Population-level association between S-gene target failure and the relationship between cases, hospitalisations and deaths of Covid-19

Work in progress - not peer reviewed

Sam Abbott, Sebastian Funk on behalf of the CMMID Covid-19 Working Group

29 January, 2021

For correspondence: sam.abbott@lshtm.ac.uk

Abstract

Background: Individual-level data indicates that the B.1.1.7 variant may be associated with increased case-fatality ratios but has so far provided inconclusive evidence on the relationship with hospitalisation rates, and hospitalisation-fatality ratios. Whilst an individual-level approach represents the most robust method for observational analysis population-level analysis may be helpful for triangulating findings, especially when individual data sources are confidential, difficult to obtain, or only partially representative. In this analysis, we use multiple approaches to evaluate public population-level data for evidence of an association between S-gene target failure (SGTF), a correlate of the B.1.1.7 variant, and Covid-19 case outcomes.

Method: We explored the association between the proportion of samples that were S-gene negative and the case-fatality ratio, case-hospitalisation ratio, and hospitalisation-fatality ratio of Covid-19 aggregated at the Upper-tier local authority (UTLA) and National Health Service England (NHSE) region level. Two approaches were used with the first assuming a fixed lag between cases/admissions and admissions/deaths with the lag optimised using the Pearson's correlation coefficient. The second approach assumed that admissions/deaths could be estimated using a convolution of cases/admissions and a distributional delay multiplied by a scaling factor. For the fixed lag analysis we investigated both additive and negative effects of SGFT and for the convolution approach we explored delays between observations that varied spatially. We present estimates from all approaches unadjusted for confounders as well as adjusted for location specific variation, and adjusted for case/admission burden.

Results: We found consistent evidence of an association between an increase in the population-level case-fatality ratio of Covid-19 and SGTF. At the ULTA level, using the fixed lag approach, we estimated that SGTF was associated with an additive increase in the case-fatality ratio of 0.016 (0.014, 0.017) or a multiplicative increase of 2.6 (2.3, 2.8). The convolution approach reduced the estimated effect to 2 (1.8, 2.1) at the UTLA level. There was also evidence for an association with an increase in population-level case-hospitalisation ratios (CHRs) and hospitalisation-fatality ratios (HFRs), with the effect on CHRs (1.7 (1.6, 1.8)) being larger than that on HFRs (1.2 (1.1, 1.3)) and more robust to the approach used and confounders adjusted for.

Conclusions: Population-level surveillance data supports findings from other studies using individual level data that SGTF, and by implication the B.1.1.7 variant, is associated with an increase in the case-fatality ratio. Our findings also indicate that the majority of this effect is likely linked to an increase in the case-hospitalisation ratio rather than the hospitalisation-fatality ratio. Our methods and code are available and may be extended or generalised to other research settings where individual data is not currently available.

Introduction

Preliminary evidence, based on individual level line list data, indicates that the B.1.1.7 variant is likely associated with an increased case fatality ratio.[1] However, this data represents only a fraction of notified Covid-19 cases and deaths and may suffer from bias due to potentially being unrepresentative of the wider population of Covid-19 cases. Whilst individual-level data represents the most robust method for observational analyses population-level analyses may be helpful for triangulating findings and provide initial evidence to guide further investigation.

In this study, we explore the evidence for an association between S-gene target failure (SGTF), a correlate of the B.1.1.7 variant, and case-fatality ratios, case-hospitalisation ratios, and hospitalisation-fatality ratios using population-level surveillance data aggregated at both the NHSE region and upper-tier local authority UTLA level. We make use of two approaches, with the first being to standardise the reference dates of weekly cases/admissions with admissions/deaths using am optimised fixed delay and the second being to assume that admissions/deaths can be predicted using a scaling (for deaths and cases this would be the case-fatality ratio) and a distributional delay from case/admission to death/admission. We explore both additive and multiplicative effects of SGTF and include the results from a range of models with differing assumptions. We aim to provide an additional evidence source to help triangulate the effect of the B.1.1.7 variant on case outcomes and provide a generalisable approach for studying case outcomes using population level data.

Method

Data

We used 4 main sources of data: test positive Covid-19 notifications by UTLA,[2] hospitalisations with Covid-19 by UTLA,[3] deaths linked to Covid-19 notification within 28 days of notification,[2] and S-gene status from PCR tests by local authority provided by Public Health England (PHE)[4]. We aggregated the data at the weekly, or daily, level and restricted the analysis to the period beginning Monday, 5 October. PCR testing data was only available aggregated to weeks by specimen date.

Statistical analysis

We calculated the weekly proportion of positive tests that were S-gene negative over time by local authority and NHS region. We estimated the proportion of tests that were S-gene positive by date of infection by shifting all estimates by date of specimen back by a week. We then conducted two analyses. In the first analysis lags between cases, admissions and deaths were estimated by maximising Pearson's correlation coefficient and all data was then adjusted using these lags to date of infection. In the second analysis the delay between observations (for example deaths and cases) was assumed to be log normal with this then being estimated in model either globally ("global convolution") or locally (i.e. at the NHS region or UTLA level) using a random effect ("local convolution"). All analyses were repeated at NHS region and upper-tier local authority (UTLA) scales. Further details of each analysis are given in the following sections.

Fixed lag analysis

We assumed that the observed number of Covid-19 hospitalisations or deaths within 28 days (henceforth: hospitalisations or deaths) that occurred on day t ($D_{i,t}$) were a function of Covid-19 case notifications or hospitalisations d days earlier, ($C_{i,t-d}$), scaled by the case-fatality ratio (or case-hospitalisation ratio/hospitalisation-fatality ratio) of S-gene positive cases/hospitalisations (c^+) and S-gene negative cases/hospitalisations (c^-),

$$D_{i,t+d} \sim \text{NB}\left(c^+_{i,w}\left(1-f_{i,w}\right)C_{i,t} + c^-_{i,t}f_{i,w}C_{i,t} + \epsilon, \phi\right)$$

where i indicates UTLA or NHS region, w the week containing day t, ϵ is an error term that accounts for imported deaths/admissions not linked to local cases/admissions, $f_{i,w}$ is the fraction of cases that were found to be S-gene negative by UTLA each week, and ϕ is the overdispersion of admissions/deaths. The case-fatality ratio (or hospitalisation-fatality / case-hospitalisation ratio, respectively) of S-gene negative

cases was then assumed to be a function of local variation, and case/admissions burden reflecting additional mortality that might occur due to hospital pressure,

$$c^+_{i,t} = \text{logit}^{-1}\left(\gamma_i + \beta C_{i,t}\right)$$

where γ_i is an intercept corresponding to the baseline case-fatality ratio per UTLA or NHS region, and β corresponds to the effect of case burden with non-linearity incorporated using a thin plate spline. The S-gene negative case-fatality ratio was then assumed to be related to the S-gene positive case-fatality ratio via a multiplicative relationship,

$$c^-_{i,t} = \alpha c^+_{i,t}$$

or an additive relationship

$$c^-_{i,t} = \alpha + c^+_{i,t}$$

where α represents either the multiplicative change in case-fatality ratio or the additive change. These alternative parametrisations represent either a population wide effect for the former parametrisation or a subpopulation effect in the latter parametrisation.

Convolution analysis

We assumed that the observed number of Covid-19 admissions/deaths ($D_{i,t}$) by date of report were a function of past Covid-19 notifications/admissions ($C_{i,t-\tau}$, $\tau \in \{0, ..., 30\}$), each convolved by a log-normal delay and scaled by a factor ($c_{i,t}$, which when using cases and deaths is the case-fatality ratio),

$$D_{i,t} \sim \text{NB}\left(c_i \sum_{\tau=0}^{30} \xi_i(\tau|\mu,\sigma) C_{i,t-\tau}, \phi\right)$$

where i indicates UTLA or NHS region, t day on which a specimen was taken, $\xi_i(\tau|\mu,\sigma)$ is the probability mass function of a discretised log-normal distribution and may be either estimated across all locations or estimated by location using a random effect, μ and σ are the log mean and standard deviation of the log-normal distribution, and τ indexes days prior to t. As for the fixed lag approach ϕ is the overdispersion of admissions/deaths. $c_{i,t}$ is the location and time specific case-fatality ratio (or hospitalisation-fatality ratio / case-hospitalisation ratio, respectively) and is estimated using,

$$c_{i,t} = \text{logit}^{-1}\left(\alpha f_{it} + \beta C_{i,t} + \gamma_i\right)$$

where, as for the fixed lag analysis, γ_i is either an intercept corresponding to the baseline case-fatality ratio per UTLA or NHS region, f_{it} is the fraction of cases that were found to be S-gene negative by UTLA or NHS region each week, α represents the multiplicative change in case-fatality ratio, and β corresponds to the effect of case burden with non-linearity incorporated using a thin plate spline.

All models were implemented using the `brms` package in `R` and the custom `stan` extension code.[5,6] A minimum of 2000 iterations per chain were used with 4 chains per model and convergence assessed using the R hat diagnostic.[6] All code required to reproduce this analysis is available from https://github.com/epiforecasts/covid19.sgene.utla.rt/. All intervals presented are 95% credible intervals.

Results

Visual inspection of the relationship between S-gene negativity and the case-fatality ratio, case-hospitalisation ratio, and hospitalisation-fatality ratio revealed an unclear relationship at both the NHS region and UTLA level due to the large amount of variance both between areas and over time (Figure 1). However, aggregating to the NHS region level gave some indication of a relationship between an increase in the proportion of samples that were S-gene negative and an increase in negative outcomes for Covid-19 cases.

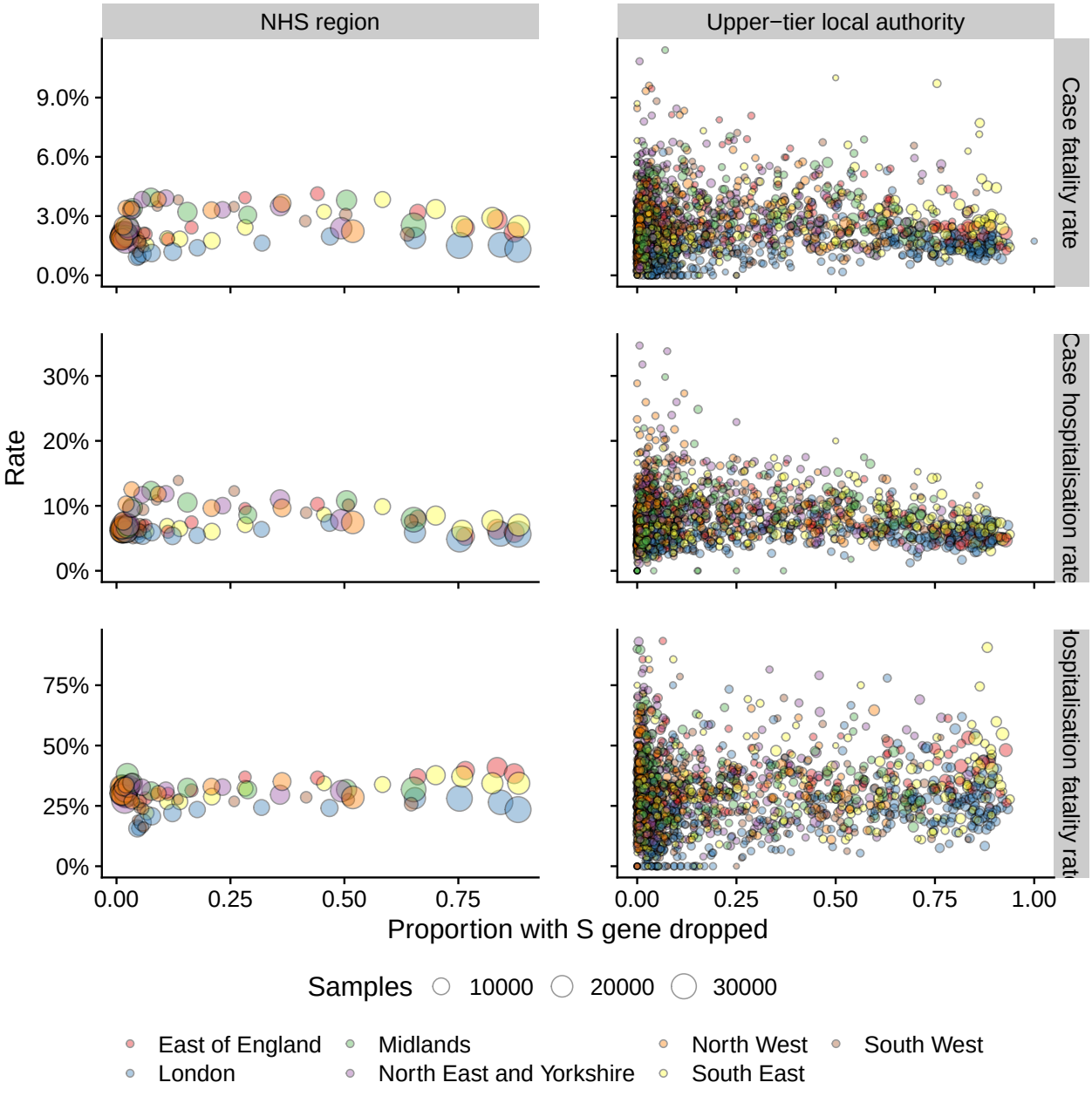


Figure 1: Proportion with S gene dropped compared to the adjusted severity ratios each week beginning Monday the 5th of October by NHS region and upper-tier local authority (UTLA). Each point represents one NHS region or UTLA and one week, with the size of the point given by the number of PCR tests.

Using our modelling framework, we found consistent evidence of an association between S-gene negativity and an increase in the case-fatality ratio (CFR) of Covid-19 though the strength of the effect varied across methods and spatial aggregation (Figure 2, Table 1, and Supplementary Table 1). In general, the strength of the effect was increased by adjusting for spatial differences (both for the delay between cases and deaths and spatial variation in the case-fatality ratio), and by current cases as a proxy for burden. However, reducing the level of aggregation from NSHE region to UTLA level reduced the estimated effect across all approaches. When the effect of SGTF was assumed to be additive we estimated that the associated additive percentage increase in the CFR due to SGTF was 0.016 (0.014, 0.017) using the optimised fixed lag approach at the UTLA level. When the effect was instead assumed to be multiplicative SGTF was associated with an increase in the CFR of 2.6 (2.3, 2.8) using the fixed lag approach at the UTLA level, 2 (1.8, 2.1) using the global convolution approach at the UTLA level, and 2.4 (2.1, 2.6) at the NHS region level. For all methods the unadjusted effect was lower but was still likely to be greater than 1.

The associated effect of S-gene negativity on the case-hospitalisation ratio and the hospitalisation-fatality ratio presented the same spatial patterns as for the case-fatality ratio with estimates for both being broadly consistent with those of the effect on the case-fatality ratio across models (Figure 2, Table 1, and Supplementary Table 1). Across all models that at least adjusted for location specific intercepts the effect on the case-hospitalisation ratio was higher than the effect on the hospitalisation-fatality ratio. When all hypothesised confounders were accounted for we found that the minimum estimated effect on the case-hospitalisation ratio associated with SGTF was 1.6 (1.5, 1.8) when data was aggregated to the NHS region level and a local convolution was assumed. Using the same method indicated little evidence of a associated effect of SGTF on the hospitalisation-fatality ratio (0.95 (0.83, 1.1)). though dropping the assumption of a locally varying delay between hospitalisation and death and instead aggregating to the UTLA level increased this to 1.2 (1.1, 1.3). Visual inspection supports the direction of these findings and the increased uncertainty in estimates for the effect on the hospitalisation-fatality ratio (Figure 1 and Figure 2).

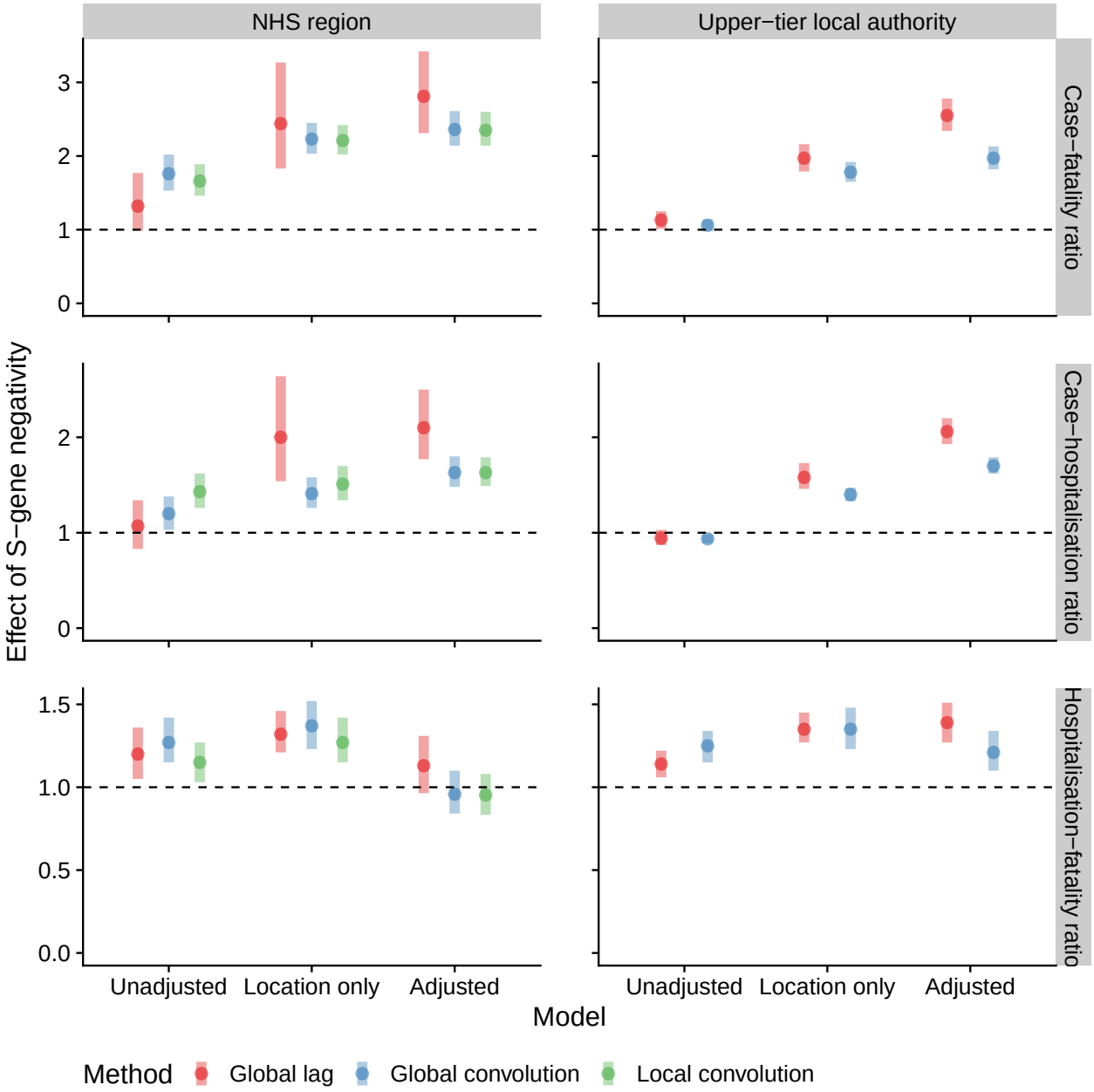


Figure 2: Estimated effect of S-gene negativity on severity ratios (median with 95\% credible intervals) assuming a multiplicative effect across spatial aggregations, delay adjustment methods, and confounder adjustment. The effect can be interpreted as a scaling of the S-gene negativity ratio. Confounders included in the adjusted model are location variability and current case/hospital admissions.

Table 1: Estimated effect of S-gene negativity on severity ratios (median with 95% credible intervals) assuming a multiplicative effect across spatial aggregations, delay adjustment methods, and confounder adjustment. The effect can be interpreted as a scaling of the S-gene negativity ratio. Confounders included in the adjusted model are location variability and current case/hospital admissions.

Method	Aggregation	Model	Case-fatality ratio	Case-hospitalisation ratio	Hospitalisation-fatality ratio
Global lag	NHS region	Unadjusted	1.3 (0.99, 1.8)	1.1 (0.83, 1.3)	1.2 (1, 1.4)
		Location only	2.4 (1.8, 3.3)	2 (1.5, 2.6)	1.3 (1.2, 1.5)
		Adjusted	2.8 (2.3, 3.4)	2.1 (1.8, 2.5)	1.1 (0.96, 1.3)
	UTLA	Unadjusted	1.1 (1, 1.2)	0.94 (0.87, 1)	1.1 (1.1, 1.2)
		Location only	2 (1.8, 2.2)	1.6 (1.5, 1.7)	1.4 (1.3, 1.5)
		Adjusted	2.6 (2.3, 2.8)	2.1 (1.9, 2.2)	1.4 (1.3, 1.5)
Global convolution	NHS region	Unadjusted	1.8 (1.5, 2)	1.2 (1, 1.4)	1.3 (1.1, 1.4)
		Location only	2.2 (2, 2.4)	1.4 (1.3, 1.6)	1.4 (1.2, 1.5)
		Adjusted	2.4 (2.1, 2.6)	1.6 (1.5, 1.8)	0.96 (0.84, 1.1)
	UTLA	Unadjusted	1.1 (0.98, 1.1)	0.94 (0.89, 0.98)	1.2 (1.2, 1.3)
		Location only	1.8 (1.7, 1.9)	1.4 (1.3, 1.5)	1.4 (1.2, 1.5)
		Adjusted	2 (1.8, 2.1)	1.7 (1.6, 1.8)	1.2 (1.1, 1.3)
Local convolution	NHS region	Unadjusted	1.7 (1.5, 1.9)	1.4 (1.3, 1.6)	1.1 (1, 1.3)
		Location only	2.2 (2, 2.4)	1.5 (1.3, 1.7)	1.3 (1.1, 1.4)
		Adjusted	2.4 (2.1, 2.6)	1.6 (1.5, 1.8)	0.95 (0.83, 1.1)

Discussion

We studied the relationship between SGTF (as a proxy for the new variant of concern) and the association between Covid-19 cases, hospitalisations and deaths adjusted using multiple approaches. We found consistent evidence across modelling approaches of an association between an increase in case-fatality ratio of Covid-19 and S-gene negativity with both additive and multiplicative assumptions though these associations were difficult to visualise in the unadjusted data. In general, the strength of the effects were increased by adjusting for spatial variation and decreased by reducing the aggregation level of the data (from NHS region to UTLA level). At the ULTA level, using the fixed lag method, we estimated that SGTF was associated with an increase in the case-fatality ratio of 0.016 (0.014, 0.017) or a multiplicative increase of 2.6 (2.3, 2.8). The convolution approach reduced the estimated effect to an increase of 2 (1.8, 2.1). We also found consistent evidence for an increase in case-hospitalisation ratios and weaker evidence of an effect on hospitalisation-fatality ratios across all methods and models evaluated. When data was aggregated at the UTLA level and a global convolution approach was used we found that SGTF was associated with a 1.7 (1.6, 1.8) increase in the case hospitalisation and a 1.2 (1.1, 1.3) increase in the hospitalisation-fatality ratio.

Our estimates for the association between SGTF and the case-fatality ratio were broadly comparable to preliminary estimates from several research groups presented to New and Emerging Respiratory Virus Threats Advisory (NERVTAG).[1] All groups used a similar study design but the estimated effect of B.1.1.7 on the case-fatality ratio varied with LSHTM estimating a 1.35 (95% CI 1.08-1.68) increase in hazard of death,[7] Imperial estimating a 1.36 (95% CI 1.18-1.56) increase in the hazard of death,[8] Public Health England (PHE) estimating a death risk ratio of 1.65 (95% CI 1.21-2.25),[9] and Exeter estimating a 1.91 (95% CI 1.35 - 2.71) increase in hazard of death. Our unadjusted estimates, aggregated at the NHS level using the fixed lag approach, were most consistent with the LSHTM and Imperial estimates whilst the Exeter estimate was more in line with our adjusted estimates and those at the UTLA level using all methods. The estimate from PHE sat between those of the other groups and were comparable to our findings from several methods and models though this was partially driven by the wide confidence intervals of the PHE estimate. Whilst these groups used a range of approaches all analyses were based on the same core dataset of SGTF cases identified through Pillar 2 testing linked to the PHE COVID-19 deaths line list.[1,9] Though providing individual level data this line list was based on a subset of deaths (<10%) and of those who died only 26% had a Pillar 2 test and of those only 30% had S-gene data.[1] This means that all of these studies may be based on a non-representative population which may lead to bias. However, our estimates are also likely biased both due to unmeasured confounding and because they are based on data that does not directly link Covid-19 notifications with hospital admissions or deaths. Individual level studies on the association between B.1.1.7 or SGTF and case-hospitalisation risk or hospitalisation-fatality risk are currently limited. However, a rapid analysis using data from the CO-CIN study did not identify an increase in the risk of death in hospitalised B.1.1.7 cases.[1,10] Whilst our study did find evidence of an association between SGTF and the hospitalisation-fatality ratio the effect size was small and sensitive to the choice of method and covariates indicating that it may not be a robust finding. However, CO-CIN data is currently unevenly distributed and not fully recorded for the period of interest meaning that the evidence for an effect may change as more data is collected.

Our results are indicative only as they make use of aggregated data that is potentially subject to bias. However, they may act as useful support for other sources of evidence, and potentially may be generalised to scenarios where individual level data is not available. Whilst we adjusted for location and case/admissions as a proxy for health service strain we could not adjust for multiple confounders such as the age of cases/admissions, and more direct indicators of health service burden. We aimed to create these analysis with public data only - if more relevant data aggregated to UTLA and NHS region were to become publicly available this could be incorporated into our framework and potentially may reduce bias in our findings. We also only made use of weekly point estimates for the proportion of cases that were SGTF in each UTLA. Because of this, uncertainty in our regression coefficients are underestimated, and probably considerably so. In addition to this, SGTF data was only available by specimen date which meant that we had to use fixed mappings to shift estimates to the relevant case/admission date which may have introduced bias. Unfortunately, fitting a delay distribution varying at UTLA-level was not possible once covariates were included due to computational costs. Hopefully, further work may allow us to improve on the efficiency of

the model and associated code to allow for more complex models to be calibrated in a reasonable time frame. Finally, SGTF does not uniquely identify the B.1.1.7 variant and therefore our study may be biased by the inclusion of other variants, although recently B.1.1.7 has made up almost all SGTF samples[11]. Our results should be used to triangulate the effect of B.1.1.7 rather than as standalone evidence and will hopefully be made redundant by improved data sources and linkages.[1]

Population-level surveillance data supports findings from other studies using individual-level data that SGTF is associated with an increase in the case-fatality ratio and indicates that the majority of this effect is likely linked to an increase in the case-hospitalisation ratio rather than the hospitalisation-fatality ratio. This evidence is indicative only due to the potential for unadjusted confounding and should be supported by further studies using individual-level data. Our methods are available as an open-source extension to a Bayesian regression package and may be used to study case-fatality ratios, case-hospitalisation ratios and hospitalisation-fatality ratios in other settings where individual data is not available or biased.

References

1. Peter Horby, N. D., Catherine Huntley. (2021). *NERVTAG note on b.1.1.7 severity.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/955239/NERVTAG_paper_on_variant_of_concern___VOC___B.1.1.7.pdf.

2. *Coronavirus (covid-19) in the uk.* (2021). https://coronavirus.data.gov.uk/details/healthcare.

3. Meakin, S., Abbott, S., & Funk, S. (2021). *NHS trust level covid-19 data aggregated to a range of spatial scales.* https://doi.org/10.5281/zenodo.4447465

4. England, P. H. (2020). *Investigation of novel sars-cov-2 variant: Variant of concern 202012/01.* https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201.

5. Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

6. Stan Development Team. (2020). *RStan: The r interface to stan.* http://mc-stan.org/

7. Davies, D.-O., N. (2021). *Relative fatality hazard in pillar 2 tested individuals with voc.* unpublished analysis.

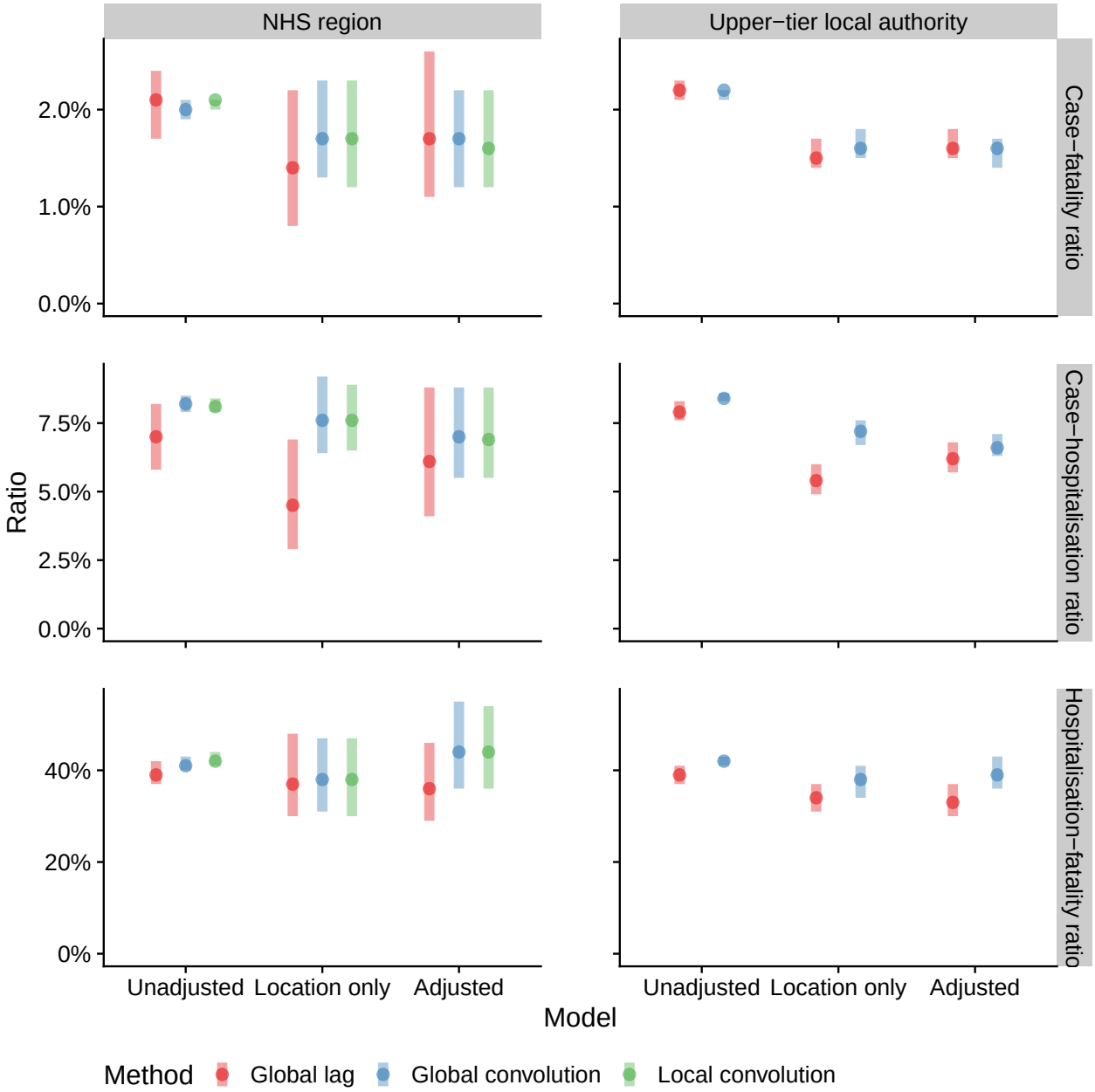
8. Ferguson, N. (2021). *Non-parametric analysis of fatal outcomes associated with b1.1.7.* unpublished analysis.

9. England, P. H. (2021). *Variant of concern 202012/01: Technical briefing 2. Investigation of novel sars-cov-2 variant.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949639/Technical_Briefing_VOC202012-2_Briefing_2_FINAL.pdf.

10. Docherty A., E., Harrison. (2021). *Hospital case fatality and emergence of variant of concern b.1.1.7.* unpublished analysis.

11. England, P. H. (2021). *Variant of concern 202012/01: Technical briefing 3.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/954978/s1014-phe-voc-202012-01-technical-briefing-3.pdf.

Supplementary information



Supplementary Figure 1: Baseline estimated national severity ratios expressed as percentages adjusted for S-gene negativity. The median with 95% credible is show for each estimate. Confounders included in the adjusted model are location variability and current case/hospital admissions.

Aggregation	Model	Case-fatality ratio	Case-hospitalisation ratio	Hospitalisation-fatality ratio
NHS region	Unadjusted	0.0065 (-0.00021, 0.014)	0.0031 (-0.0092, 0.016)	0.01 (-0.0067, 0.028)
	Location only	0.02 (0.015, 0.024)	0.03 (0.016, 0.044)	0.028 (0.01, 0.046)
	Adjusted	0.021 (0.018, 0.025)	0.041 (0.029, 0.051)	0.008 (-0.011, 0.026)
UTLA	Unadjusted	0.0028 (0.00044, 0.0053)	-0.0037 (-0.0095, 0.002)	0.017 (0.0026, 0.032)
	Location only	0.013 (0.011, 0.014)	0.028 (0.023, 0.032)	0.054 (0.041, 0.068)
	Adjusted	0.016 (0.014, 0.017)	0.043 (0.04, 0.047)	0.049 (0.034, 0.064)

Supplementary Table 1: Estimated effect of S-gene negativity on severity ratios (median with 95% credible intervals) assuming an additive effect using the fixed delay method across spatial aggregations, and confounder adjustment. The effect can be interpreted as a direct change in the ratio related to S-gene negativity. Confounders included in the adjusted model are location variability and current case/hospital admissions.