Estimating the increase in reproduction number associated with the Delta variant using local area dynamics in England

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

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

Abstract

Background: Local estimates of the time-varying effective reproduction number (Rt) of Covid-19 in England become increasingly heterogeneous in England during April and May. This variation may have been attributed to the spread of novel SARS-CoV-2 variant "Delta." In this report we aimed to investigate the changes in the proportion of positive cases that were S-gene positive, an indicator of the novel variant in front of the background of the previously predominant variant "Alpha," 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 positive 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 S-gene positivity and weekly mean R estimate. Results are shown with 90% credible intervals.

Results: We found evidence of an association between increased mean Rt estimate and the proportion of S-gene positives across all models evaluated with the magnitude of the effect increasing as model flexibility was decreased. Models that adjusted for either national level or NHS region level residual variation over time were found to fit the data better, suggesting a potential role of unexplained confounders.

Conclusions: Our results indicate that even after adjusting for residual variation over time between NHS regions S-gene positivity 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 the reproduction number will remain above 1 at current levels of behaviour.

Limitations: Associations are purely observational and therefore subject to bias due to confounders not included in the model. Surge testing and infections brought in by travellers may have affected the case numbers and, consequently, reproduction number estimates. Our analysis does not distinguish between increased reproduction numbers due to transmission advantage and transmission advantage due to vaccine escape and therefore cannot distinguish between the two. Whilst we did investigate two different generation interval distributions we did not consider a scenario where the generation times were different between S-gene positives and negatives.

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 provided by Public Health England (PHE), Google mobility data stratified by context,[2] and data on the timings of national easing. We aggregated the data at the weekly level and restricted the

analysis to the period beginning Tuesday, 23 February and ending Tuesday, 25 May. We further restricted the analysis to UTLA/week combinations where more than 20% of reported cases and more than 20 in total had an S-gene status reported.

Statistical analysis

We calculated the weekly proportion of positive tests that were S-gene positive over time by local authority. We estimated reproduction numbers using the method described in[3] and[4] and implemented in the `EpiNow2` R package.[5] 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)[5,6] or with a mean of 5.5 days (SD: 0.5 days), and SD of 2.1 days (SD: 0.25 days),[7] respectively.

We then built a separate model of the expected reproduction number in UTLA i during week t starting in the week beginning 23 February, 2021, as a function of local restrictions, mobility indicators, residual temporal variation, and proportion of positive tests S-gene positive:

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

where R_t is an UTLA-level intercept corresponding to Rt during national lockdown in February/March, T_{ijt} is 1 if intervention j (out of: January lockdown, reopening phase 1/2) is in place and 0 otherwise, G_{ikt} is the relative mobility in context k (home, workplace, public transport) 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"). We considered the model with only a static regional parameter our baseline model as it yielded the most directly interpretable parameter estimates, assuming reproduction numbers could be completely explained by the relaxation steps and spread of S-gene positivity. The spline versions were designed to capture counfounding due to unmeasured covariates over time and were therefore considered as lower bounds on effect estimates. We lastly fitted the model for each t separately, i.e. at fixed time slices ("Time-sliced").

The key parameter is α, the relative change in reproduction number in the presence of s-gene positivity 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 negative, 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 Student's t-distribution observation model with a single variance parameter and a single degrees of freedom parameter. All models were implemented using the `brms`[8] package in `R`. All code required to reproduce this analysis is available from https://github.com/epiforecasts/covid19.sgene.utla.rt/ (`may-update` branch).

Results

We found consistent evidence of an association between S-gene positivity and increased UTLA level reproduction number estimates. The association became more apparent over time. The outer 90% credible intervals of the estimates in the time-sliced analysis (Fig. 2) ranged from a -10% increase to a 113% increase, in April/May, depending on the assumption about generation times and the model used, as the proportion of of tests that were S-gene positive increased heterogeneously across NHS regions. Out of the models fitted to all time, ones that adjusted for residual variation over time on both a national and NHS region level fit

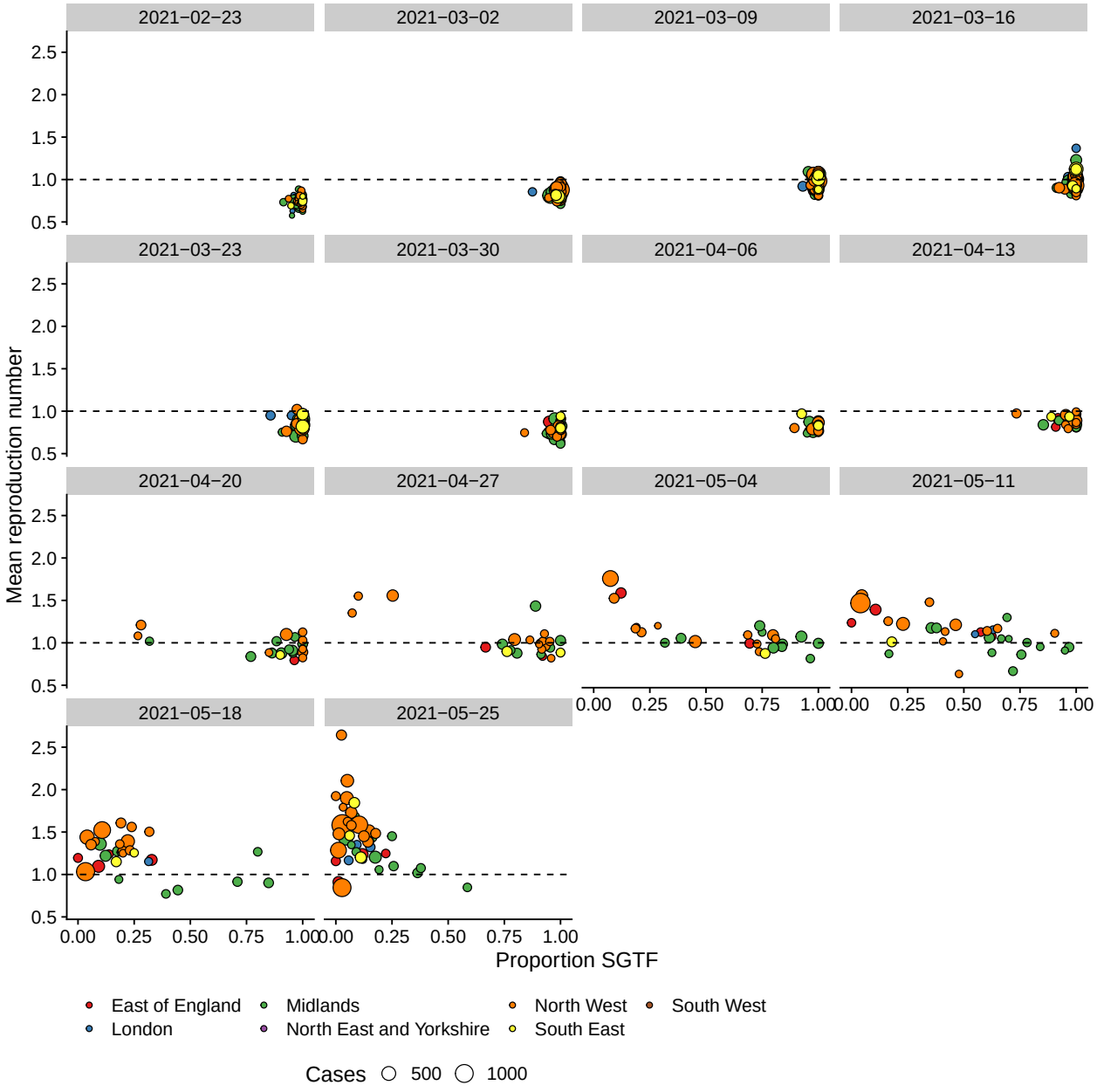

Figure 1: *Mean reproduction numbers using a generation time with a mean of 5.5 days since the week beginning 23 February, 2021, compared to the proportion of all test-positives tested for S-gene that tested S-gene positive/negative that week. Each point represents one UTLA, with the size given by the number of cases in the week following the week of the given reproduction number to account for the delay from infection to testing. Only UTLAs with sufficient coverage of S-gene results at least 20% of cases tested and at least 20 results in total)*

Figure 2: *Estimates of α from the time-sliced model.*

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

Model	ELPD difference	Standard error
Short generation time		
Regional time-varying	0	0
National time-varying	-30	11
Regional static	-119	14
Long generation time		
Regional time-varying	0	0
National time-varying	-32	12
Regional static	-122	15

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 positive cases.*

Model	Estimate
Short generation time	
Regional static	0.28–0.4
National time-varying	0.21–0.35
Regional time-varying	0.2–0.33
Long generation time	
Regional static	0.38–0.55
National time-varying	0.28–0.47
Regional time-varying	0.27–0.46

the data better than those that did not (Tab. 1). However, all models had evidence of increased Rt with S-gene positivity with the best fitting model yielding a lower bound of 20% higher Rt of S-gene positive cases with a short generation time (Tab. 2), higher than the model that only adjusted for national level residual variation over time (lower bound: 21%) and lower than the model that did not adjust for residual variation over time (lower bound: 28%). With a longer generation time, these lower bounds changed to 27%, 28%, and 38%, respectively. The upper bound of the increase in Rt varied from 33% to 55% in models with different assumed generation times.

The model that did not for residual variation appeared to reproduce estimated reproduction numbers relatively well over time (Fig. 3) although there are notable outliers especially in early March, when some of the R estimates may have been affected by the results of mass testing in schools entering the case data, and in some of the later weeks when Delta was increasingly prevalent. This model also yielded estimates of the relative impact of the different steps of reopening on R, with a relatively small effect in the order of 5-20% each time, but combinedly possibly lifting R close to 1 at the time of the third step of reopening even with the previously circulating Alpha variant.

Discussion

We studied the relationship between S-gene positivity (as a proxy for the new variant of concern Delta) 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 that associated all temporal variation with interventions, mobility and SGTF produced the largest increases (central estimates: 46% and 34% increase with long and short generation interval, respectively). A more conservative, if more difficult to interpret, model with region-specific splines yielded central estimates of a 26% and 36% increase in R with short and long generation interval, respectively. While this model yielded the best fit to the data the estimate of increased R associated with SGTF may well be an underestimate

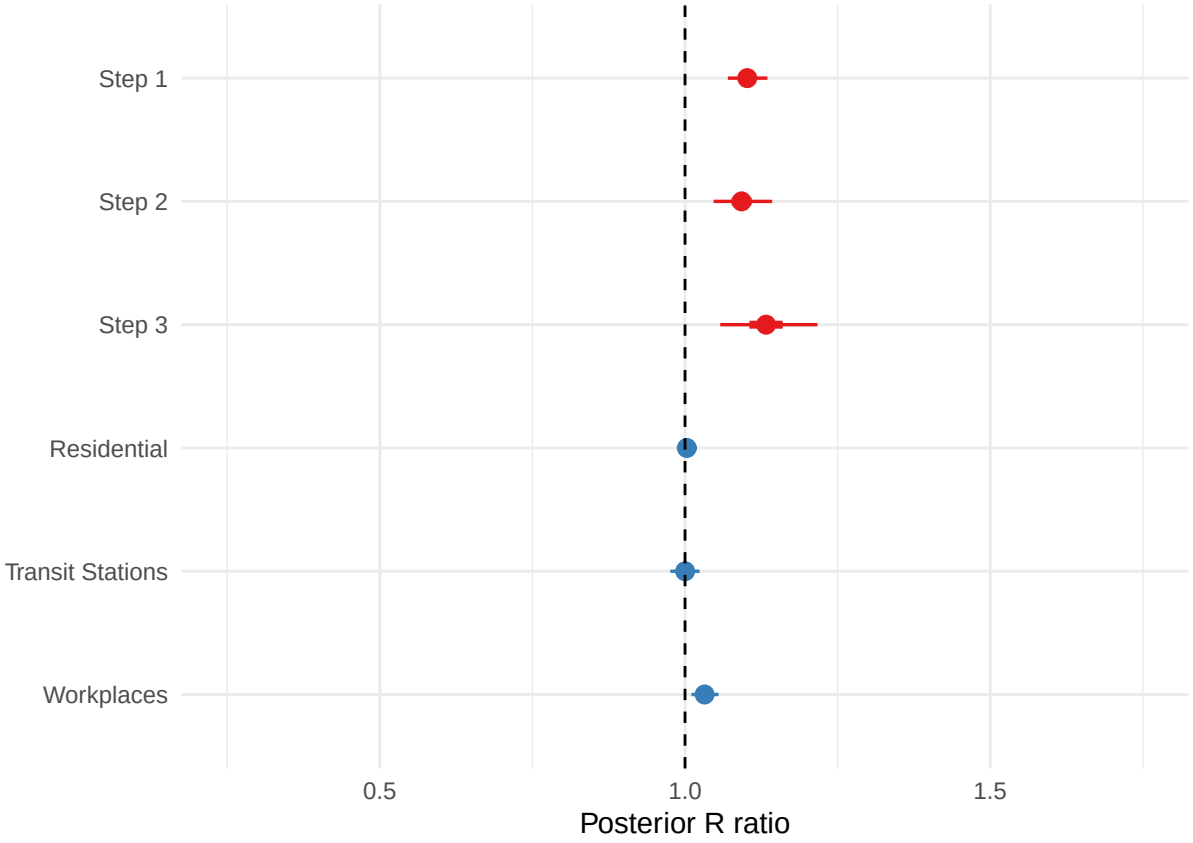

Figure 3: *Parameter estimates (R ratios) of the Regional static model. These correspond to a multiplicative effect on a baseline R of 0.74–0.81). Intervention indicators are coloured in red, and mobility indicators in blue.*

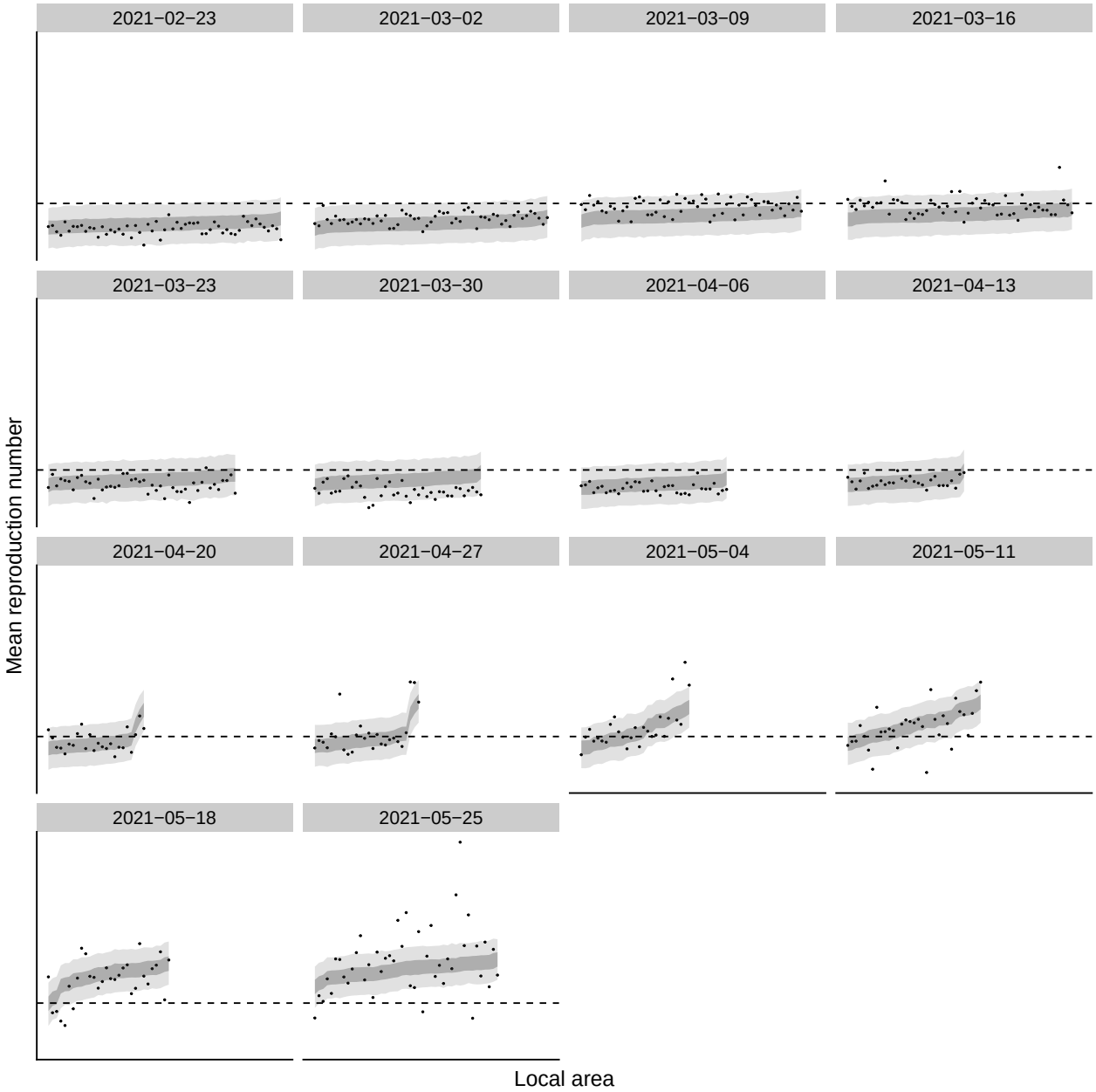

Figure 4: *Predictions of the regional static model with long generation interval compared to the data (solid dots). Dark grey: central 50% prediction interval; light grey: central 90% prediction interval. Areas are ordered each week according to predicted median.*

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-region UTLA-level differences. On the other hand, the fact that this model yielded a better fit than the model without splines suggests that that model is affected by confounders not included in the model.

Our estimates with the longer generation interval without adjusting for additional residual variation over time are consistent with ones from other preliminary rapid analyses using similar generation intervals (between 5-6 days), which were in the order of a 13-57%,[9] 20-60%[10] increase, yet lower than another estimate of 50-100%.[11] They are also consistent with increased secondary attack rates from contact tracing during relevant time periods, increased from 8.2% (8.0%-8.4%) for Alpha to 12.4% (11.7%-13.2%), corresponding with an increased reproduction number of around 50%.[12] That said, other model variants yield lower estimates consistent with the data, as do shorter generation intervals, which generally lead to reproduction numbers closer to 1 and thus lower estimates of a multiplicative effect. This may be particularly relevant where the effect of the variant cause the reproduction number to cross 1.

Our results should be treated with caution as several caveats apply: we assumed that S-gene positive and negative cases had the same generation interval, while a complementary hypothesis might be that the new variant shortened the generation interval, affecting our estimates.[13] 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 as sensitivity analysis, 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 positives observed in every UTLA per week, which were treated as fixed point estimates. Because of this, uncertainty in our regression coefficients are underestimated, and probably considerably so. Improving our inference method to incorporate these uncertainties is a future aim of our research. Lastly, whilst Delta increased rapidly throughout England over the period investigated, other variants as well as, on occasion, the previously predominant Alpha variant may yield S-gene positives. We therefore cannot rule out that our effect includes a component not related to the new variant.

This analysis was done rapidly when Delta started increasing in England, in response to a need to gather scientific evidence for the scale of the transmission advantage. A previous version was considered by the Scienctific Pandemic Influenza Group on Modelling, SPI-M.[14] We found consistent evidence that S-gene positivity 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 at the levels of restrictions on contacts preent in England though April to June were not sufficient to reduce the reproduction number below 1 with the emerging Delta variant.

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. 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
4. 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
5. Abbott, S., Hellewell, J., Sherratt, K., Gostic, K., Hickson, J., Badr, H. S., DeWitt, M., Thompson, R., Epi-Forecasts, & Funk, S. (2020). *EpiNow2: Estimate real-time case counts and time-varying epidemiological*

parameters. https://doi.org/10.5281/zenodo.3957489
6. 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
7. 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
8. 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
9. Keeling, M. (2021). *Estimating the transmission advantage for b.1.617.2*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993156/S1269_WARWICKTransmission_Advantage.pdf.
10. Adam Kucharski, R. E., Nicholas Davies. (2021). *Dynamics of b.1.617.2 in england NHS regions from importations, traveller-linked and non-traveller-linked transmission*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993232/S1272_LSHTM_Modelling_Paper_B.1.617.2.pdf.
11. Ferguson, N. (2021). *B.1.617.2 transmission in england: Risk factors and transmission advantage*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993159/S1270_IMPERIAL_B.1.617.2.pdf.
12. Public Health England. (2021). *SARS-CoV-2 variants of concern and variants under investigation in england. Technical briefing 14*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/991343/Variants_of_Concern_VOC_Technical_Briefing_14.pdf.
13. Sang Woo Park, S. F., Benjamin M. Bolker. (2021). Roles of generation-interval distributions in shaping relative epidemic strength, speed, and control of new SARS-CoV-2 variants. *medRxiv*. https://doi.org/10.1101/2021.05.03.21256545
14. Sam Abbott, C. W. G., Sebastian Funk. (2021). *Local area reproduction numbers and s-gene positivity*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993228/S1271_Local_area_reproduction_numbers_and_S-gene_positivity.pdf.