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Work in progress - not peer reviewed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: sebastian.funk@lshtm.ac.uk

Abstract

Background: Individual level data indicates S-gene target failure (SGTF) may be associated with increased case fatality rates, increased hospitalisation rates, and increased hospitalisation fatality rates. Whilst an individual level data approach represents the gold standard for observational analysis population level analysis may be helpful to triangulate findings, especially when individual data sources are confidential, or only partially representative. In this analysis, we use multiple approaches to evaluate public population level data for evidence of an association between SGTF and severity measures.

Method: We explored the association between the proportion of samples that were S-gene negative and the case fatality rate, hospitalisation rate, and hospitalisation fatality rate of Covid-19 aggregated at the UTLA and NHS region level. Two approaches were used with the first assuming a fixed lag between primary and secondary observations with the lag optimised using the Pearson's correlation coefficient. The second approach assumed that the secondary observations could be estimated using a convolution of primary observations multiplied by some scaling factor. For the fixed lag analysis we investigated both additive and negative effects of being S-gene negative and for the convolution approach we explored delays between observations that varied spatially. We present both univariate and multivariate estimates with the latter adjusted for spatial and temporal variation.

Results: We found consistent evidence of an association between an increase in case fatality rate of Covid-19 and S-gene negativity. At the ULTA level, using the fixed lag approach, we estimated that SGTF was associated with an increase in the case fatality rate 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 case hospitalisation rates (CHRs) and hospitalisation fatality rates (HFRs) with the effect on CHRs (1.7 (1.6, 1.8)) being larger than that on HFRs (1.2 (1.1, 1.3)).

Conclusions: Population level surveillance data supports findings from other studies using individual level data that SGTF is associated with an increase in the case fatality rate and indicates that the majority of this effect is likely linked to an increase in the case hospitalisation rate rather than the hospitalisation fatality rate. 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 may be associated with an increased case fatality rate. 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.

Whilst individual level data represents the gold standard for observational analysis population level analyses may be helpful to triangulate findings and provide initial evidence to guide further investigation.[1]

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 rates, increased hospitalisation rates, and increased hospitalisation fatality rates using population level surveillance data aggregated at both the NHS 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 rate) and a distributional delay from case/admission to death/admission. We explore both additive and multiplicate effects of the 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.

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 shifting all estimates 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 using a random effect ("local convolution"). All analyses were repeated at NHS region and upper-tier local authority (UTLA) scales. Futher details of each analysis are given in the following sections.

Fixed lag analysis

We assumed that the observed number of Covid-19 admissions/deaths ($D_{i,t}$) within 28 days by date of infection were a function of Covid-19 notifications/admissions ($C_{i,t}$) by date of infection scaled by the case fatality rate of S-gene positive cases (c^+) and S-gene negative cases (c^-),

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

where i indicates UTLA or NHS region, t week of infection, ϵ is an error term that accounts for imported deaths/admissions not linked to local cases/admissions, and f_{it} is the fraction of cases that were found to be S-gene negative by UTLA each week. The case fatality rate (or hospitalisation-fatality rate / case-hospitalisation rate, respectively) of S-gene negative cases then assumed was then assumed to be a function of static local, and temporal variation.

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

where γ_i is an intercept corresponding to the baseline case fatality rate per UTLA or NHS region, and β corresponds to the effect of case burden with non-linearity incorporated using a thin plate spline. In

other words, we stratify the data set either by UTLA or by week and determine whether differences in the associations between cases, admissions and deaths are explained by changes in proportion of cases that are SGTF over time and space, respectively.

The S-gene negative case fatality rate was then assumed to be related to the S-gene positive case fatality rate via a multiplicative relationship,

$$c^- = \alpha c^+$$

or an additive relationship

$$c^- = \alpha + c^+$$

where α represents either the multiplicative change in case fatality rate or the additive change. These alternative parameterisations represent either a population wide effect for the former parameterisation or a subpopulation effect in the latter parameterisation.

Convolution analysis

We assumed that the observed number of Covid-19 admissions/deaths ($D_{i,t}$) by date of report were a function of Covid-19 notifications/admissions ($C_{i,t}$) by date of report, convolved by a log normal delay, and scaled by a rate (which when using cases and deaths is the case fatality rate),

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

where i indicates UTLA or NHS region, t day of report, $\xi_{i,\tau}$ is the probability mass function of a log normal distribution and may be either be static across all locations or vary by location, and τ indexes days prior to t. c_i is the location specific case fatality rate (or hospitalisation-fatality rate / case-hospitalisation rate, respectively). c_i is then estimated using,

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

where, as for the fixed lag analysis, γ_i is either an intercept corresponding to the baseline case fatality rate per UTLA or NHS region, 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`[5] package in `R` and the custom `stan` extension code. 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 rate, case hospitalisation rate, and hospitalisation fatality rate was difficult 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 negative outcomes.

Using our modelling framework, we found consistent evidence of an association between S-gene negativity and an increase in the case fatality rate (CFR) of Covid-19 though the strength of the effect varied across method and spatial aggregation (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 rates), and

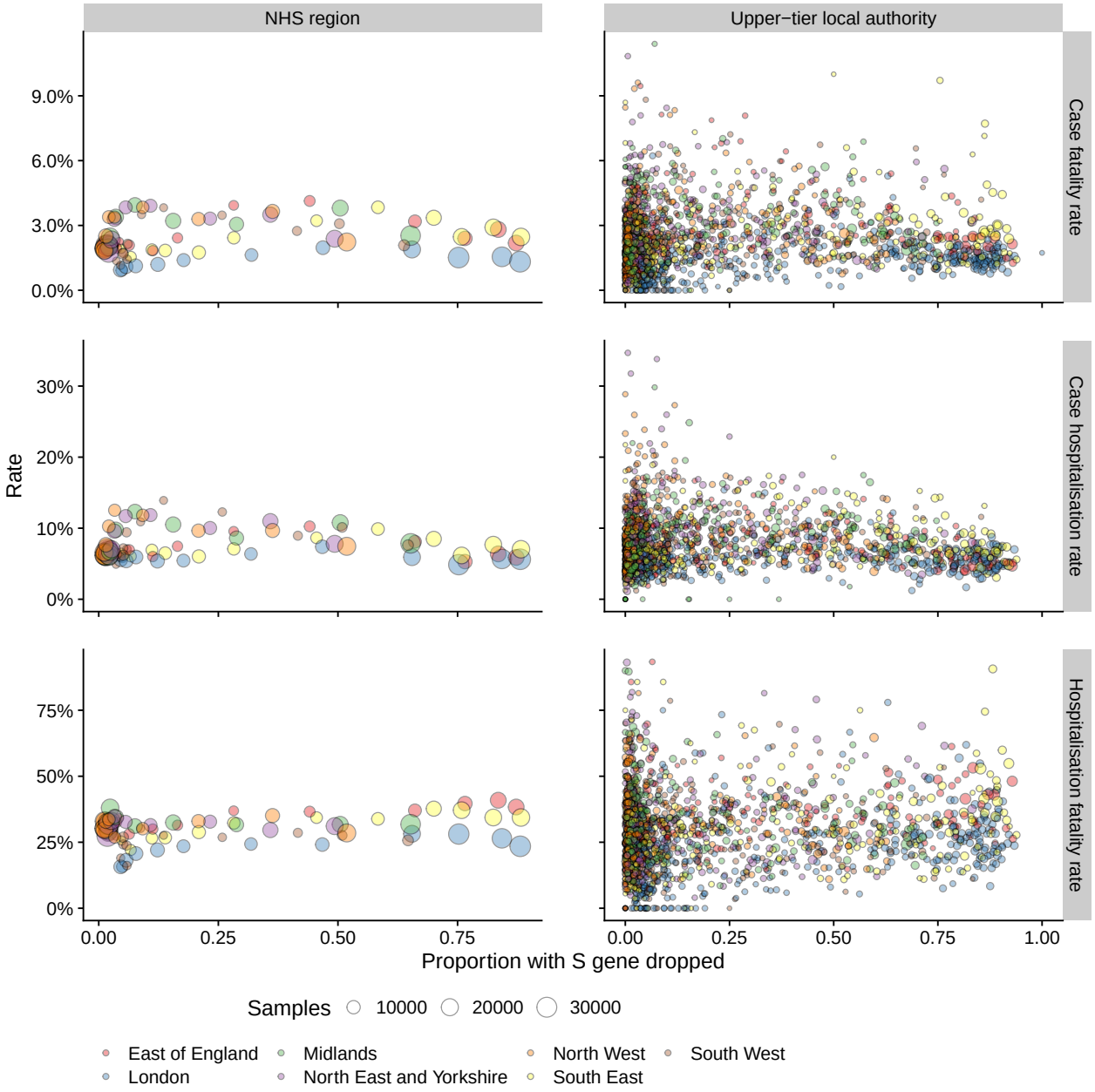

Figure 1: Proportion with S gene dropped compared to the adjusted severity rates each week beginning Monday the 5th of October by NHS region and upper-tier local authority (UTLA). Each point represents one NHS region or UTLA and one week, with the size of the point given by the number of PCR tests.

by current cases/admissions. However, reducing the level of aggregation from NSH region to UTLA level reduced the observed effect across all approaches. When the effect of SGTF was assumed to be additive we estimated that the associated 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 substantially greater than 1 using either of the convolution approaches or using data aggregated to the UTLA level rather than to NHS region.

The effect of S-gene negativity on the case hospitalisation rate and the hospitalisation rate presented the same spatial patterns as for the case fatality rate with estimates for both being broadly consistent with those of the effect on the case fatality rate across models (Table 1). Across all models that at least adjusted for location specific intercepts the effect on the case hospitalisation rate was higher than the effect on the hospitalisation fatality rate. When all hypothesised confounders were accounted for we found that the minimum estimated effect on the case hospitalisation rate 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 an effect of SGTF on the hospitalisation fatality rate (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 uncreased uncertainty in estimates for the effect on the hospitalisation fataility rate (Figure 1).

Table 1: Estimated effect of S-gene negativity on severity rates (median with with 95% credible intervals) for the both additive and multiplicative assumptions across spatial aggregations, delay adjustment methods, and confounder adjustment. In additive models the effect can be interpreted as a direct change in the rate related to S-gene negativity whilst in multiplicative model the effect can be interpreted as a scaling of the S-gene positive rate. Confounders included in the adjusted model are location variability and current case/hospital admissions.

Effect type	Method	Aggregation	Model	Case fatality rate	Case hospitalisation rate	Hospitalisation fatality rate
Additive	Global lag	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)
Multiplicative	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 rate 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 rate 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 rates and weaker evidence of an effect on hospitalisation fatality rates 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 aa 1.7 (1.6, 1.8) increase in the case hospitalisation and a 1.2 (1.1, 1.3) increase in the hospitalisation fatality rate.

Our estimates for the association between SGTF and the case fatality rate were broadly comparable to preliminary estimates from several research groups presented to the 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 rate varied with LSHTM estimating a 1.35 (1.08-1.68) increase in hazard of death, Imperial estimating a 1.36 (1.18-1.56) increase in the hazard of death,[6] Public Health England (PHE) estimating a death risk ratio of 1.65 (1.21-2.25),[7] and Exeter estimating a 1.91 (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.[1,7] Though providing individual level data this line list was based on a subset of deaths (<10%) and of those who died only 26% had 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 Covid-19 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 C0-CIN study did not identify an increase in the risk of death in hospitalised B.1.1.7 cases.[1,8] Whilst our study did find evidence of an association between SGTF and the hospitalisation fatality rate the effect size was small and sensitive to the choice of method and covariates indicating that it may not be a robust finding. However, C0-CIN data is currently unevenly distributed and not fully recorded for the study period meaning that the evidence for an effect may strengthen over time in-line with our findings.

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, individual level approaches, 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/admission. If more relevant data aggregated to UTLA and NHS region becomes publicly available this can be incorporated which may reduce any 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 relevent case/admission date which may have introduced bias. Finally, SGTF does not uniquely identify the B.1.1.7 and therefore our analysis may be biased by the inclusion of other variants. Our results should be used to triangulate the effect of B.1.1.7 rather than as standalone evidence and will hopefully be made redudent by improved data sources and linkages.

Population level surveillance data supports findings from other studies using individual level data that SGTF

is associated with an increase in the case fatality rate and indicates that the majority of this effect is likely linked to an increase in the case hospitalisation rate rather than the hospitalisation fatality rate. 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 rates, case hospitalisation rates and hospitalisation fatality rates in other settings where individual data is not available.

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

7. 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.

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