covariate interactions with time since positive test														
	sgtf	1.59	1.41	1.80	d28	SGTF	L	S	2020-11-01		10	LTLA:date	age:tstop	cc

	sgtf	1.68	1.47	1.93	d28	SGTF	L	S	2020-11-01		10	LTLA:date	age:tstop	ipw
	sgtf	1.58	1.40	1.79	d28	SGTF	S	S	2020-11-01		10	LTLA:date	sex:tstop	cc
	sgtf	1.67	1.46	1.91	d28	SGTF	S	S	2020-11-01		10	LTLA:date	sex:tstop	ipw
	sgtf	1.58	1.39	1.78	d28	SGTF	S	L	2020-11-01		10	LTLA:date	imd:tstop	cc
	sgtf	1.66	1.45	1.90	d28	SGTF	S	L	2020-11-01		10	LTLA:date	imd:tstop	ipw
	sgtf	1.58	1.40	1.79	d28	SGTF	S	S	2020-11-01		10	LTLA:date	eth:tstop	cc
	sgtf	1.67	1.46	1.90	d28	SGTF	S	S	2020-11-01		10	LTLA:date	eth:tstop	ipw
	sgtf	1.58	1.40	1.79	d28	SGTF	S	S	2020-11-01		10	LTLA:date	res:tstop	cc
	sgtf	1.67	1.46	1.91	d28	SGTF	S	S	2020-11-01		10	LTLA:date	res:tstop	ipw
No registration cutoff														
	sgtf	1.59	1.42	1.78	d28	SGTF	S	S	2020-11-01		0	LTLA:date		cc
	sgtf	1.61	1.42	1.83	d28	SGTF	S	S	2020-11-01		0	LTLA:date		ipw
Adjustment for, not stratification by, region and time														
	sgtf	1.50	1.37	1.65	d28	SGTF	S	S	2020-11-02	2020-01-24	10		NHSE:week	cc
	sgtf	1.48	1.28	1.71	d28	SGTF	S	S	2020-11-02	2020-01-24	10		NHSE:week	ipw
Subjects with full 28-day follow-up only														
	sgtf	1.46	1.22	1.74	d28	SGTF	S	S	2020-11-01	T - 38	0	LTLA:date		cc

	sgtf	1.52	1.28	1.79	d28	SGTF	S	S	2020-11-01	T - 38	0	LTLA:date		ipw
Asymptomatic screening indicator included as covariate														
	sgtf	1.58	1.39	1.79	d28	SGTF	S	S	2020-11-01		10	LTLA:date	asymptomatic	cc
	sgtf	1.66	1.45	1.90	d28	SGTF	S	S	2020-11-01		10	LTLA:date	asymptomatic	ipw
Alternative misclassification adjustment														
	p_voc	1.64	1.44	1.86	d28	pVOC2	S	S	2020-11-01		10	LTLA:date		cc
	p_voc	1.71	1.48	1.97	d28	pVOC2	S	S	2020-11-01		10	LTLA:date		ipw

Supplementary Figures

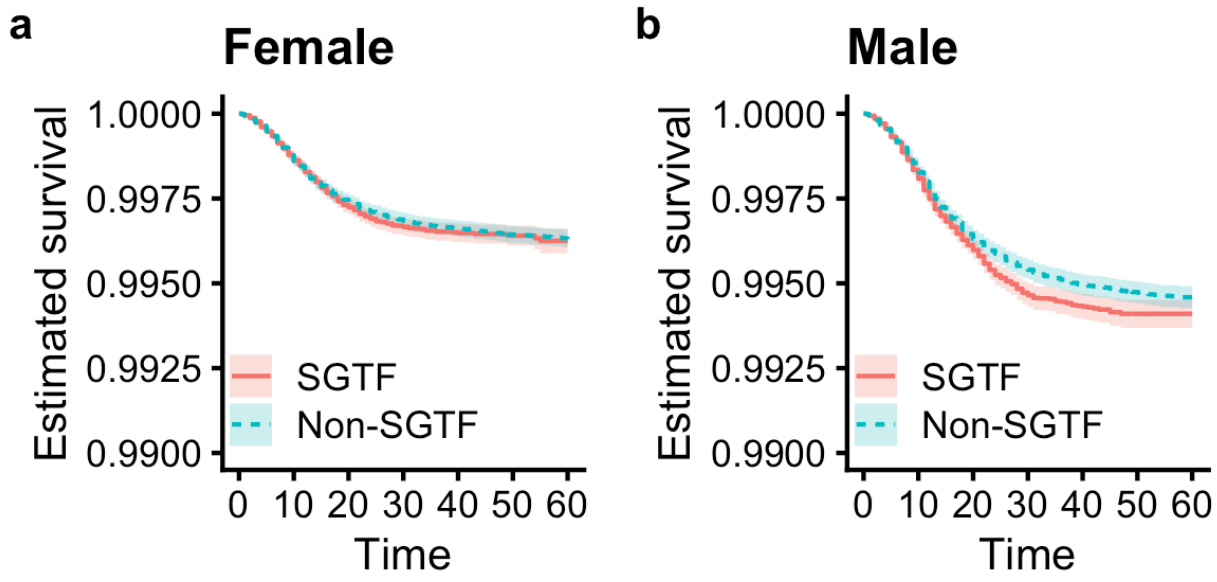

Fig. S1. Kaplan-Meier plots of survival within 60 days of positive test for SGTF versus non-SGTF by sex.

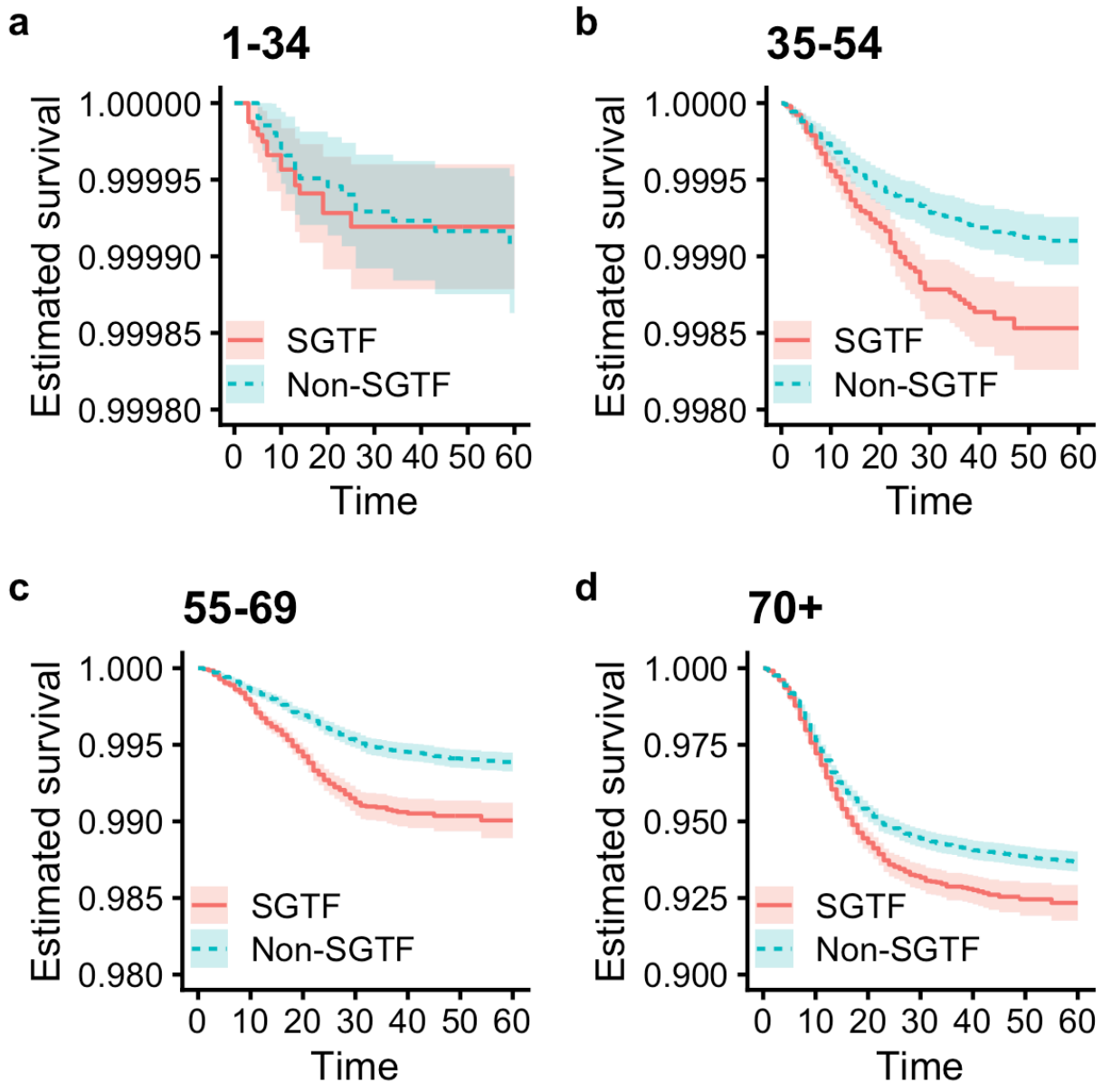

Fig. S2. Kaplan-Meier plots of survival within 60 days of positive test for SGTF versus non-SGTF by age group. Note that the Y axis differs for each panel.

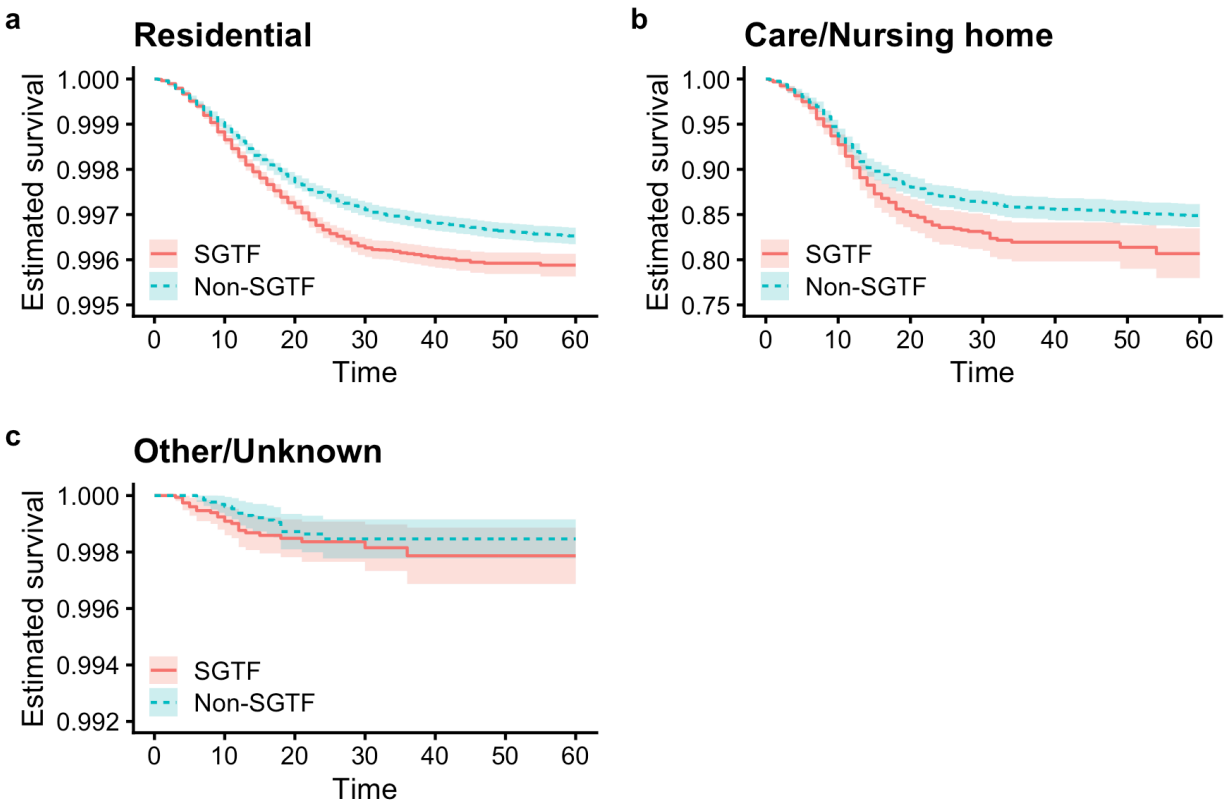

Fig. S3. Kaplan-Meier plots of survival within 60 days of positive test for SGTF versus non-SGTF by residence. Note that the Y axis differs for each panel.

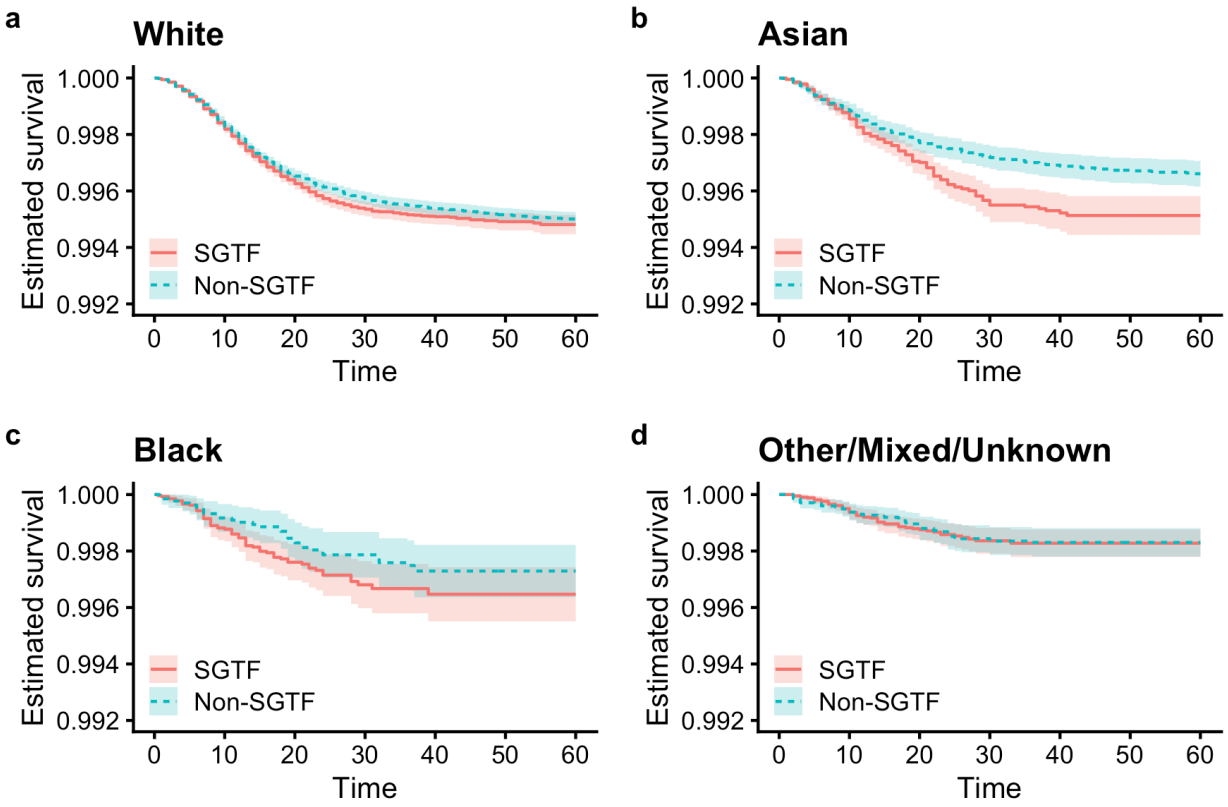

Fig. S4. Kaplan-Meier plots of survival within 60 days of positive test for SGTF versus non-SGTF by ethnicity.

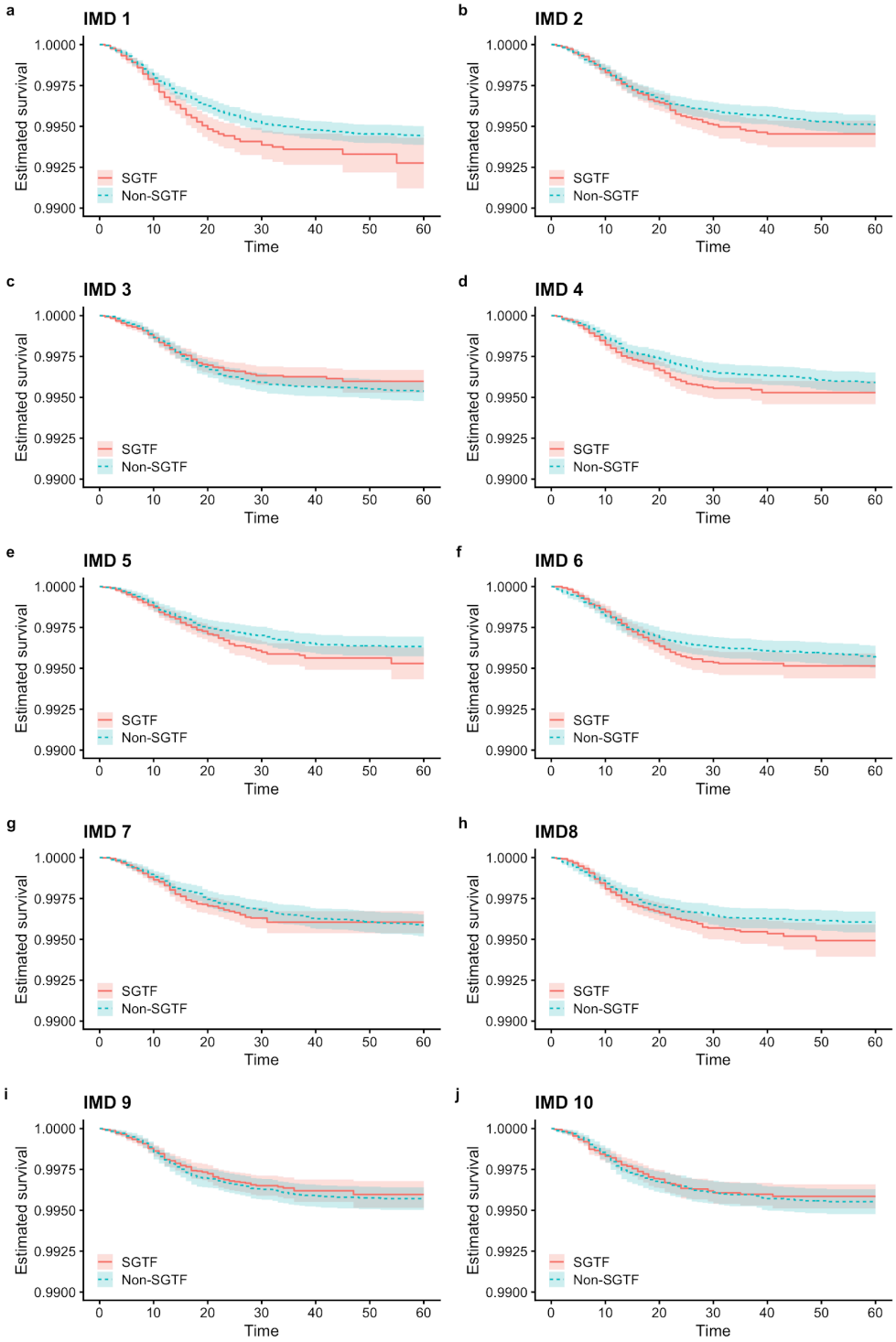

Fig. S5. Kaplan-Meier plots of survival within 60 days of positive test for SGTF versus non-SGTF by Index of Multiple Deprivation (IMD) decile.

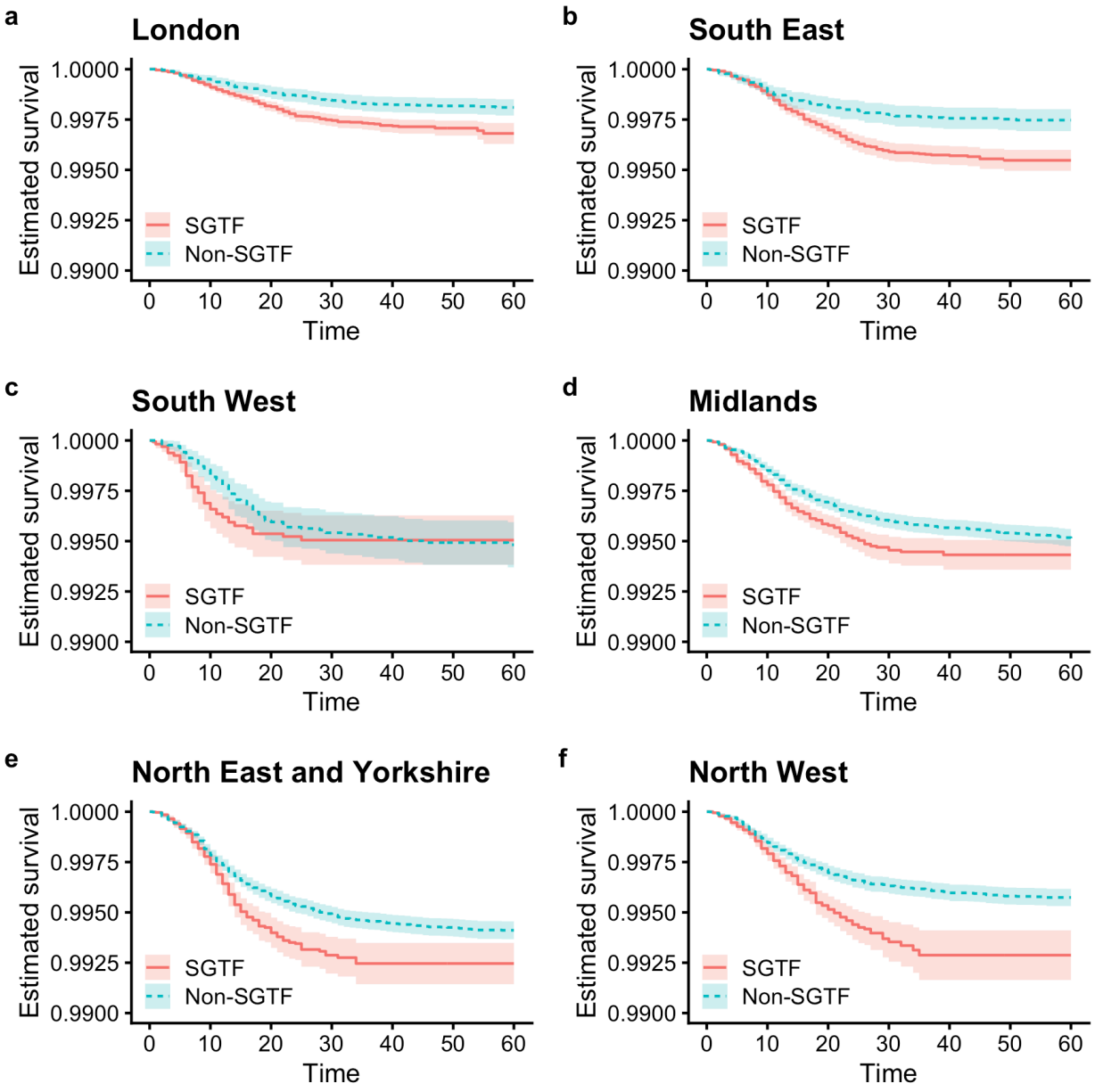

Fig. S6. Kaplan-Meier plots of survival within 60 days of positive test for SGTF versus non-SGTF by NHS region.

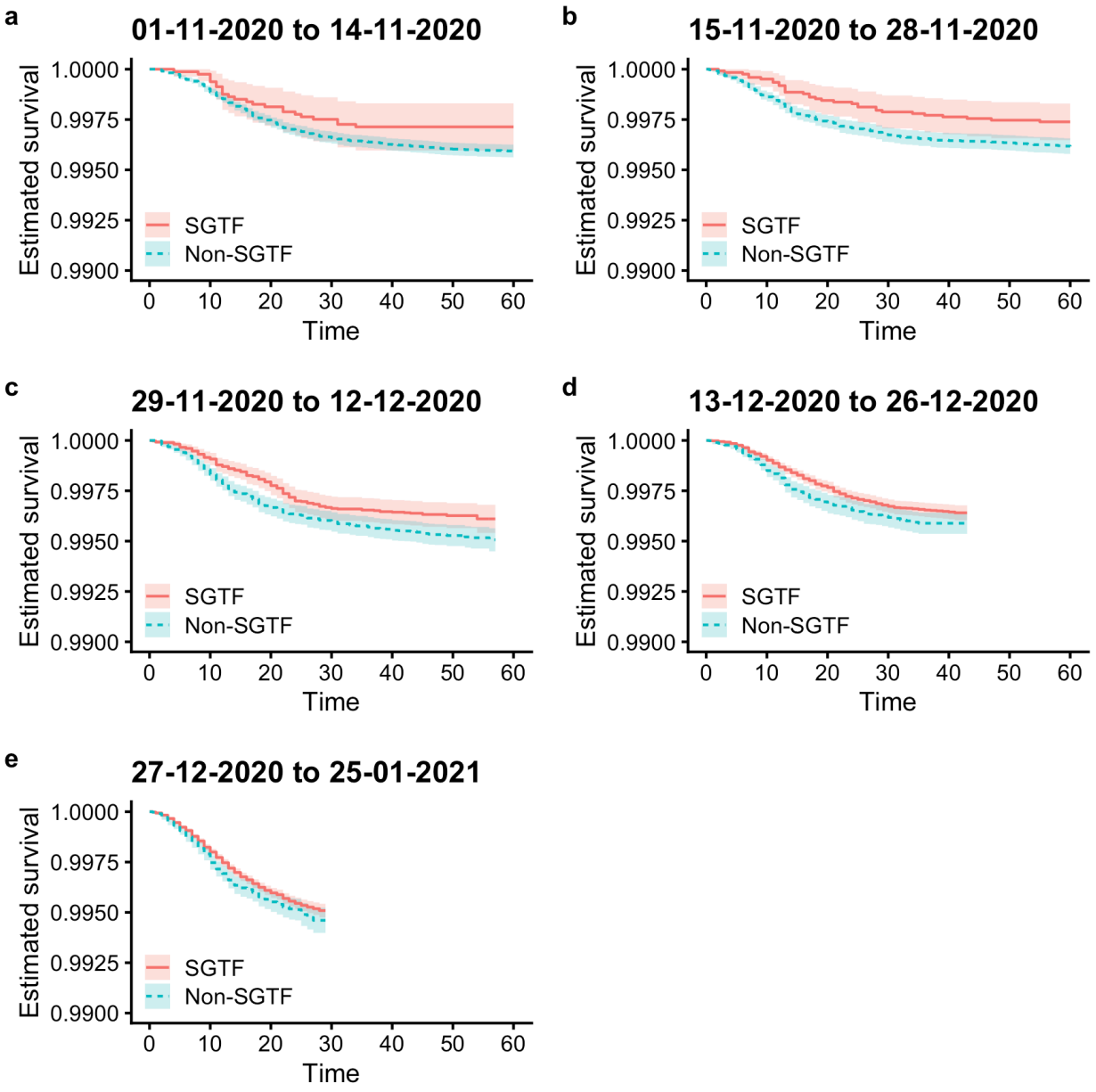

Fig. S7. Kaplan-Meier plots of survival within 60 days of positive test for SGTF versus non-SGTF by specimen date.

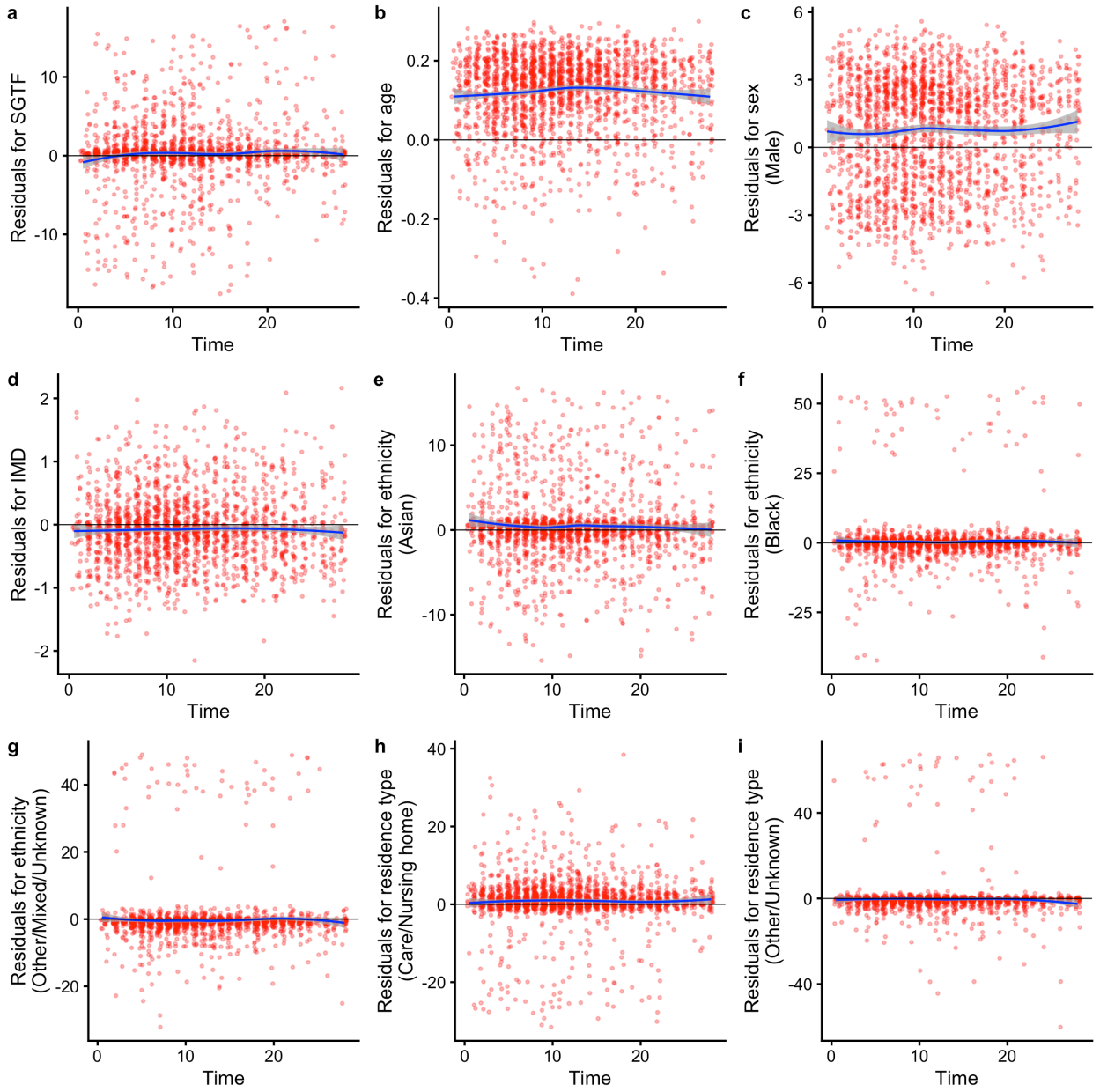

Fig. S8. Schoenfeld residuals for survival model by SGTF stratified by LTLA and specimen date . Model uses linear terms for age and IMD a 28-day followup using complete cases. Schoenfeld residual tests give P = 0.031 for SGTF; P = 0.425 for age; P = 0.170 for sex; P = 0.603 for IMD decile; P = 0.410 for ethnicity; P = 0.728 for residence type; and P = 0.244 globally.

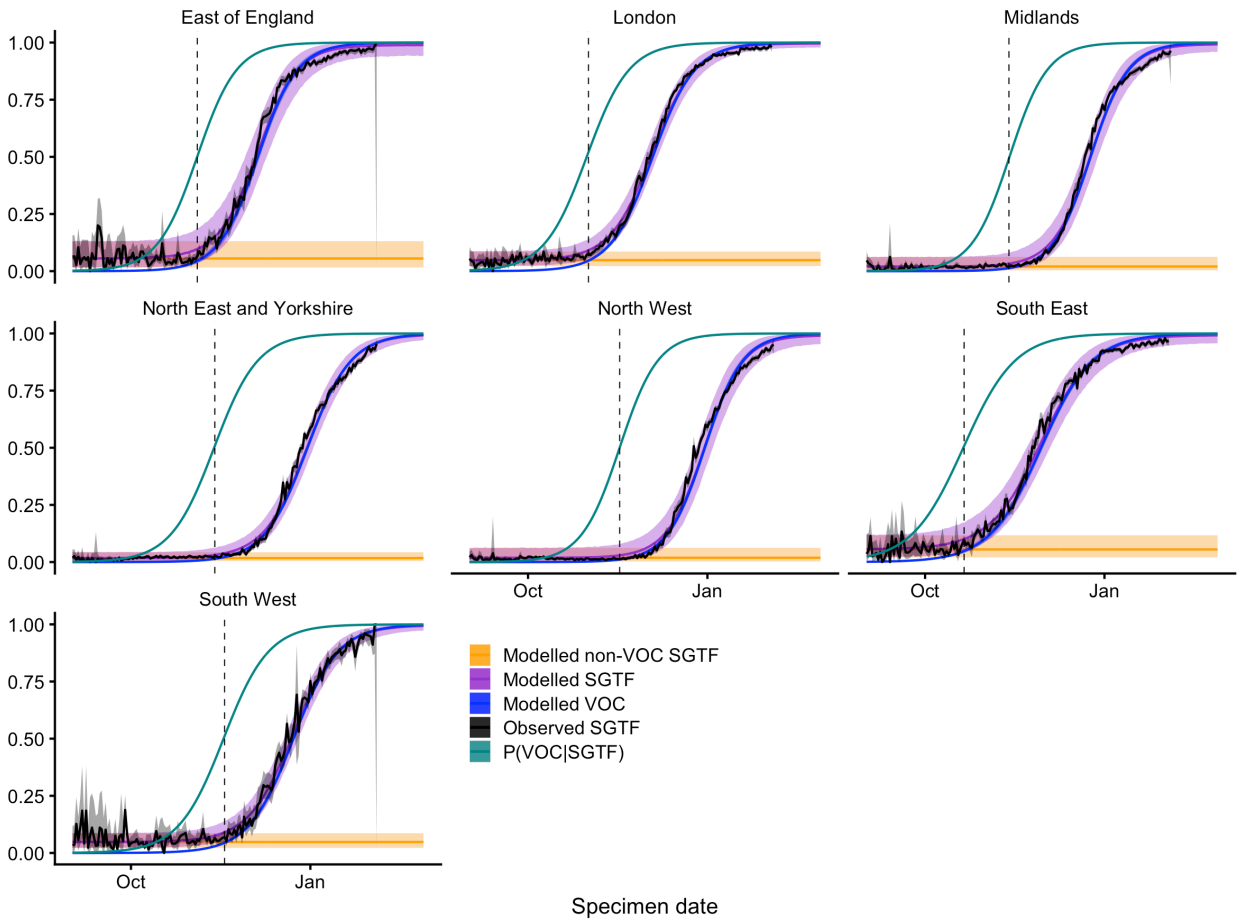

Fig. S9. Misclassification model. For each NHS England region, we fit a beta-binomial model (purple, Modelled SGTF) to the observed SGTF frequencies among Pillar 2 tests (black, Observed SGTF), which estimates a constant proportion of "false positive" SGTF samples among non-VOC 202012/01 specimens (orange, Modelled non-VOC SGTF) and a logistically growing proportion of VOC 202012/01 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(VOC|SGTF)). For our misclassification survival analysis, p_{VOC} = 0 for non-SGTF specimens and p_{VOC} = P(VOC|SGTF) for SGTF specimens.

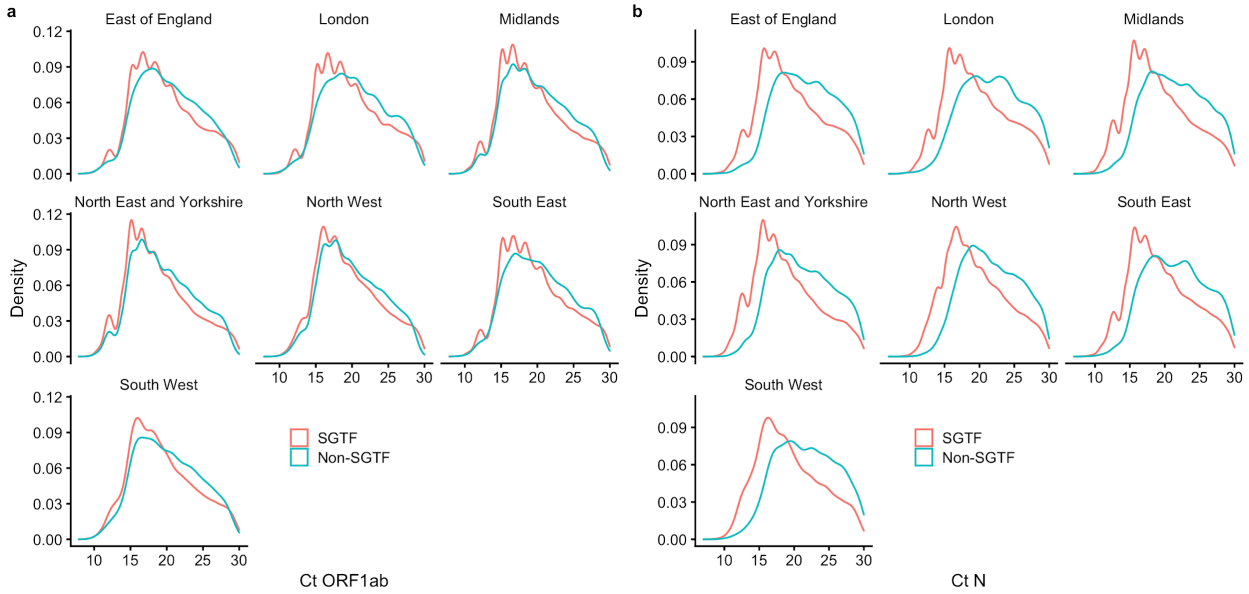

Fig. S10. Ct values for SGTF versus non-SGTF. The distribution of Ct values for (**a**) ORF1ab and (**b**) N gene targets among specimens collected between 1–25 January 2021.

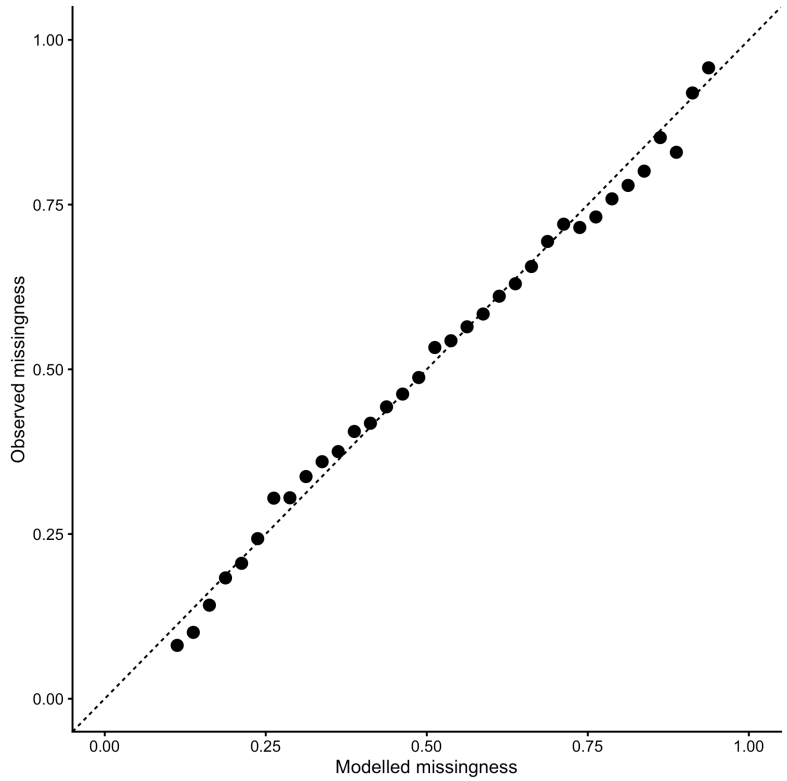

Fig. S11. Q-Q plot assessing the fit of the final missingness model (Cauchit link).

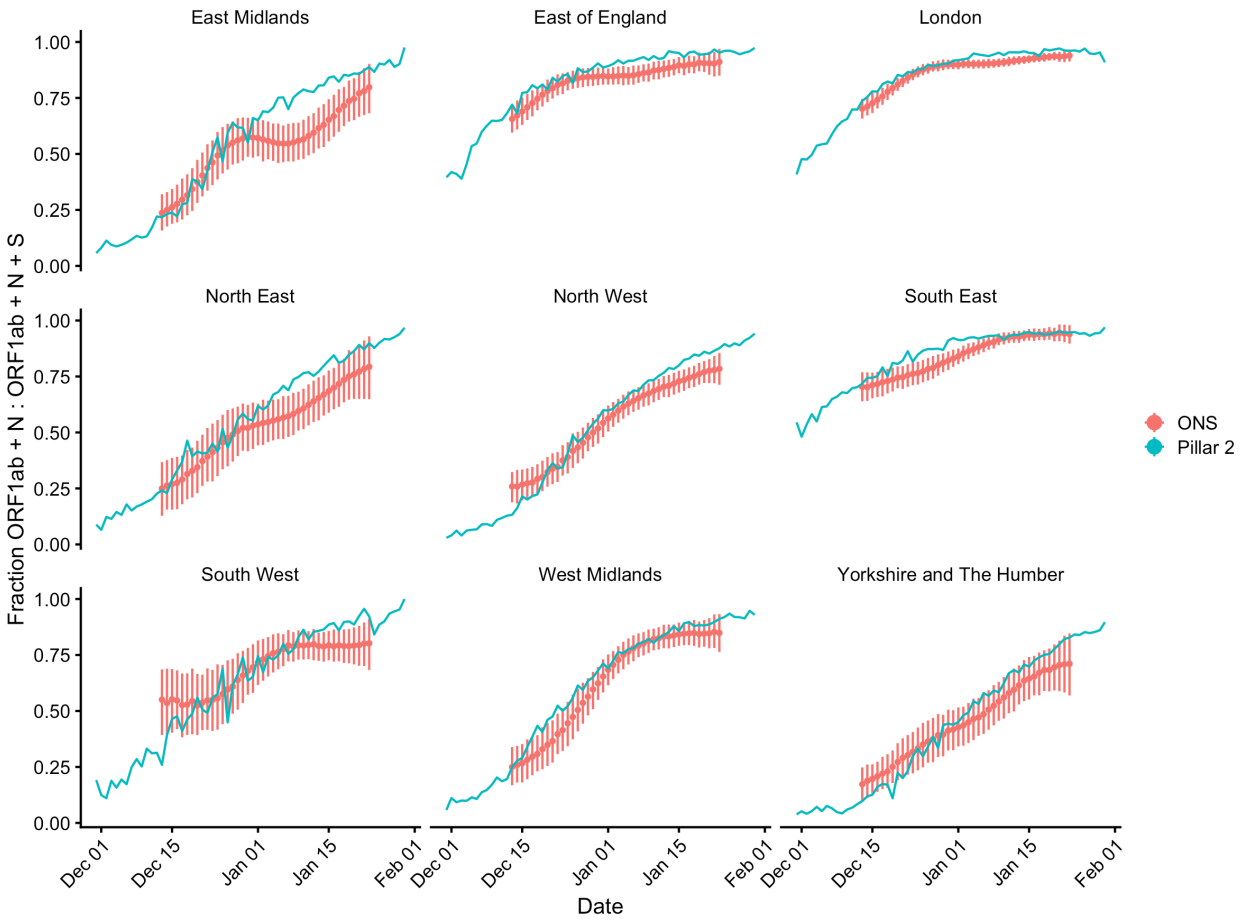

Fig. S12. Comparison of the proportion of samples with S gene dropout in our Pillar 2 sample (testing data) compared to ONS (random sampling of the community).

References

1. Ministry of Housing & Communities & Local Government. English indices of deprivation 2019: technical report. https://www.gov.uk/government/publications/english-indices-of-deprivation-2019-technical-report (2019).
2. Cox, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34**, 187–202 (1972).
3. Public Health England. *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 (2020).
4. 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).
5. Davies, N. G. *et al.* Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. *bioRxiv* (2020) doi:10.1101/2020.12.24.20248822.
6. Department of Health and Social Care. NERVTAG paper on COVID-19 variant of concern B.1.1.7. https://www.gov.uk/government/publications/nervtag-paper-on-covid-19-variant-of-concern-b117 (2021).
7. 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.
8. Seaman, S. R. & White, I. R. Review of inverse probability weighting for dealing with missing data. *Stat. Methods Med. Res.* **22**, 278–295 (2013).
9. Levin, A. T. *et al.* Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur. J. Epidemiol.* **35**, 1123–1138 (2020).

10. Lumley, T. & Scott, A. Partial likelihood ratio tests for the Cox model under complex sampling. *Stat. Med.* **32**, 110–123 (2013).
11. Home - COG-UK Consortium. https://www.cogconsortium.uk/.
12. Imperial College London. Microreact - Open data visualization and sharing for genomic epidemiology. https://microreact.org/.