Local area reproduction numbers and S-gene dropouts

Preliminary analysis

Sam Abbott, CMMID Covid-19 Working Group, Sebastian Funk

5 January, 2020

This is preliminary analysis. Every attempt was made to avoid errors and acknowledge limitations but the code and analysis have not yet undergone peer review.

Abstract

Background: Regional Covid-19 effective reproduction number estimates have become increasing 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 which may be more transmissible than the wild type variant. In this report we aim to investigate the changes in the proportion of positive cases that were S-gene negative, an indicator of the novel variant, and correlate them with changes in the estimated reproduction number at the level of lower tier local authorities (LTLA).

Method: We calculated the weekly proportion of Covid-19 cases that were negative for the S-gene over time by local authority. We then explored the relationship between the proportion of cases that were S-gene negative and the effective reproduction number over time. Effective reproduction numbers were estimated using the `EpiNow2` R package independently for each local authority using test positive cases and two estimates of the generation time. The `brms` R package was used to fit a range of regression models with a custom probability distribution used to estimate a multiplicative effect of S-gene negativity. Models were compared using the expected log-predictive density though a range of models accounting for several scenarios were presented in the results.

Results: We found evidence for increased transmissibility across all models evaluated with the magnitude of the effect increasing as model flexibility was decreased. Models that did not adjust for residual variation over time found a comparable effect to 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 that S-gene negativity was associated with a reduced, but still large, increase in transmissibility.

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 transmissibility 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 was a 19% increase in transmissibility which implies that NPI measures implemented since September in England may not be sufficient to reduce the reproduction number below 1.

Method

Data

We used 4 main sources of data: test positive lower-tier local authority Covid-19 notifications, S-gene status from PCR tests by local authority, Google mobility data stratified by context, and a database of non-

pharmaceutical interventions (NPI). The data used for this analysis is available from https://github.com/epiforecasts/covid19.sgene.ltla.rt/.

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 (Abbott, Hellewell, et al. 2020b) and (Sherratt et al. 2020) and implemented in the `EpiNow2` R package (Abbott, Hellewell, et al. 2020a). 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) (Abbott, Hellewell, et al. 2020a; Ganyani et al. 2020) or with a mean of 5.5 days (SD: 0.5), and SD of 2.1 days (SD: 0.25 days) (Ferretti et al. 2020), respectively.

We then built a separate model of the expected reproduction number in LTLA i during week t starting in the week beginning 12 October, 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 LTLA-level intercept corresponding to R 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 LTLA 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 LTLA is

$$R_{t,i} = (1 + \alpha s_{it})R_{t,i}^+ + s_{it}R_{t,i}^-$$

where $R_{t,i}^-$ is the S-gene negative reproduction number of $R_{t,i}^+$ is the S-gene positive reproduction number.

We used a Gaussian observation model with a single variance. The model was implemented using the `brms` (Bürkner 2018) package in `R`. All code required to reproduce this analysis is available from https://github.com/epiforecasts/covid19.sgene.ltla.rt/.

Results

We found consistent evidence of an association between S-gene negativity and increased LTLA 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 2) but all models had evidence of increased transmissibility due to S-gene negativity with a lower bound in the best fitting model of 19% (Table 3). Using a longer generation time, a model that only adjusted for national level residual variation over time, or a model that did not adjust for residual variation over time increased this lower bound to 26%, 21% and 33% respectively. Similarly, the upper bound of the increase in transmissibility varied from 37% to 52% in models that fit the data comparably, and increased to a maximum of 78% in a model, with a generation time of 5.5 days, that assumed all regional variation not otherwise adjusted for was due to S-gene negativity.

The best fitting model appeared to reproduce estimated reproduction numbers to an acceptable degree (Figure 2) though it could not fully capture the observed variation in the data.
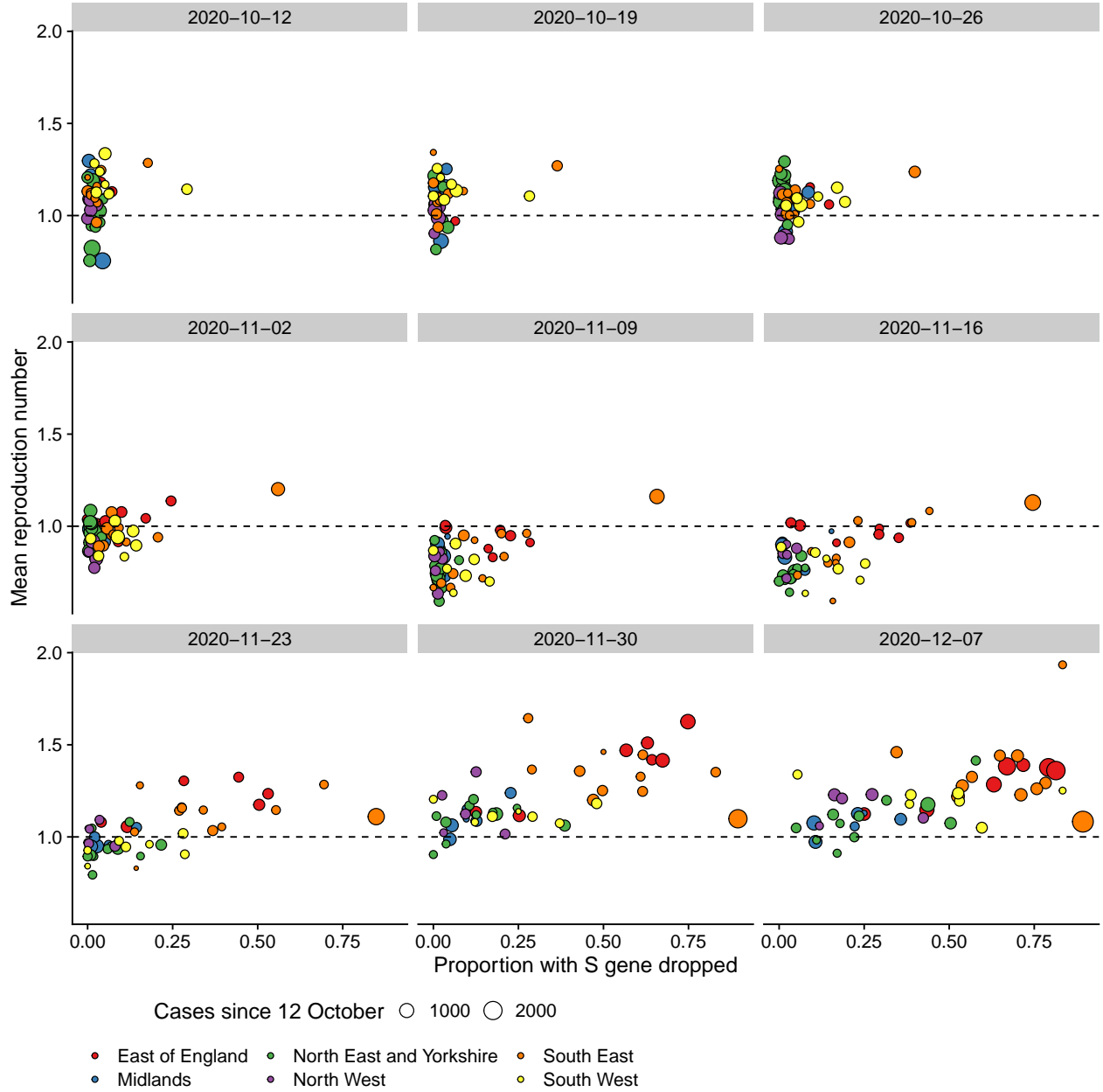

Figure 1: Mean reproduction numbers using a generation time with a mean of 3.6 days since the week beginning 12 October, 2020, compared to the proportion of all test-positives tested for S-gene that tested S-gene positive/negative that week. Each point represents one LTLA.

Model	ELPD difference
Regional static	-96.31
National time-varying	-6.03
Regional time-varying	0.00

Table 2: Model comparison by difference in expected log-predictive density.

Model	Estimate (short GT)	Estimate (long GT)
Regional static	0.42 (0.33-0.52)	0.64 (0.5-0.78)
National time-varying	0.3 (0.21-0.39)	0.45 (0.32-0.58)
Regional time-varying	0.28 (0.19-0.37)	0.38 (0.26-0.52)

Table 3: 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.
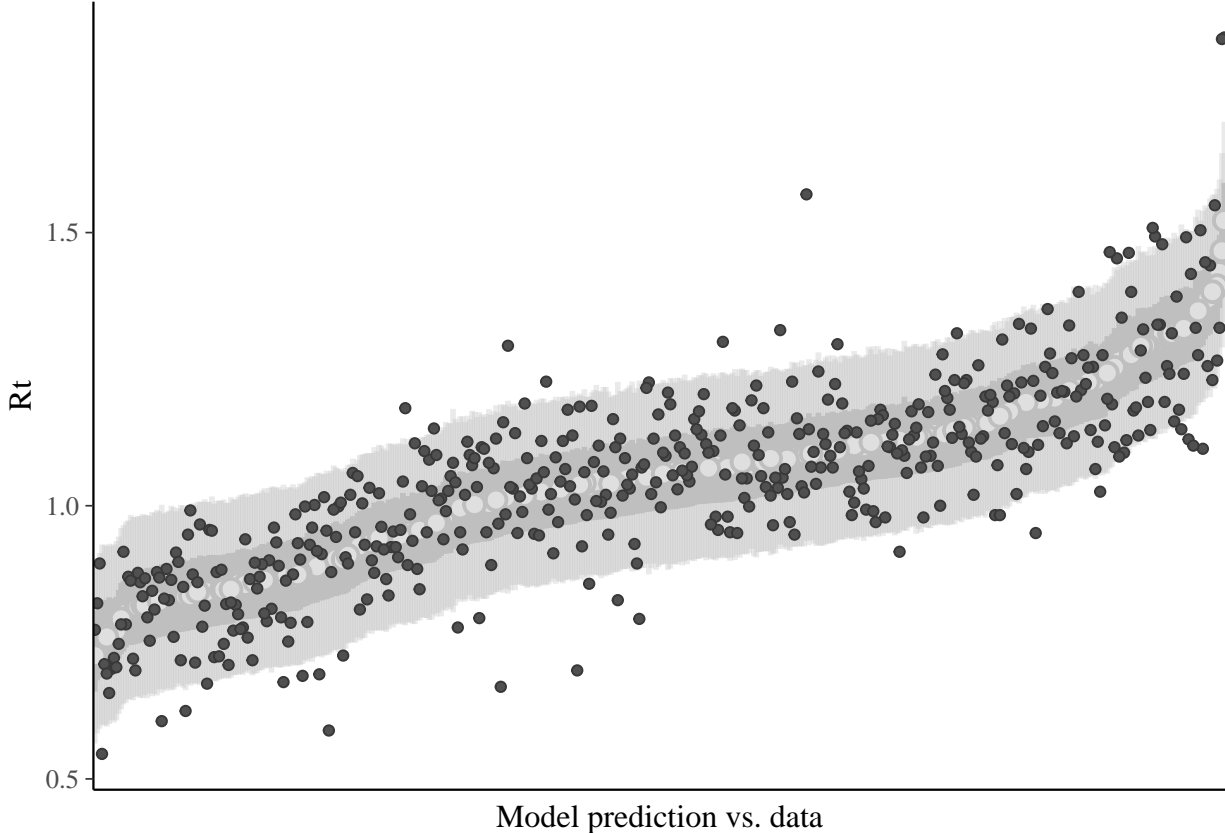

Figure 2: Predictions of the best fitting short generation interval model (per-region spline) compared to the data (solid dots).

Discussion

We studied the relationship between S-gene dropout (as a proxy for the new variant of concern) and reproduction numbers using three related models that had varying degrees of flexibility in ascribing changes in the reproduction numbers to factors not explained by the proportion of cases with S-gene dropout. The model with region-specific splines suggests the smallest effect (central estimate: 28% increase and 38% with short and long generation interval, respectively). This might be an underestimate as it can explain regional differences by unmodelled factors, i.e. ones beyond interventions, mobility, and distribution of S-gene dropouts, and is therefore largely a model of within-regions LTLA-level differences. The largest effect of S-gene dropout (central estimate: 42% increase and 64% increase with short and long generation interval, respectively) was seen in the model that modelled all temporal variation as related to interventions, mobility and S-gene dropout. 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% (Davies et al. 2020) or 50-75% (Volz et al. 2020). 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.
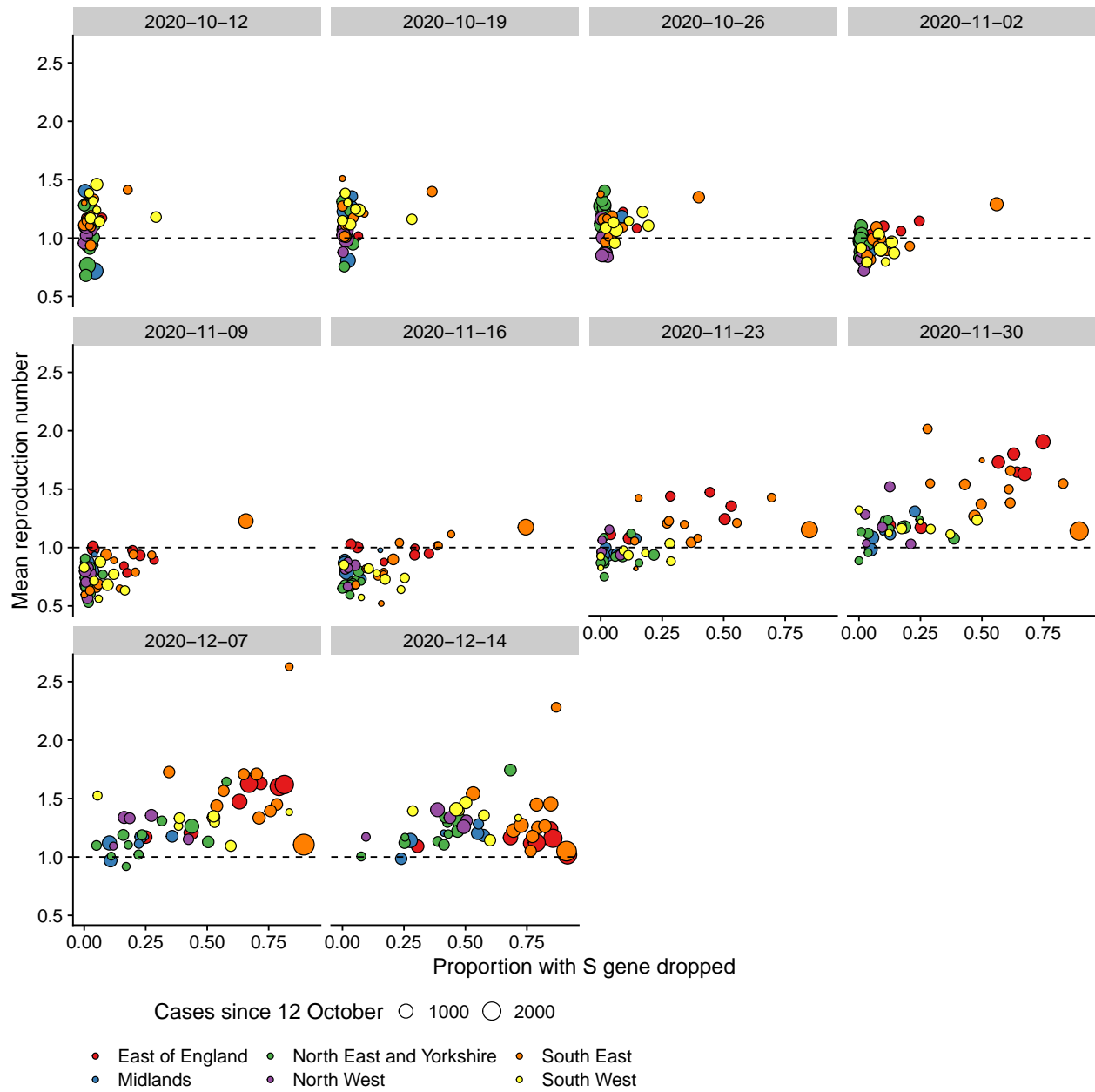
Our results should be treated with caution as several caveats apply: we have not observed any local authorities in which all tests were S-gene negative and therefore are extrapolating beyond the available data. We assumed that the effect of tiers and lockdown applied uniformly across the country. While we did allow for a flexible regional-level behaviour through our use of regression splines, there may be LTLA 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 LTLA 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 transmissbility 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

Abbott, Sam, Joel Hellewell, Katharine Sherratt, Katelyn Gostic, Joe Hickson, Hamada S. Badr, Michael DeWitt, Robin Thompson, EpiForecasts, and Sebastian Funk. 2020a. *EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters.* https://doi.org/10.5281/zenodo.3957489.

Abbott, Sam, J Hellewell, RN Thompson, K Sherratt, HP Gibbs, NI Bosse, JD Munday, et al. 2020b. "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.

Bürkner, Paul-Christian. 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.

Davies, Nicholas G., Rosanna C. Barnard, Christopher I. Jarvis, Adam J. Kucharski, James Munday, Carl A. B. Pearson, Timothy W. Russell, et al. 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.

Ferretti, Luca, Alice Ledda, Chris Wymant, Lele Zhao, Virginia Ledda, Lucie Abeler- Dorner, Michelle Kendall, et al. 2020. "The Timing of Covid-19 Transmission." *medRxiv.* https://doi.org/10.1101/2020.09.04.20188516.

Ganyani, Tapiwa, Cécile Kremer, Dongxuan Chen, Andrea Torneri, Christel Faes, Jacco Wallinga, and Niel Hens. 2020. "Estimating the Generation Interval for Coronavirus Disease (Covid-19) Based on Symptom Onset Data, March 2020." *Eurosurveillance* 25 (17). https://doi.org/https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257.

Sherratt, Katharine, Sam Abbott, Sophie R Meakin, Joel Hellewell, James D Munday, Nikos Bosse, Mark Jit, and Sebastian Funk. 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.

Volz, E, S Mishra, M Chand, JC. Barrett, R Johnson, L Geidelberg, WR Hinsley, et al. 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 12 October, 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 LTLA.
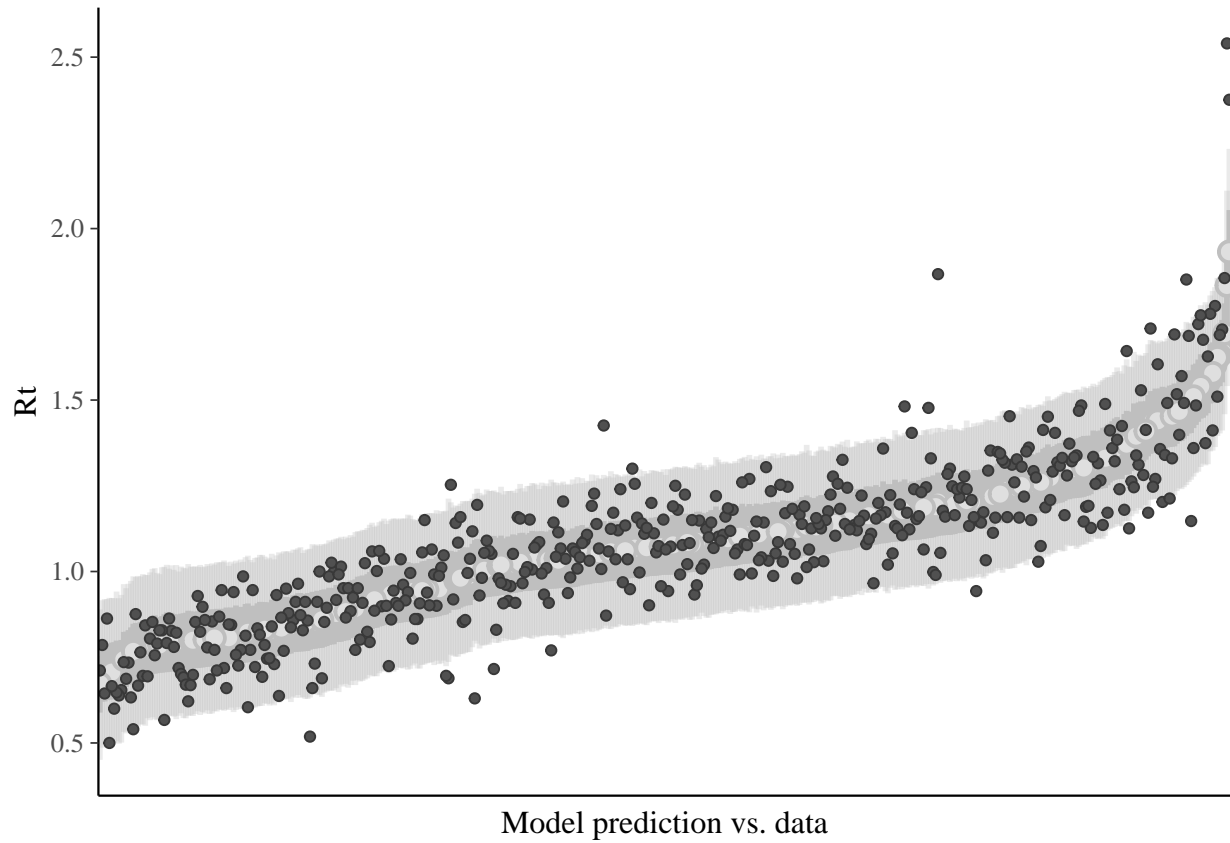
Supplementary Figure 2: Predictions of the best fitting long generation interval model (per-region spline) compared to the data (solid dots).