Local area reproduction numbers and S-gene target failure

Preliminary analysis - not peer reviewed

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

6 January, 2020

Abstract

Background: Local estimates of the time-varying effective reproduction number (Rt) of Covid-in England have become increasingly heterogeneous since the introduction of the November national lockdown in England. This variation may be attributed to the spread of a novel Covid-19 variant. In this report we aimed to investigate the changes in the proportion of positive cases that were S-gene negative ("S-gene target failure", SGTF), an indicator of the novel variant, and to correlate them with changes in the estimated reproduction number at the level of upper tier local authorities (UTLA).

Method: We explored the relationship between the proportion of samples that were S-gene negative and the effective reproduction number of test positive cases over time. Effective reproduction numbers were estimated using the `EpiNow2` R package independently for each local authority using two different estimates of the generation time. The `brms` R package was used to fit a range of regression models to estimate a multiplicative relationship between SGTF and weekly mean R. Models were compared using the expected log-predictive density.

Results: We found evidence of an association between increased mean Rt and the proportion of SGTF across all models evaluated with the magnitude of the effect increasing as model flexibility was decreased. Models found an effect consistent with the low ends of estimates that reported elsewhere when a generation time with a mean of 5.5 days was used. Models that adjusted for either national level or NHS region level residual variation over time were found to fit the data better and found yielded consistent results of the association of Rt with SGTF.

Conclusions: Our results indicate that even after adjusting for between NHS region residual variation over time S-gene negativity was associated with an increase in the effective reproduction number of Covid-19. These findings were robust across a range of models and generation time assumptions though the specific effect size was variable depending on the assumptions used. The lower bound of the estimated effect indicated that NPI measures implemented between September and January 1st in England may not be sufficient to reduce the reproduction number below 1.

Method

Data

We used 4 main sources of data: test positive Covid-19 notifications by UTLA[1], S-gene status from PCR tests by local authority, Google mobility data stratified by context[2], and two publicly available data bases of of non-pharmaceutical interventions by UTLA[3].

Statistical analysis

We calculated the weekly proportion of positive tests that were S-gene negative over time by local authority. We estimated reproduction numbers using the method described in[5] and[6] and implemented in the `EpiNow2` R package[7]. Daily updated estimates can be downloaded at https://github.com/epiforecasts/covid-rt-estimates/blob/master/subnational/united-kingdom-local/cases/summary/rt.csv. We used two sets of estimates, obtained using uncertain, gamma distributed, generation interval distributions with a mean of 3.6 days (standard deviation (SD): 0.7), and SD of 3.1 days (SD: 0.8)[7,8] or with a mean of 5.5 days (SD: 0.5 days), and SD of 2.1 days (SD: 0.25 days)[9], respectively.

We then built a separate model of the expected reproduction number in UTLA i during week t starting in the week beginning 14 September, 2020, as a function of local restrictions, mobility indicators, residual temporal variation, and proportion of positive tests S-gene negative:

$$R_{i,t} = (1 + \alpha f_{it}) \exp\left(s(t) + \sum_j \beta_j T_{ijt} + \sum_k \gamma_k G_{ikt} + \log R_i\right)$$

where R_t is an UTLA-level intercept corresponding to Rt during national lockdown in November, T_{ijt} is 1 if intervention j (out of: no tiers, tier 1/2/3) is in place and 0 otherwise, G_{ikt} is the relative mobility in context k (home, parks, workplace, etc.) at time t in UTLA i as measured by Google, and $s(t)$ is a time-varying component, modelled either as a region-specific thin-plate regression spline ("Regional time-varying"), the sum of a static regional parameter and a national spline ("National time-varying"), or only a static regional parameter ("Regional static"). The key parameter is α, the relative change in reproduction number in the presence of the variant that is not explained by any of the other variables, where f_{it} is the proportion out of all positive tests for SARS-CoV-2 where the S-gene was tested that came back negative for the S-gene, and the reproduction number in any given UTLA is

$$R_{t,i} = (1 - f_{it})R^+_{t,i} + f_{it}R^-_{t,i}$$

where $R^-_{t,i}$ is the S-gene negative reproduction number, $R^+_{t,i}$ is the S-gene positive reproduction number, and it is assumed that $R^-_{t,i} = (1 + \alpha)R^+_{t,i}$

We used a Gaussian observation model with a single variance parameter. All models were implemented using the `brms`[10] package in `R`. All code required to reproduce this analysis is available from https://github.com/epiforecasts/covid19.sgene.utla.rt/.

Results

We found consistent evidence of an association between S-gene negativity and increased UTLA level reproduction number estimates. The association became more apparent over time from the middle of October through to the beginning of December (Figure 1) as the proportion of of tests that were S-gene negative increased heterogeneously across NHS regions. The association appeared to be both across NHS regions and within NHS regions. Models that adjusted for residual variation over time on both a national and NHS region level fit the data better than those that did not (Table 1) but all models had evidence of increased Rt with S-gene negativity with the best fitting model yielding a lower bound of 34% with a short generation time (Table 2), slightly higher but consistent with estimates from the models that only adjusted for national level residual variation over time (lower bound: 27%), or a model that did not adjust for residual variation over time (lower bound: 44%. With a longer generation time, these lower bounds changed to 48%, 40%, and44%, respectively. The upper bound of the increase in Rt varied from 36% to 58% in models that fit the data comparably (but had different assumed generation times).

The best fitting model appeared to reproduce estimated reproduction numbers well (Figure 2), although there are notable outliers in some of the weeks. Alternative model parametrisations fit the data less well than those presented here whilst producing comparable effect size estimates.

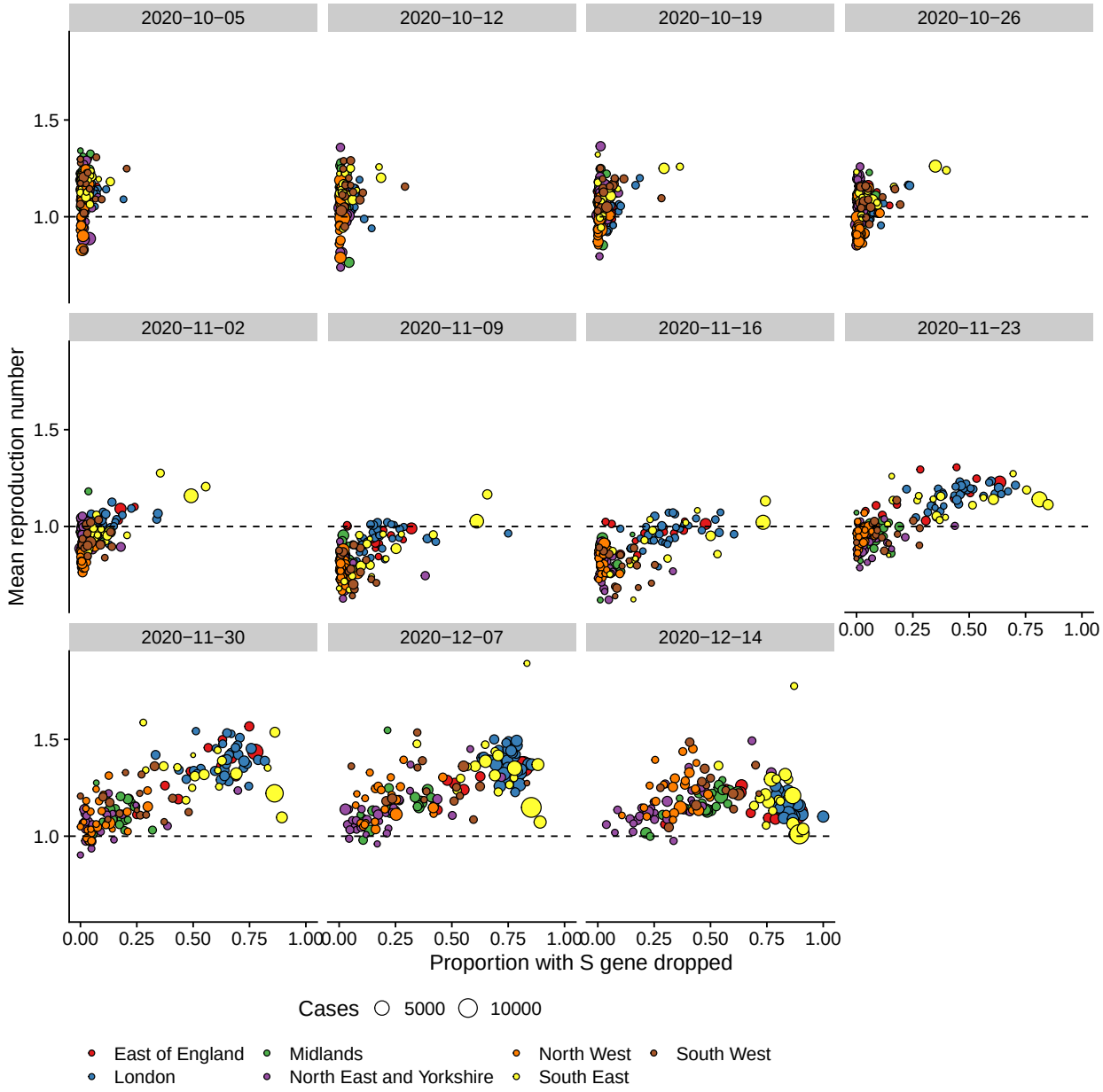


Figure 1: Mean reproduction numbers using a generation time with a mean of 3.6 days since the week beginning 14 September, 2020, compared to the proportion of all test-positives tested for S-gene that tested S-gene positive/negative that week. Each point represents one UTLA, with the size given by the number of cases in the week following the week of the given reproduction number to account for the delay from infection to testing.

Model	ELPD difference
Regional time-varying	0.00
National time-varying	-146.36
Regional static	-390.55

Table 1: Model comparison (long generation interval) by difference in expected log-predictive density.

Model	Estimate (short GT)	Estimate (long GT)
Regional static	0.33 (0.3-0.36)	0.49 (0.44-0.53)
National time-varying	0.3 (0.27-0.34)	0.46 (0.4-0.52)
Regional time-varying	0.38 (0.34-0.41)	0.53 (0.48-0.58)

Table 2: Parameter α with 95% credible intervals for the three different models of $s(t)$ for short (3.6 days mean) and long (5.5 days mean) generation intervals. The estimate corresponds to the multiplicative increase in reproduction number estimated for S-gene negative cases.

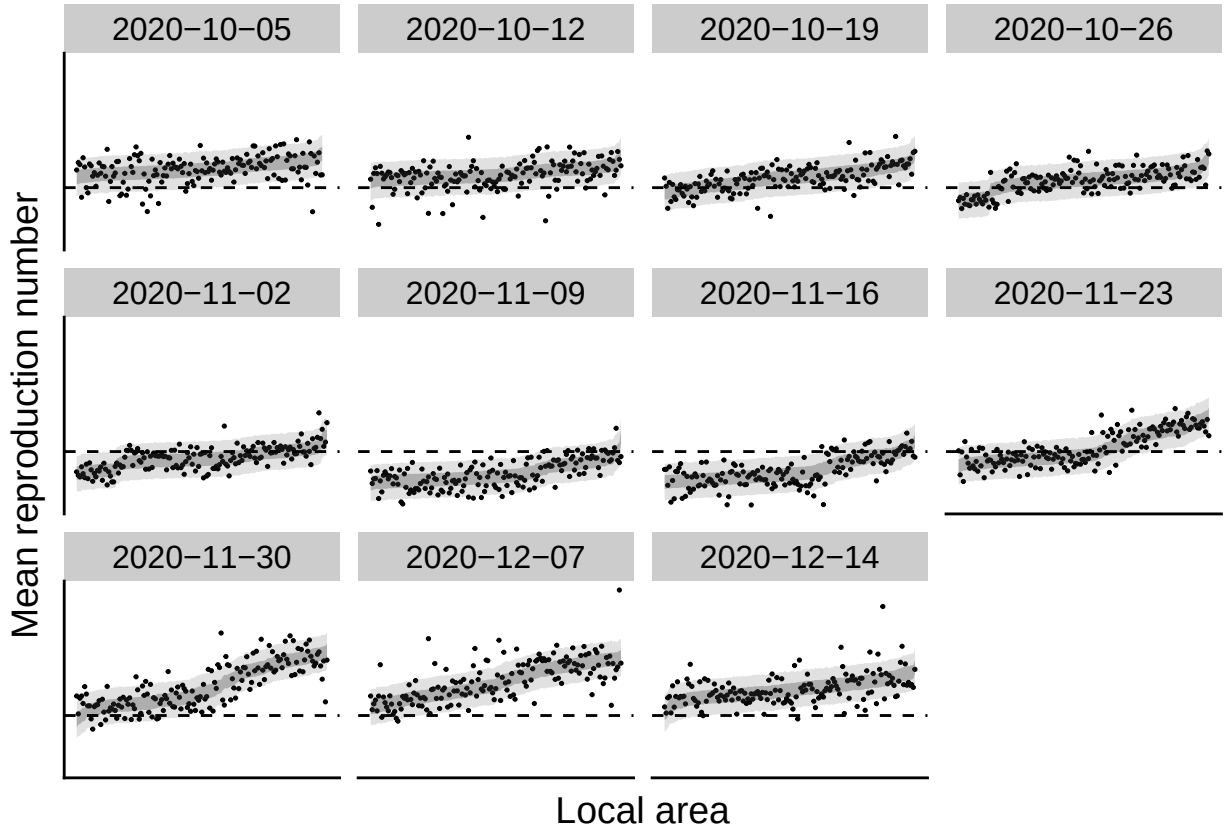


Figure 2: Predictions of the best fitting short generation interval model (per-region spline) compared to the data (solid dots). Dark grey: central 50% prediction interval; light grey: central 90% prediction interval. Areas are ordered each week according to predicted median.

Discussion

We studied the relationship between SGTF (as a proxy for the new variant of concern) and the effective reproduction number using three related models that had varying degrees of flexibility in ascribing changes in the effective reproduction number to factors not explained by the proportion of cases with SGTF. The model with region-specific splines fitted the best yielding central estimates of a 38% and 53% increase in R with short and long generation interval, respectively. In principle this could be an underestimate as it can explain regional differences by unmodelled factors, i.e. ones beyond interventions, mobility, and distribution of SGTF, and is therefore largely a model of within-regions UTLA-level differences, yet we found the results consistent with the model that modelled all temporal variation as related to interventions, mobility and SGTF (central estimates: 33% and 49% increase with short and long generation interval, respectively). This model yielded a

poorer fit to the data than the more flexible models that allowed for variation over time not ascribed to these factors.

Our estimates with the longer generation interval without adjusting for additional residual variation over time are comparable with ones from other modelling studies (both of which used a generation interval centred around 6.5 days, i.e. longer than our "long" interval) which were in the order of 50-74%[11] or 50-75%[12], if at the lower end of these estimates. Shorter generation intervals lead to reproduction numbers closer to 1 and thus possibly lower estimates of a multiplicative effect. This may be a particular issue where the effect of the variant would cause the reproduction number to cross 1.

Our results should be treated with caution as several caveats apply: we have not observed any local authorities in which all tests were S-gene negative and therefore are extrapolating beyond the available data. We assumed that the effect of tiers and lockdown applied uniformly across the country. While we did allow for a flexible regional-level behaviour through our use of regression splines, there may be UTLA level variation that we did not capture in doing so. If this could explain some of the sub-regional differences in reproduction numbers, our estimate for the increased reproduction number could biased. Lastly, we fitted the model only to the mean estimated reproduction numbers and therefore ignored uncertainty in these estimates as well as in the proportion of S-gene dropout observed in every UTLA per week. Because of this, uncertainty in our regression coefficients are underestimated, and probably considerably so. Further investigation will be necessary in order to establish the relationship between S-gene dropouts and the reproduction number.

We found consistent evidence that S-gene negativity was associated with increased reproduction numbers across a range of models and assumptions. The precise estimate of the effect size was impacted by the both the degree of flexibility allowed in the model used and the assumed generation time. However, the lower bound of the effect implies that NPI measures implemented since September in England may not be sufficient to reduce the reproduction number below 1. Our analysis is fully reproducible and all the aggregated data used is publicly available for reuse and reinterpretation.

References

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

2. *Covid-19 community mobility reports*. (2021). https://www.google.com/covid19/mobility/.

3. Gibbs, H., Jarvis, C. I., & Eggo, R. M. (2021). *UK tier data* (Version 0.0.2). Zenodo. https://doi.org/10.5281/zenodo.4421649

4. *Covid 19 local restriction tiers by area*. (2021). https://github.com/ONSgeo/Covid19_Tiers.

5. Abbott, S., Hellewell, J., Thompson, R., Sherratt, K., Gibbs, H., Bosse, N., Munday, J., Meakin, S., Doughty, E., Chun, J., Chan, Y., Finger, F., Campbell, P., Endo, A., Pearson, C., Gimma, A., Russell, T., null, null, Flasche, S., ... Funk, S. (2020). Estimating the time-varying reproduction number of sars-cov-2 using national and subnational case counts. *Wellcome Open Research*, *5*(112). https://doi.org/10.12688/wellcomeopenres.16006.2

6. Sherratt, K., Abbott, S., Meakin, S. R., Hellewell, J., Munday, J. D., Bosse, N.Jit, M., & Funk, S. (2020). Evaluating the use of the reproduction number as an epidemiological tool, using spatio-temporal trends of the covid-19 outbreak in england. *medRxiv*. https://doi.org/10.1101/2020.10.18.20214585

7. Abbott, S., Hellewell, J., Sherratt, K., Gostic, K., Hickson, J., Badr, H. S., DeWitt, M., Thompson, R., EpiForecasts, & Funk, S. (2020). *EpiNow2: Estimate real-time case counts and time-varying epidemiological parameters*. https://doi.org/10.5281/zenodo.3957489

8. Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., & Hens, N. (2020). Estimating the generation interval for coronavirus disease (covid-19) based on symptom onset data, march 2020. *Eurosurveillance*, *25*(17). https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257

9. Ferretti, L., Ledda, A., Wymant, C., Zhao, L., Ledda, V., Abeler-Dorner, L., Kendall, M., Nurtay, A.,

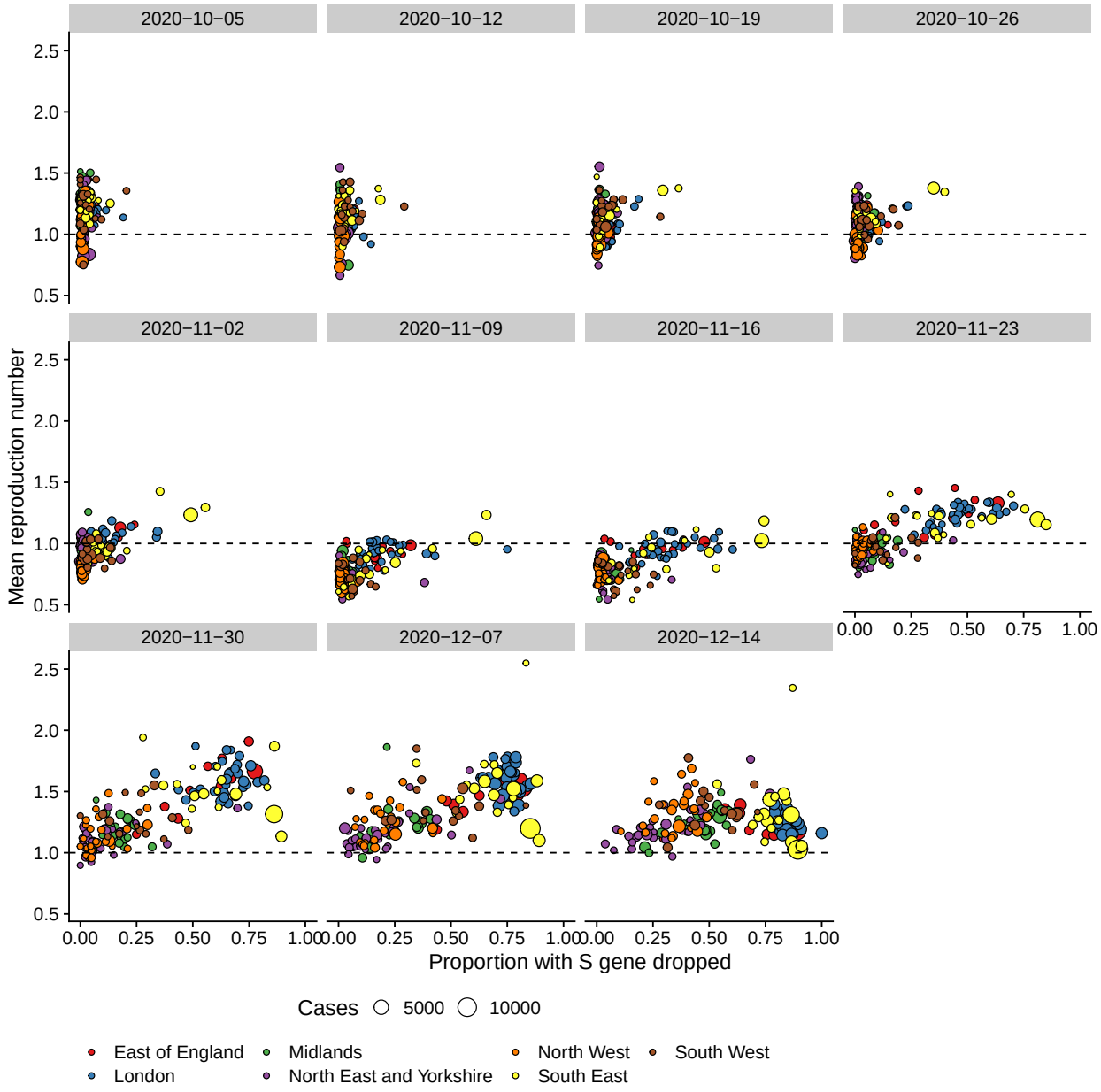
Cheng, H.-Y., Ng, T.-C., Lin, H.-H., Hinch, R., Masel, J., Kilpatrick, A. M., & Fraser, C. (2020). The timing of covid-19 transmission. *medRxiv*. https://doi.org/10.1101/2020.09.04.20188516

10. 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

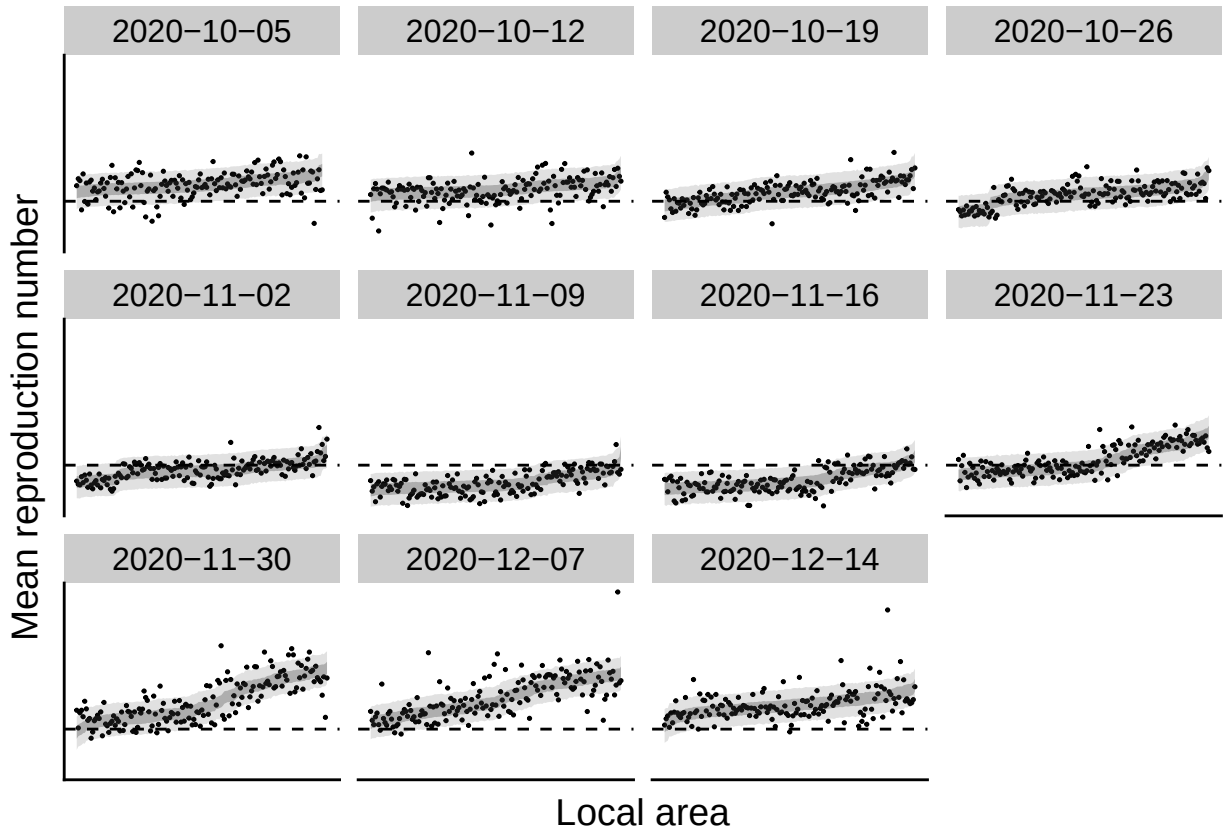
11. Davies, N. G., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J., Pearson, C. A. B., Russell, T. W., Tully, D. C., Abbott, S., Gimma, A., Waites, W., Wong, K. L., Zandvoort, K. vanEggo, R. M., Funk, S., Jit, M., Atkins, K. E., & Edmunds, W. J. (2020). Estimated transmissibility and severity of novel sars-cov-2 variant of concern 202012/01 in england. *medRxiv*. https://doi.org/10.1101/2020.12.24.20248822

12. Volz, E., Mishra, S., Chand, M., Barrett, J., Johnson, R., Geidelberg, L., Hinsley, W., Laydon, D., Dabrera, G., O'Toole, A., Amato, R., Manon Ragonnet-Cronin, I., Harrison, Jackson, B., Ariani, C., Boyd, O., Loman, N., McCrone, J., Gonçalves, S., Jorgensen, D., ... Ferguson, N. (2020). *Transmission of sars-cov-2 lineageb.1.1.7 in england: Insights from linkingepidemiological and genetic data*. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-42-sars-cov-2-variant/

Supplementary Information



Supplementary Figure 1: Mean reproduction numbers using a generation time with a mean of 5.5 days since the week beginning 14 September, 2020, compared to the proportion of all test-positives tested for S-gene that tested S-gene positive/negative that week. Each point represents one UTLA.



Supplementary Figure 2: Predictions of the best fitting long generation interval model (per-region spline) compared to the data (solid dots). Dark grey: central 50% prediction interval; light grey: central 90% prediction interval. Areas are ordered each week according to predicted median.