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Work in progress - not peer reviewed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

12 January, 2021

For correspondence: sebastian.funk@lshtm.ac.uk

Abstract

Background: Individual level data indicates S-gene target failure 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 S-gene target failure 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: TODO: summarise results.

Conclusions: TODO: Results and what it means. Our convolution regression approach may be useful for others where individual data is not available or subject to biases. It has been implemented in a generalised framework with all code made publicly available.

Method

Data

We used 4 main sources of data: test positive Covid-19 notifications by UTLA,[1] hospitalisations with Covid-19 by UTLA,[2] deaths linked to Covid-19 notification within 28 days of notification,[1] and S-gene status from PCR tests by local authority provided by Public Health England (PHE)[3]. We aggregated the data at the weekly, or daily, level and restricted the analysis to the period beginning Monday, 5 October.

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.

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

where $\gamma_{i,t}$ is either a UTLA-level intercept $\gamma_{i,t} \equiv \delta_i$ corresponding to the baseline case fatality rate per UTLA, or a temporal intercept $\gamma_{i,t} \equiv \theta_t$ corresponding to the baseline case fatality rate over time, with variation over time 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,

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

where, as for the fixed lag analysis, $\gamma_{i,t}$ is either a UTLA-level intercept $\gamma_{i,t} \equiv \delta_i$ corresponding to the baseline case fatality rate per UTLA, or a temporal intercept $\gamma_{i,t} \equiv \theta_t$ corresponding to the baseline case fatality rate over time, with time modelled using a thin plate spline.

All models were implemented using the `brms`[4] package in `R`. All code required to reproduce this analysis is available from https://github.com/epiforecasts/covid19.sgene.utla.rt/.

Results

TODO: Discuss univariate findings and evidence in scatter plot. Start at UTLA level and summarise difference at NHS region (if any). Summarise differences across models.

TODO: Discuss multivariate findings starting with UTLA and summarise difference at NHS region level (if any). Discuss impact of adjusting for covariates (and potentially subsets of covariates with these needing to be added to the SI). Discuss impact of convolution vs lag.

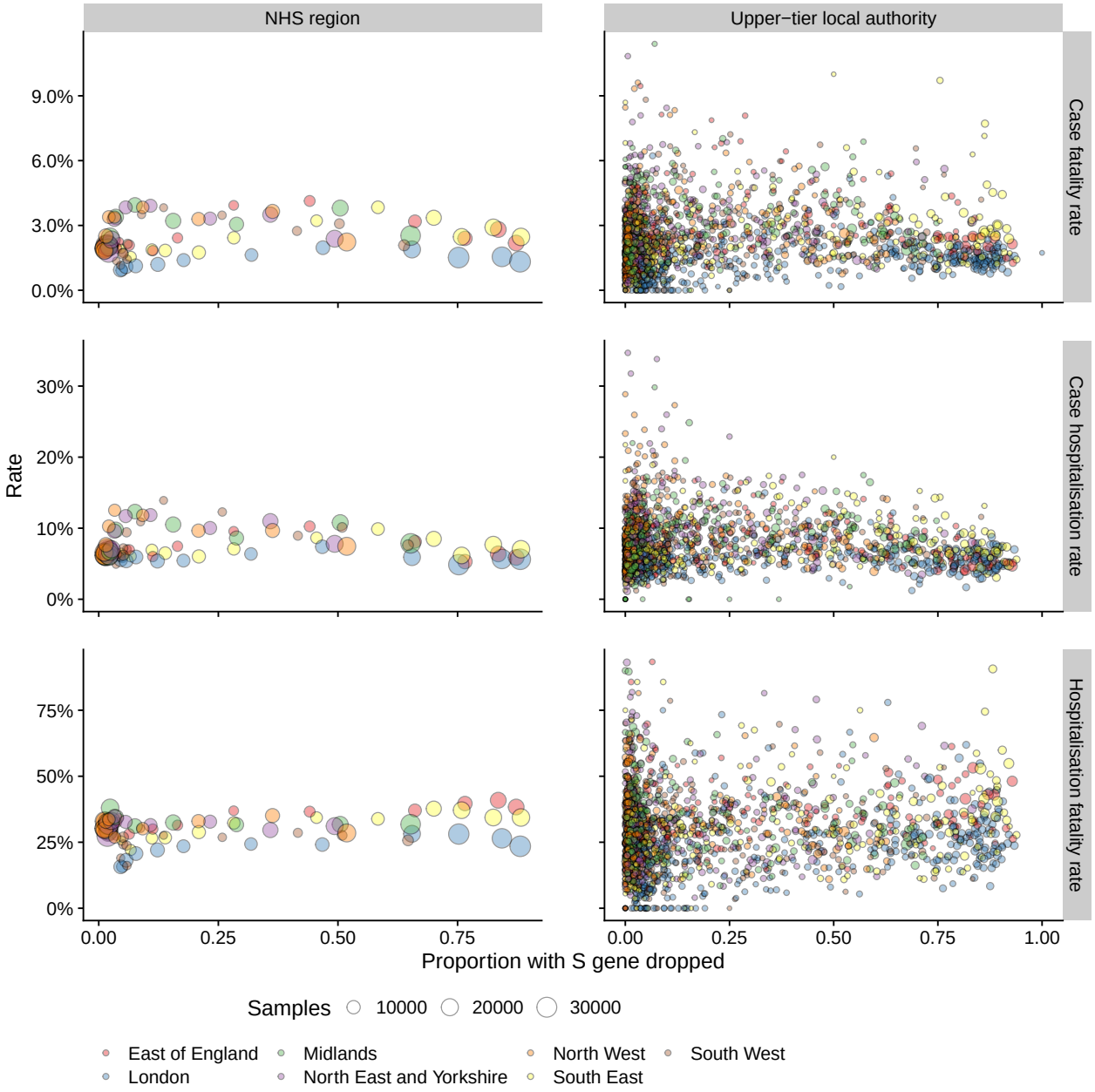

Figure 1: Proportion with S gene dropped compared to the adjusted severity rates each week beginning Monday the 5th of October by NHS region and upper-tier local authority (UTLA). Each point represents one NHS region or UTLA and one week, with the size of the point given by the number of PCR tests.

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.

TODO: Summarise convolution findings (univariate and multivariate, by UTLA and NHS region). TODO: Summarise lag findings (univariate and multivariate, by UTLA and NHS region). TODO: Summarise differences across approaches (spatial scale, and model).

Our estimates for the association between SGTF and the case fatality rate were comparable to those from individual based approaches but...

TODO: Compare to lshtm + exeters + imperials case control and survival analysis of case fatality rate. TODO: Compare to other studies looking at hospitalisation rate + hospitalisation fataility rate.

Our results are indicative only as they make use of aggregated data that is subject to a large range of confounders. 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 temporal variation we could not adjust for multiple confounders such as the age of cases/admission due to the lack of public data on the age of Covid-19 cases/admissions by NHS region or UTLA. Our inclusion of temporal variation as a confounder may also compete with the effect of S-gene target failure causing bias in our multivariate estimates. Lastly, we fitted the model only point estimates of the proportion of SGTF observed in every UTLA per week. Because of this, uncertainty in our regression coefficients are underestimated, and probably considerably so. Finally, SGTF uniquely identify the novel variant and therefore our analysis may be biased by the inclusion of other variants. Our results should be used to triangulate the effect of novel variant rather than as standalone evidence.

TODO: Improve limitations. TODO: Conclusions. What does it mean. Why is this a good way of doing it. What can we do next. Method generalisability. Wrap up.

References

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

2. 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

3. 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.

4. 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