Real-time estimation of the time-varying transmission advantage of Omicron in England using S-Gene Target Status as a Proxy

Summary report

Sam Abbott (1), Katharine Sherratt (1), Sebastian Funk (1)

December 24, 2021

1. Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, United Kingdom

Estimates are updated daily as new data is released but the summary of the situation is updated less frequently and so may not match current estimates. All data (both raw data and estimates) are available here and the report should be fully reproducible. Reports and data as available at the time of release are available from the release page. See our news file for details of what updates were made when.

Introduction

Since being highlighted by scientists in South Africa Omicron has spread rapidly globally. Using initial South African data, it has been estimated that Omicron may be both more transmissible and have greater immune escape than the previously dominant Delta variant[1].

In this work, we use S-gene target failure (SGTF) as a proxy of variant status combined with reported case counts to explore the evidence for changes in transmission advantage over time for the Omicron variant. If present this could indicate the impact of immune escape, sampling bias in SGTF data or differences in the populations within which the variants are circulating. We also report estimates for growth rates by variant and overall, case counts overall and by variant for a 14 day forecast window assuming constant future growth, the date at which Omicron will become dominant in England and in each UKHSA region, and the estimated cumulative percentage of the population with a reported Omicron case. We also explore the potential for bias in S-gene target sampling by comparing cases that have a reported SGTF status to cases with an unknown SGTF status.

Methods

Data

We use S-gene status by specimen date sourced from UKHSA as a direct proxy for the Omicron variant with a target failure indicating a case has the Omicron variant. We augment this data with reported cases counts by date of specimen truncated by two days to account for delayed reporting. Data was available for England and for UKHSA region from both sources.

Models

We consider two autoregressive models with different assumptions about the relationship between growth rates for Omicron and non-Omicron cases. Both models estimate expected cases for each variant as a combination of expected cases from the previous day and the exponential of the log growth rate. The growth rate is then itself modelled as a differenced AR(1) process.

For the first model variants are assumed to have growth rates related by a fixed scaling (described from now on as the scaled model). In this model, variant growth rates then share a single differenced AR(1) process meaning that they vary over time in the same way. This model assumes that variants differ only due to a transmission advantage and that there are no time-varying biases in the reported data.

In the second model, we relax the assumption that variants co-vary using a vector autoregression structure which assume that the 1st differences of the variant growth rates are drawn from a multivariate normal distribution. This model formulation can account for variant differences other than a transmission advantage and can also better handle time-varying biases in data sources. Crucially, in the absence of evidence that variants do not co-vary, it reduces to the co-varying model.

For both models, we fit jointly to reported cases and SGTF data assuming a negative binomial and beta-binomial observation model respectively. Day of the week reporting periodicty for case counts is captured using a random effect fo r the day of the week. We initialise both models by fitting to a week of case only data where the Omicron variant is assumed to not be present (from the 17th of November to the 22nd).

A full description of the models described here can be found in the documentation for the `forecast.vocs` R package[2].

Statistical Inference

We first visualised our combined data sources (cases by specimen date and SGTF status by specimen date). We then fit both models separately to data for England and to each UKHSA region. Using these model fits we report posterior estimates from the best fitting model for the following summary statistics of epidemiological interest: growth rates by variant and overall, the time-varying transmission advantage for the Omicron variant, case counts overall and by variant for a 14 day forecast window assuming constant future growth, the date at which Omicron will become dominant in England and in each UKHSA region, and the estimated cumulative percentage of the population with a reported Omicron case.

We explored the potential for bias in S-gene target sampling by comparing cases that have a reported SGTF status to cases with an unknown SGTF status. We fit the correlated model to case counts and S-gene tested status and reported the apparent transmission advantage for those with a S-gene status over time. Any variation in this metric from 100% may be interpreted as indicating biased sampling of those with known S-gene status, which may bias our main results.

Implementation

All models were implemented using the `forecast.vocs` R package[2,3] and fit using `stan`[4] and `cmdstanr`[5]. Each model was fit using 2 chains with each chain having 1000 warmup steps and 2000 sampling steps. Convergence was assessed using the Rhat diagnostic[4]. Models were compared using approximate leave-one-out (LOO) cross-validation[6,7] where negative values indicate an improved fit for the correlated model.

Limitations

- SGTF may be an imperfect proxy for Omicron. We make no adjustment for the background rate of SGTF observed for Delta or other factors that might lead to SGTF, or a confirmed detected S-gene observed with Omicron.

- The public case data used here does not include known reinfections. If the proportion of infections that are reinfections varies over time and between variants then this could affect our estimates.

- We make a crude adjustment for the right censoring of reported case counts by specimen date by excluding the last two days of data. However, this may be insufficient to correct for the bias and so our real-time growth rates estimates may be biased downwards.

- Our analysis is based on reported case counts augmented with SGTF data. Growth rates estimated from reported case counts may be biased during periods when changes in testing practice occur or when available testing is at capacity.

- We reported the log growth rate approximation to the growth rate. This approximation becomes less valid for high absolute growth rates.

- Our models do not attempt to capture the underlying transmission process or delays in reporting. In particular, we do not consider the generation time of variants which may differ. Differences in the generation time would result in apparent differences in the growth rate (on the scale of the generation time) even in scenarios where growth rates were fixed.

- Our forecasts assume a fixed growth rate across the forecast horizon and do not account for population susceptibility or other factors such as changes in contact patterns.

- Our model only estimates an overall transmission advantage. We do not attempt to distinguish between "intrinsic" transmissibility and immune escape, a combination of which is likely to be underlying any sustained differences in growth rate.

- Our estimates are at the national or UKHSA region scale. There may be additional variation within these areas that these estimates cannot capture.

- We used an approximate form of leave-one-out cross validation to compare the performance of the models we considered. This may be inappropriate in situations where performance on held out time-series data is the target of interest.

Results

Summary

Last updated: 2021-12-23

- A model that assumed that variants growth rates varied over time relative to each other performed better in all regions than a model where the relative difference was constant over time.

- Prior to the introduction of Omicron the growth rate of non-Omicron cases was positive in most regions. In the last few weeks the growth rate for non-Omicron cases has become negative suggesting that if nothing changes Omicron will be the only circulating variant at some point in the future.

- The growth rate for Omicron cases has been consistently positive in all regions since the 23rd of November though with slightly different absolute values and with different variation over time.

- In the last week both case data and the SGTF data suggest that the growth rate of Omicron in reported cases has reduced markedly though at this time still appears positive in all regions excluding London.

- Assuming a fixed difference in growth between variants indicates an advantage for Omicron in the order of 30% to 50% across all regions. Allowing this to vary (which was shown to better represent the data) implies that this advantage generally appears to have increased over time until the last week where it fell back to initially observed levels or below in all regions.

- The variation in transmission advantage also differed by region with some regions having a fairly stable trend (such as the South West) and some (such as the North East) indicating large changes.

- Assuming growth rates stay as currently estimated we forecast that approximately 1.5% to 2.5% of the English population will have reported an Omicron case by the end of 2021, although with significant regional variation. This does not take into account any changes in testing behaviour or capacity.

- We found some evidence of bias in sampling S-gene status though this varied across regions and over time. There was particularly strong evidence of biased sampling over time in the East Midlands. In other regions, the evidence of bias was substantially weaker and therefore less likely to bias our main findings.

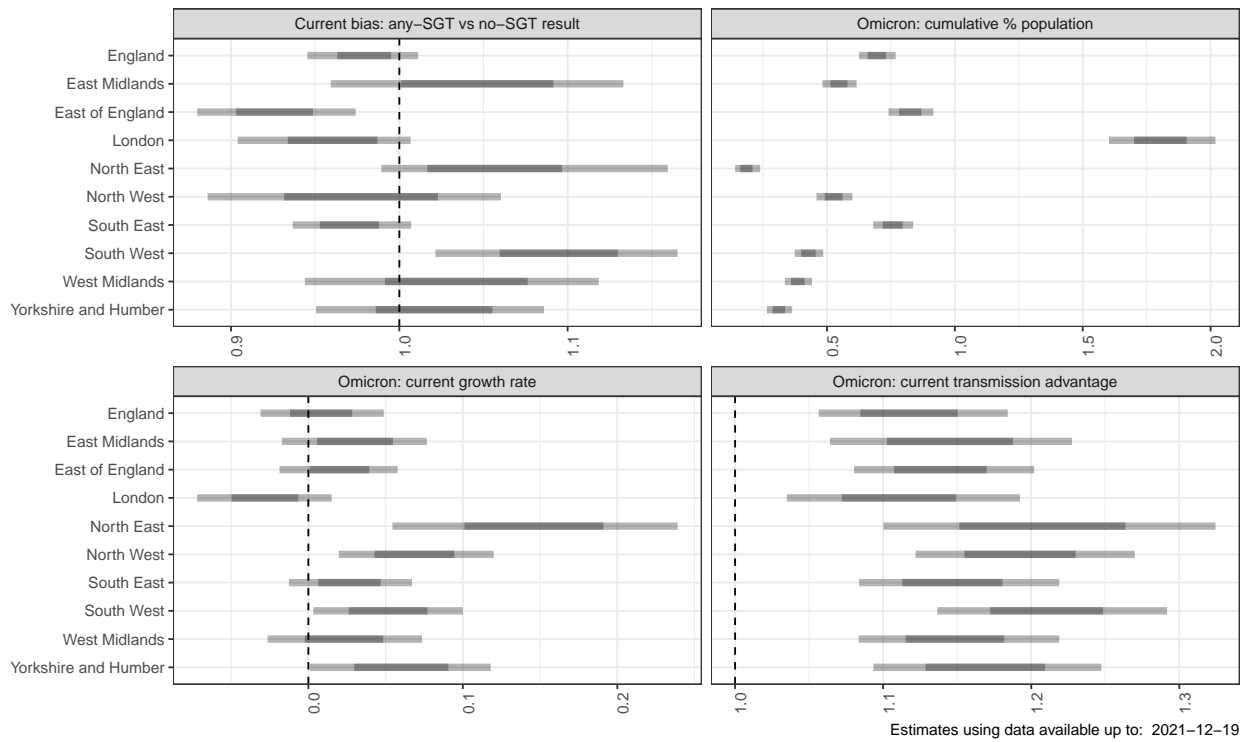**Latest available estimates using data up to: 2021-12-19**

Figure 1: Summary of latest estimates for England and by UKHSA region, showing: bias in S-gene sampling, comparing the Omicron transmission advantage among cases with any SGT result, against cases with no SGT result (at 1); cumulative % of population reported as a case with Omicron variant; current growth rate, and current transmission advantage of Omicron variant. 90% and 30% credible intervals are shown.
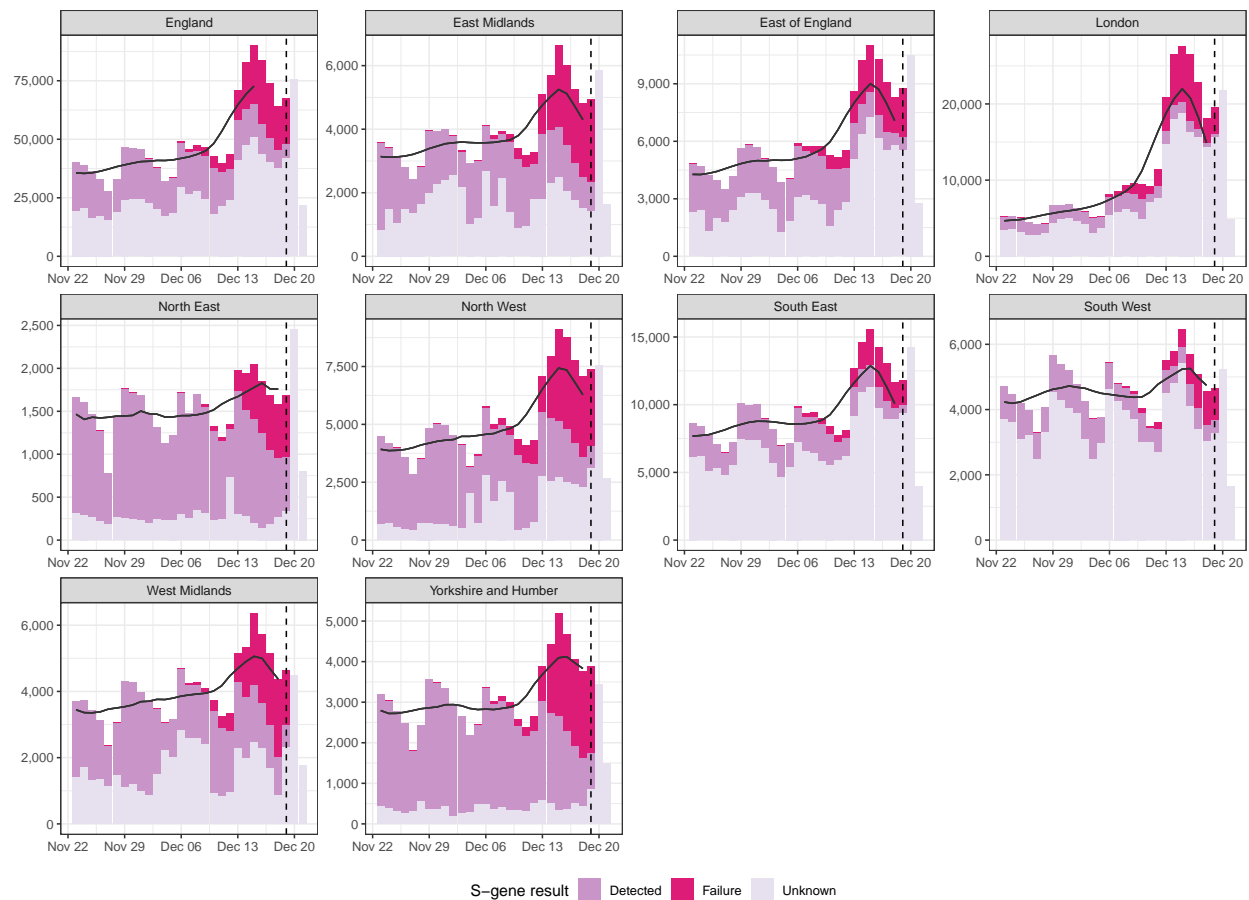
Data description

Model comparison

Figure 2: Daily cases in England and by UKHSA region, with S-gene target result (failed, confirmed detected, or unknown), and centred 7-day moving average up to date of data truncation (dotted line). Source: UKHSA and coronavirus.gov.uk; data by specimen date.

Table 1: Estimated differences in the ELPD metric (with standard errors) between the scaled and correlated models for England and UKHSA regions. Negative values indicate that the correlated model is estimated to be a better fit to the data.

Region	Elpd diff	Se diff
England	-16	6
East Midlands	-13	4
East of England	-17	4
London	-15	4
North East	-13	3
North West	-10	3
South East	-13	4
South West	-7	3
West Midlands	-21	3
Yorkshire and Humber	-20	4

Growth rate of reported cases overall, with the Omicron variant, and not with the Omicron variant

Proportion of cases with the Omicron variant on the logit scale

Natural scale

Logit scale

Date at which Omicron estimated to account for 95% of reported cases

Time-varying transmission advantage of the Omicron variant

Posterior predictions and forecasts of reported cases

Natural

Log

Cumulative percentage of the population with a reported Omicron case

Evidence of sampling bias in S-gene target results among Covid-19 test-positive cases

References

1. Pearson, C. A. B., Silal, S. P., Li, M. W. Z., Dushoff, J., Bolker, B. M., Abbott, S., Schalkwyk, C. van, Davies, N. G., Barnard, R. C., Edmunds, W. J., Bingham, J., Meyer-Rath, G., Jamieson, L., Glass, A., Wolter, N., Govender, N., Stevens, W. S., Scott, L., Mlisana, K., . . . Pulliam, J. R. C. (2021). *Bounding the levels of transmissibility & immune evasion of the omicron variant in south africa.* https://www.sacmcepidemicexplorer.co.za/downloads/Pearson_etal_Omicron.pdf
2. Abbott, S. (2021). Forecast.vocs: Forecast case and sequence notifications using variant of concern strain dynamics. *Zenodo.* https://doi.org/10.5281/zenodo.5559016
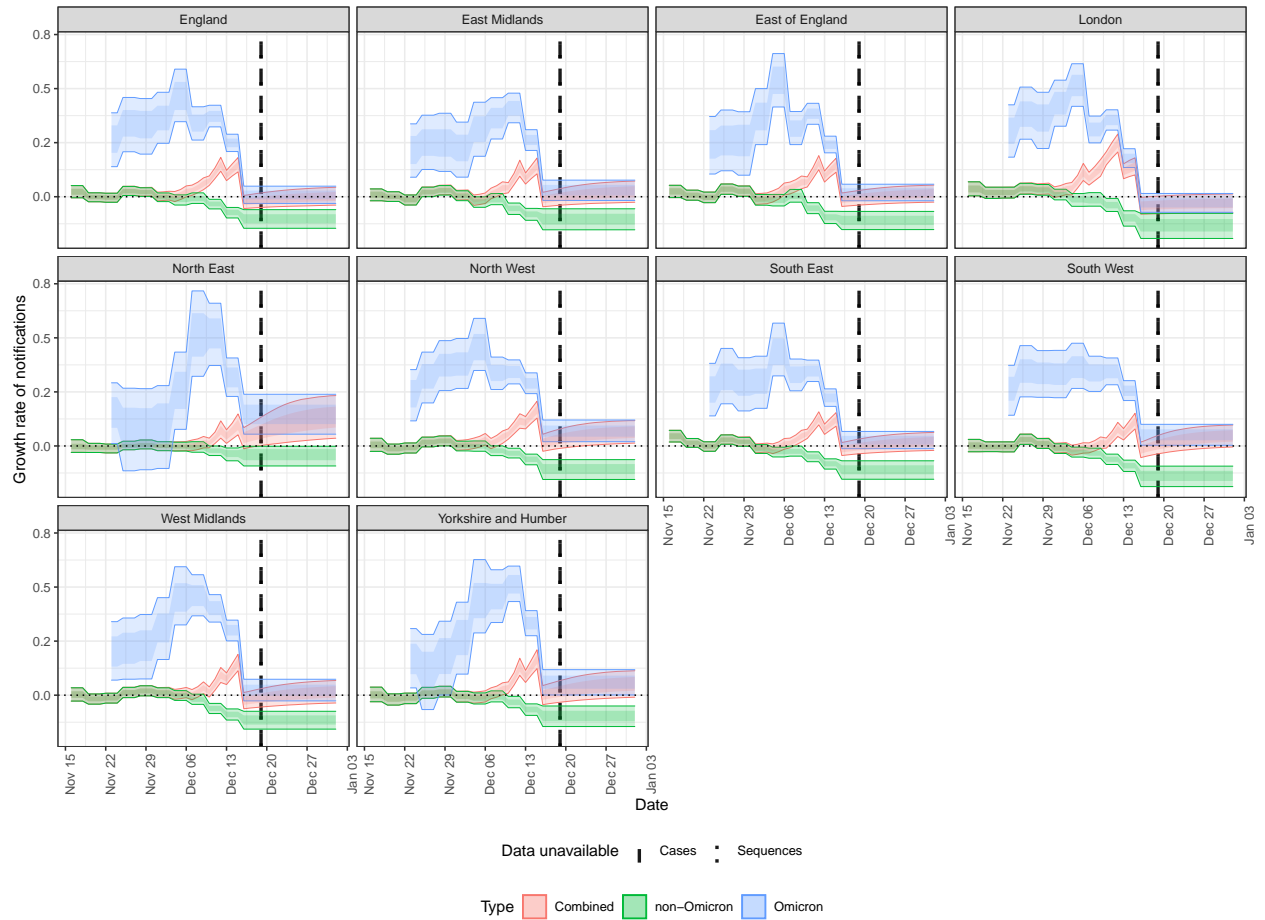
Figure 3: Estimates of the time-varying daily growth rate for England and by UKHSA region (using the log scale approximation) for reported cases overall, cases with the Omicron variant, and non-Omicron cases. Growth rates are assumed to be piecewise constant with potential changes every 3 days. Growth rates are assumed to be constant beyond the forecast horizon (vertical dashed line). 90% and 60% credible intervals are shown.
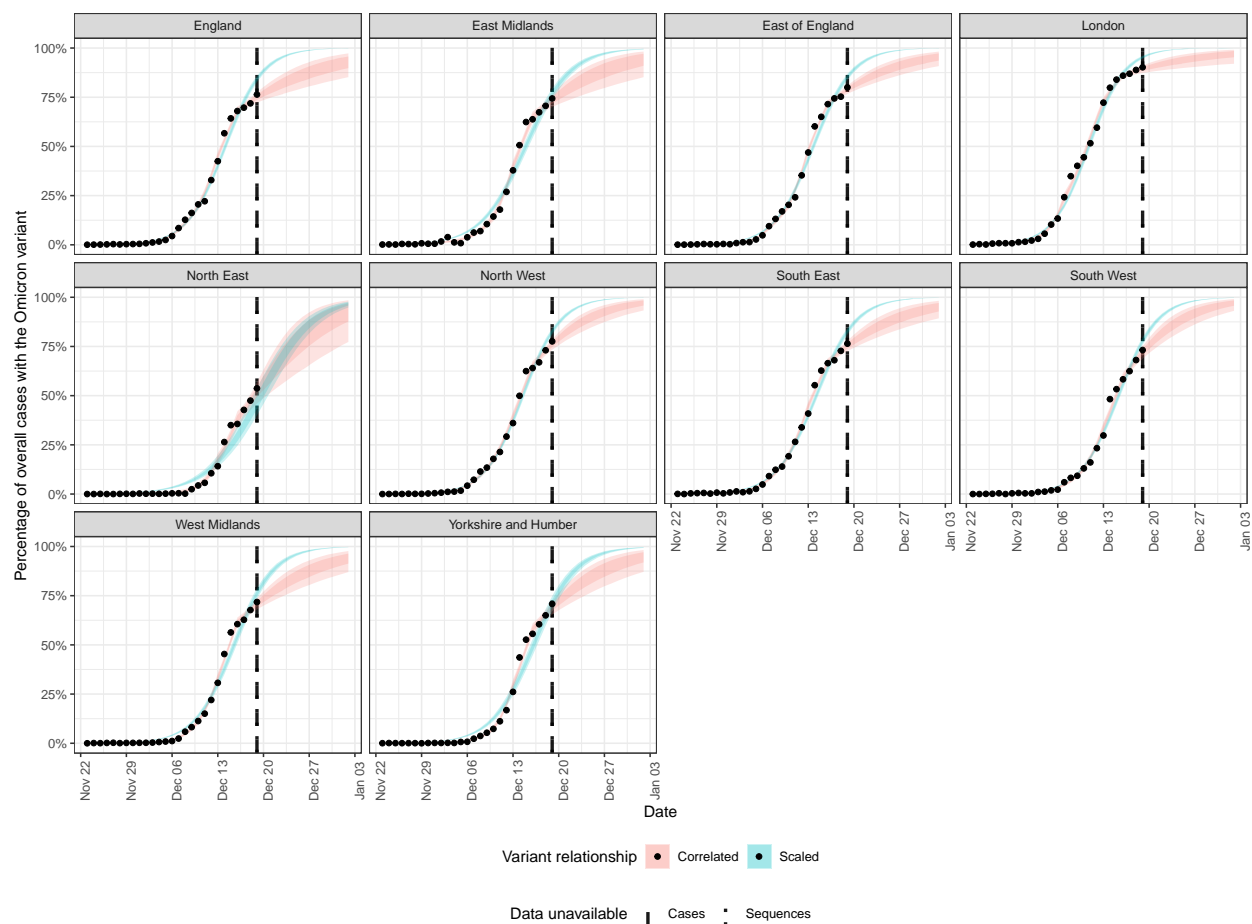
Figure 4: Estimates of the proportion of reported cases with the Omicron variant for England and by UKHSA region using both the scaled model and the correlated model on the natural scale. The forecast horizon is indicated using a dashed line. 90% and 60% credible intervals are shown. Observation error is not shown meaning that the posterior predictions may look overly precise.
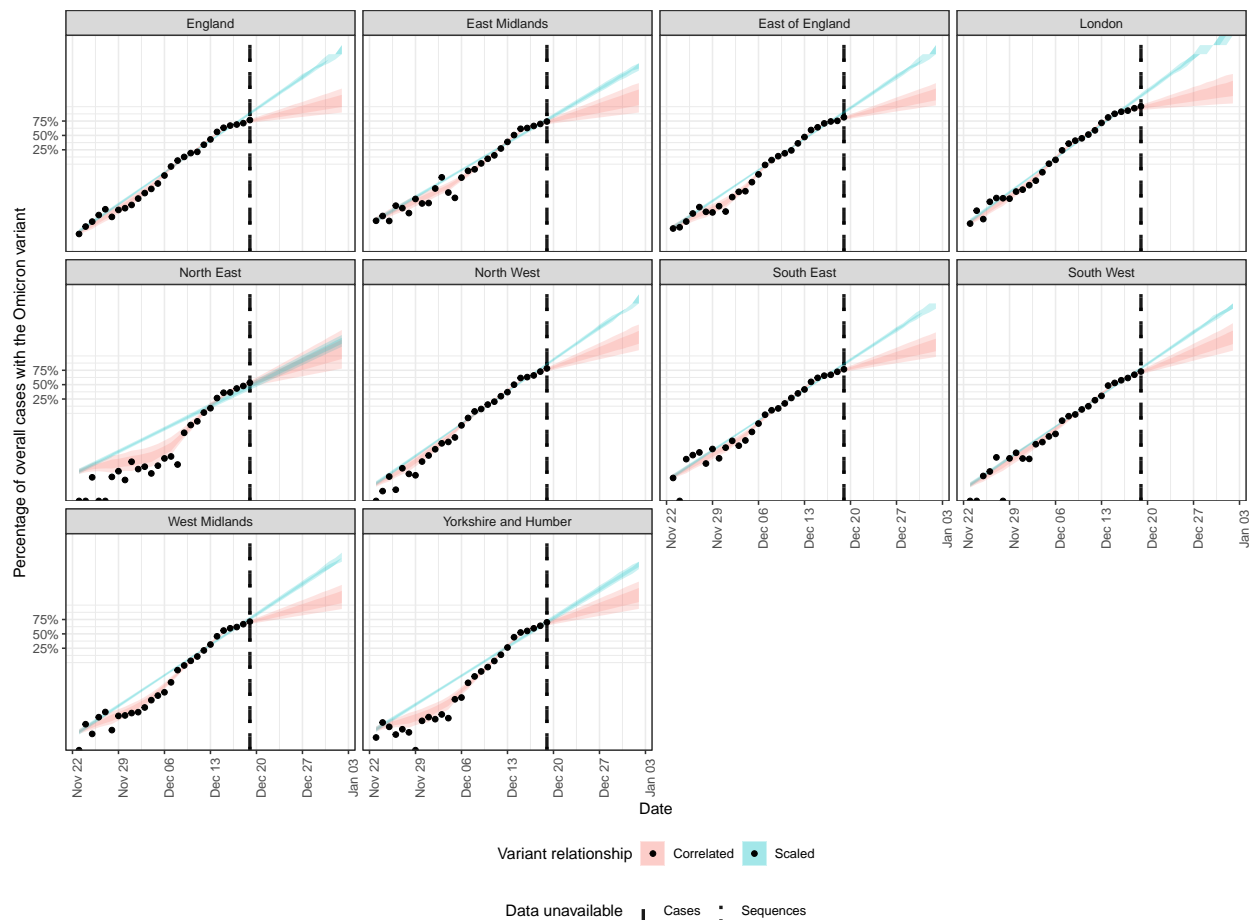
Figure 5: Estimates of the proportion of reported cases with the Omicron variant for England and by UKHSA region using both the scaled model and the correlated model on the logit scale. Note that scaled model is effectively assuming a linear relationship on this scale. The forecast horizon is indicated using a dashed line. 90% and 60% credible intervals are shown. Observation error is not shown meaning that the posterior predictions may look overly precise.
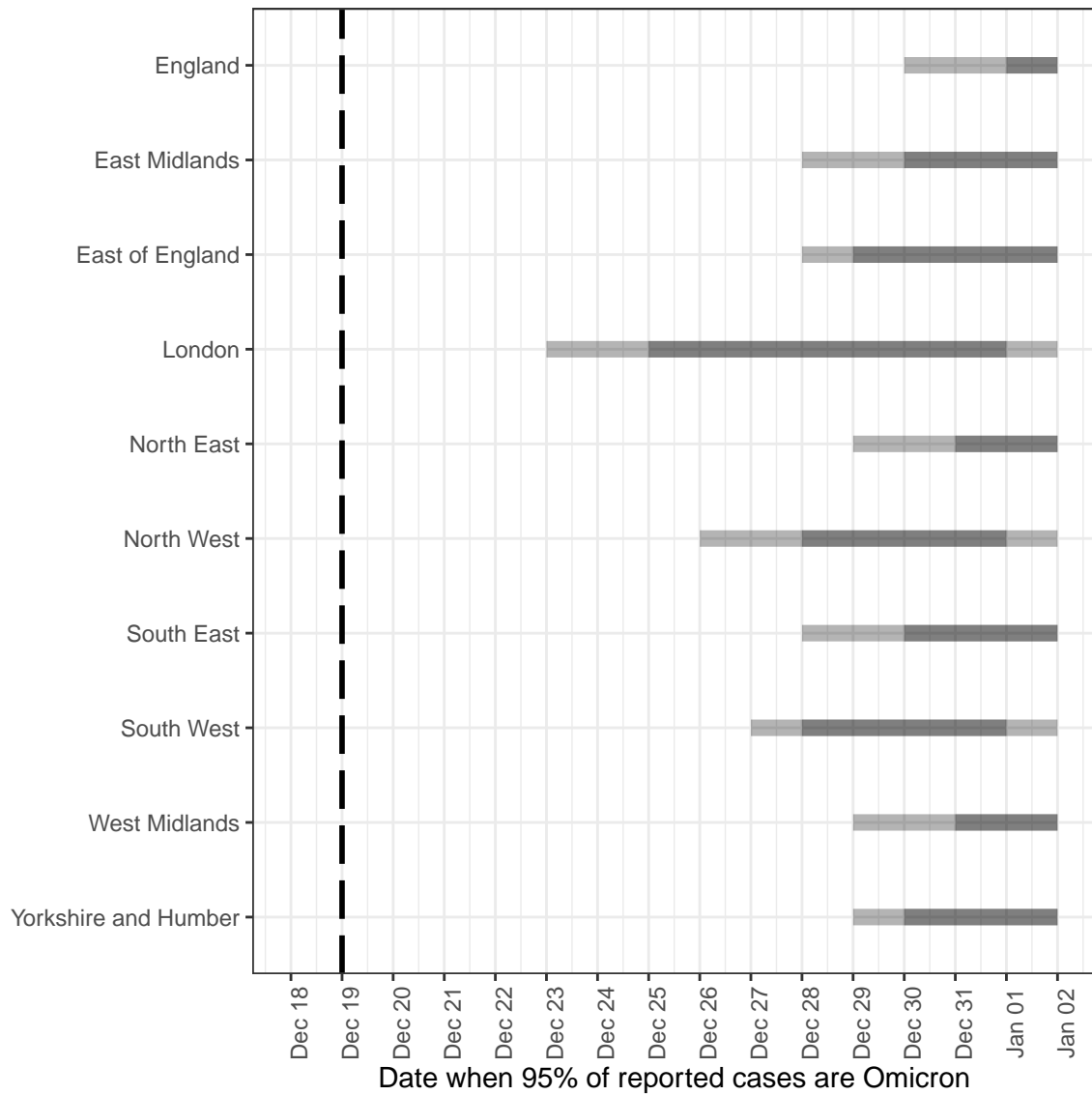
Figure 6: Estimated dates at which 95% of reported cases will have the Omicron variant for England and UKHSA regions. 90% and 60% credible intervals are shown,
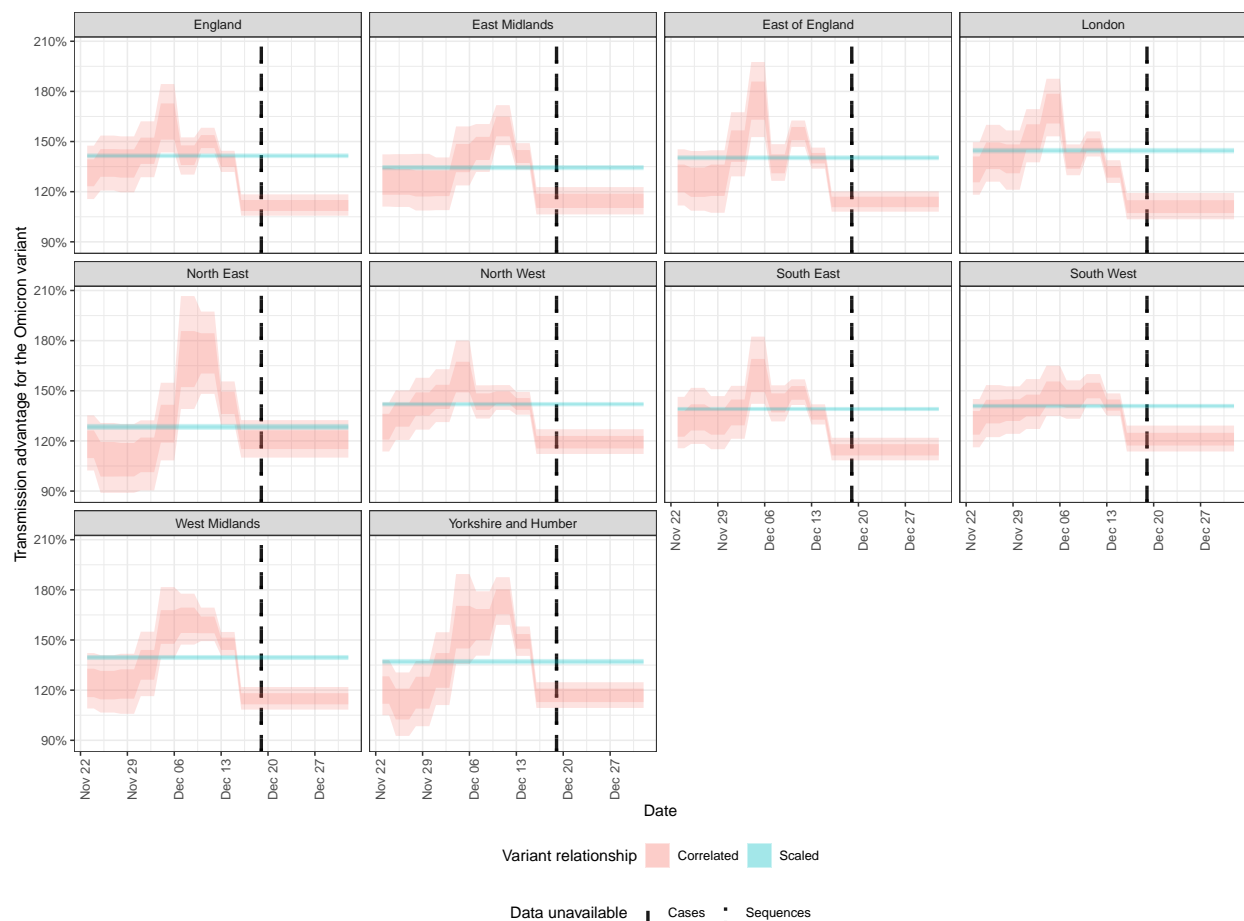
Figure 7: Estimates of the time-varying daily transmission advantage for the Omicron cases vs non-Omicron cases for England and by UKHSA region. Estimates are shown for both the scaled model and the correlated model. Variation in the advantage over time could be an indicator of biases in the SGTF data, differences in the populations in which the variants are circulating or as a result of immune escape. Transmission advantage is assumed to be constant beyond the forecasting horizon (dashed line). 90% and 60% credible intervals are shown. Advantage is calculated as the exponential of Omicron's growth rate minus non-Omicron's growth rate.
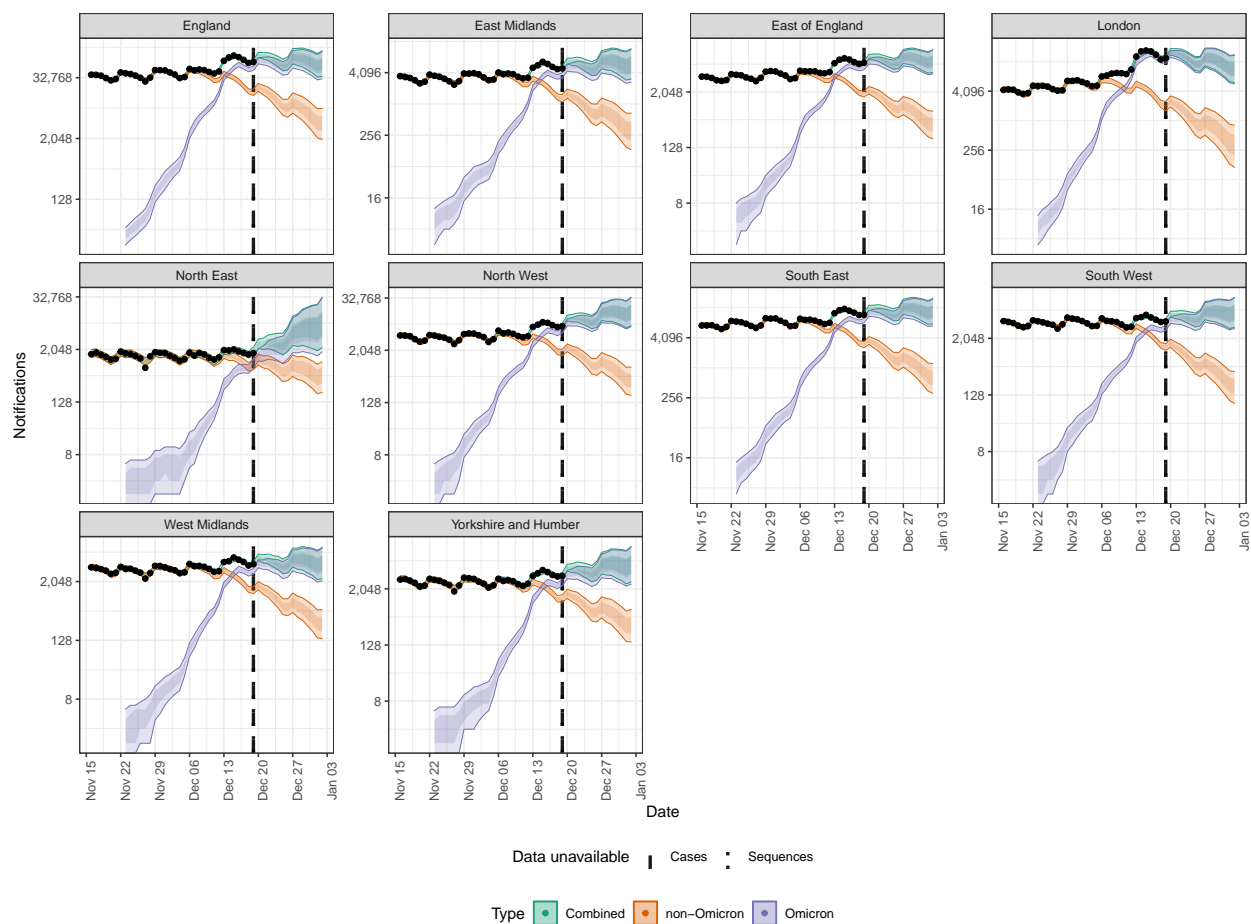
Figure 8: Estimates of the number of reported cases by variant and overall for England and each UKHSA region. Estimates are shown both within the window supported by the data and for 14 days afterwards as a forecast where it is assumed that growth rates for each variant are stable. The forecast horizon is indicated using a dashed line. 90% and 60% credible intervals are shown.

Figure 9: Estimates of the number of reported cases by variant and overall for England and each UKHSA region on the log 2 scale. Estimates are shown both within the window supported by the data and for 14 days afterwards as a forecast where it is assumed that growth rates for each variant are stable. The forecast horizon is indicated using a dashed line. 90% and 60% credible intervals are shown.
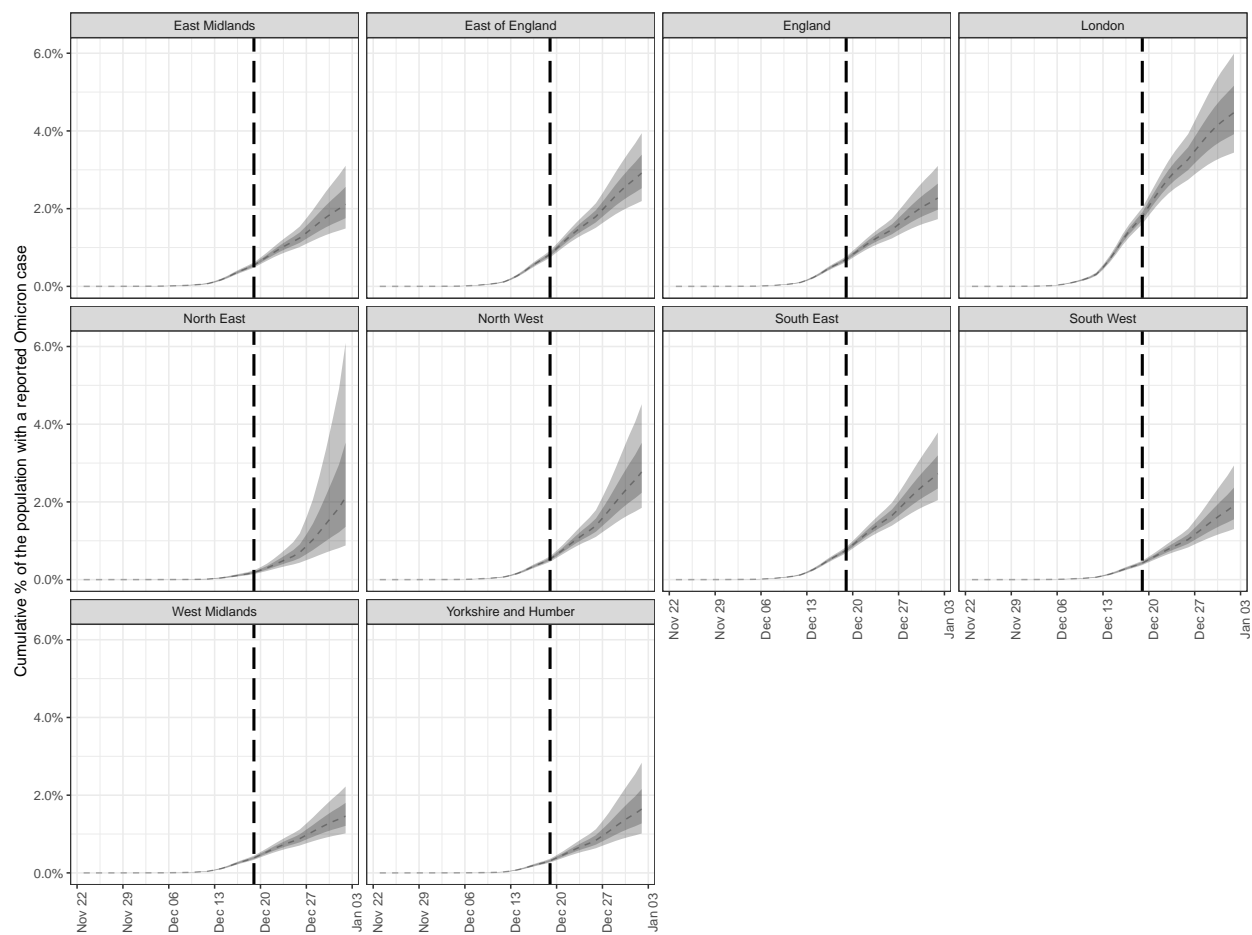
Figure 10: Estimates of the cumulative percentage of the population with reported Omicron cases in England and by UKHSA. The forecast horizon is indicated using a dashed line. 90% and 60% credible intervals are shown.
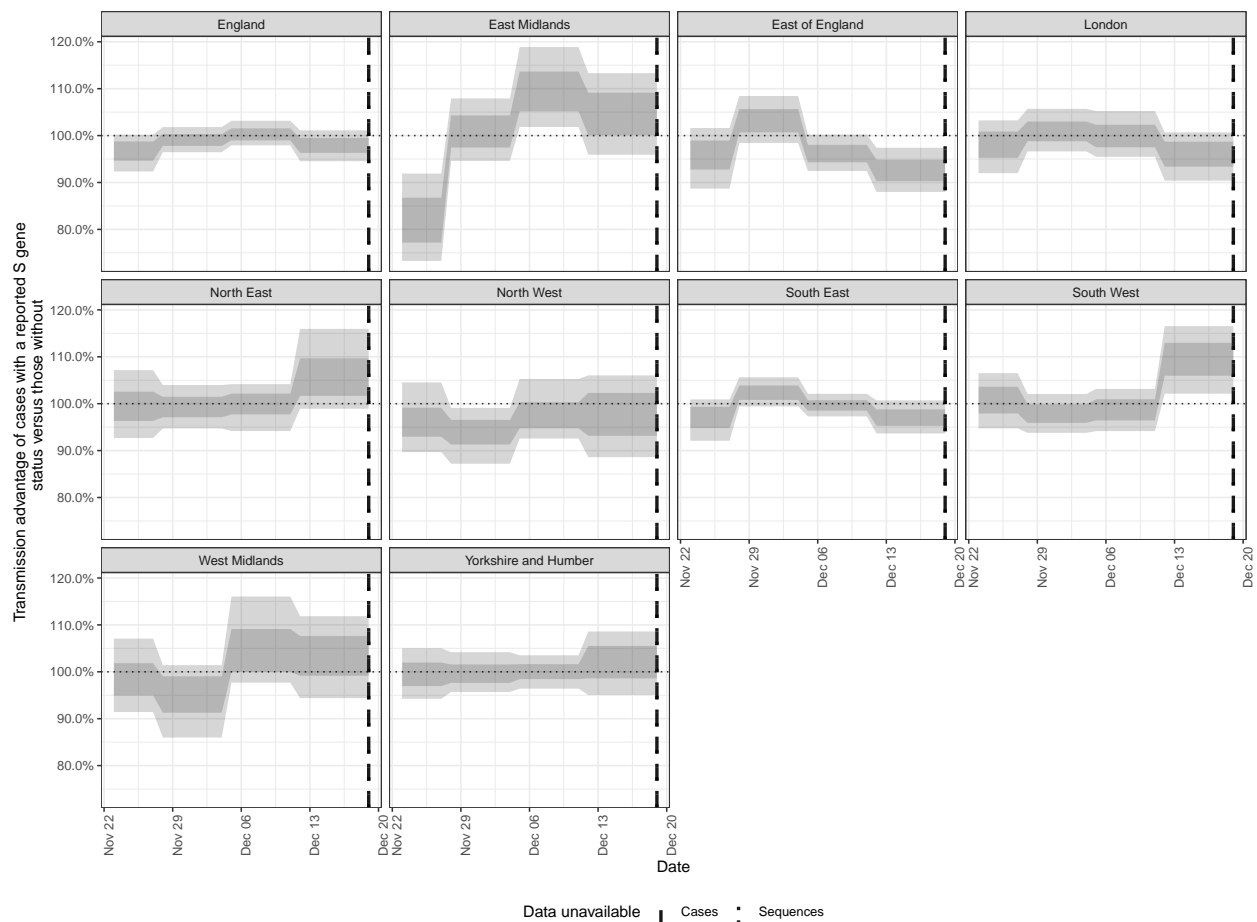
Figure 11: Estimates of the time-varying daily transmission advantage for cases with a reported S-gene status vs those without for England and by NHS region. Deviation from 100% (dashed horizontal line) indicates that growth rates differ in the two populations which is an indication of sampling bias. If present this sampling bias could impact our main results. Transmission advantage is assumed to be constant beyond the forecasting horizon (dashed line vertical line). 90% and 60% credible intervals are shown. Note that here a 7-day piecewise constant growth rate has been used.

3. R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

4. Team, S. D. (2021). *Stan modeling language users guide and reference manual, 2.28.1.*

5. Gabry, J., & Češnovar, R. (2021). *Cmdstanr: R interface to 'CmdStan'.*

6. Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2020). *Loo: Efficient leave-one-out cross-validation and WAIC for bayesian models.* https://mc-stan.org/loo/

7. Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4