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

Epiforecasts

16 December 2021

Background

When processing a positive test result for Covid-19 it is also possible to detect the S-gene target (SGT). A test for SGT can give either a failure or a positive result. The proportion of cases that are tested for SGT and result in SGT-Failure is a useful proxy for the proportion of the Omicron variant in comparison to the Delta variant.

Only a sample of positive Covid-19 test results are also tested for SGT. The ability to test for SGT depends on type of Covid-19 test (only PCR), and laboratory resources. If these constraints are associated with factors influencing transmission, this would bias estimates of transmission made only from SGT data and confound the relationship between SGT results and all test-positive cases.

We aim to explore bias in the availability of S-gene target results among cases testing positive for Covid-19. Here we use a branching process model to identify any difference in growth rate between cases which have an SGT result and those which do not.

Methods

We used a two-strain branching process model to compare all test-positive cases to test-positive cases with any SGT result. We used data for all England test-positive cases. We used only the most recent three weeks of data between 2021-11-01 and 2021-12-12.

We estimated a daily growth rate and allowed weekly piecewise constant variation. We used a uniform prior for the initial growth rate advantage of SGT-result cases compared to no-SGT-result cases. We estimated the transmission advantage as the difference in growth rate between the two sets of cases on a daily scale. We used the `forecast.vocs` package[1] and fit the model using MCMC with 4000 samples.

We fit to data that contained a day-of-week reporting artefact. To test the imapct of this, we repeated the analysis using a 7-day moving average of total cases and those with SGT results. We compared results for transmission advantage and scored each model fit.

Results

On average, 40% cases had an SGT result. We found a median advantage for SGT-result cases of 1.01 (95% credible interval 0.99 - 1.03), scaled against cases without an SGT result. The growth rate increased over time at a similar rate between SGT-result and no-SGT-result cases (figure **??**).

We compared fitting to raw data against fitting to smoothed daily data (see Appendix). We found no substantial difference in transmission advantage when using smoothed data, with better predictive performance from the model using raw unsmoothed data.

Figure 1: Growth rate for all cases (Combined), those with an S-gene target result (SGT-result) and those without
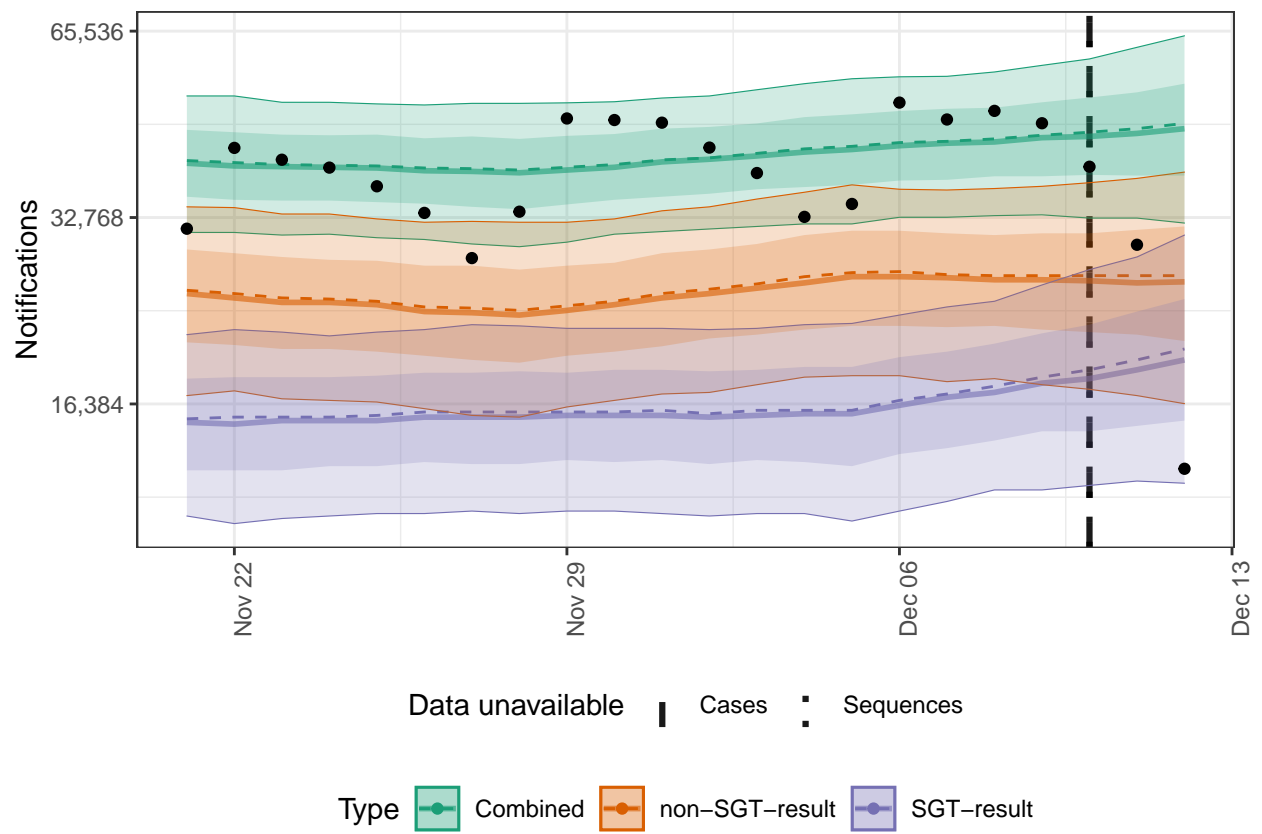
Figure 2: Estimates of all cases (Combined), those with an S-gene target result (SGT-result) and those without

Discussion

We adapted a two-strain branching process model to explore sampling bias in cases with and without an S-gene target result. We assumed that a sampling bias in cases tested for SGT would be associated with transmissibility. This meant we could detect any bias as a transmission advantage for cases with an SGT result compared to cases without an SGT result. For example, labs with capacity for S-gene target detection may be used preferentially to test clusters of test-positive contacts of a known Omicron variant case. This could bias estimates of transmission made only from SGT data and confound the relationship between SGT results and all test-positive cases. This in turn could create a bias in comparisons of Omicron compared to Delta.

However, we found a near-identical estimated transmission advantage with very similar trends in increasing growth rate over time between cases that have been tested for S-gene target and those that have not. Therefore, proportions of SGTF compared to all SGT-result cases are likely to be representative of the wider case population.

This analysis has limitations. Our estimates may be inaccurate where we used a fixed relationship over the three weeks to describe whether cases were given a SGT-result or not. This relationship could change over time depending on sampling strategies: for example, an increase in testing in clusters around Omicron cases, or differential testing capacity by laboratories that can test for the S-gene target. We could further explore this using a model that allows the relationship between the two case groups to vary over time depending on data.

We have not considered sampling biases not associated with transmissibility, or factors beyond the presence or absence of an SGT test. While SGT-failure is used as a proxy for Omicron strain BA.1, a modification of Omicron can also result in SGT-positivity (BA.2). If BA.2 accounts for a significant proportion of Omicron cases, this reduces the overall usefulness of SGTF data for comparing Omicron to Delta as the dominant SGT-positive strain.

To conclude, we find no evidence of a sampling bias between SGT-result cases compared to non-SGT result cases, suggesting that proportions of SGTF compared to all SGT-result cases are likely to be representative of the wider case population.

Appendix

We compared models using both raw data with a day-of-week effect, and smoothed data using a 7-day centred moving average on total cases and SGT-result cases.

Case data

:
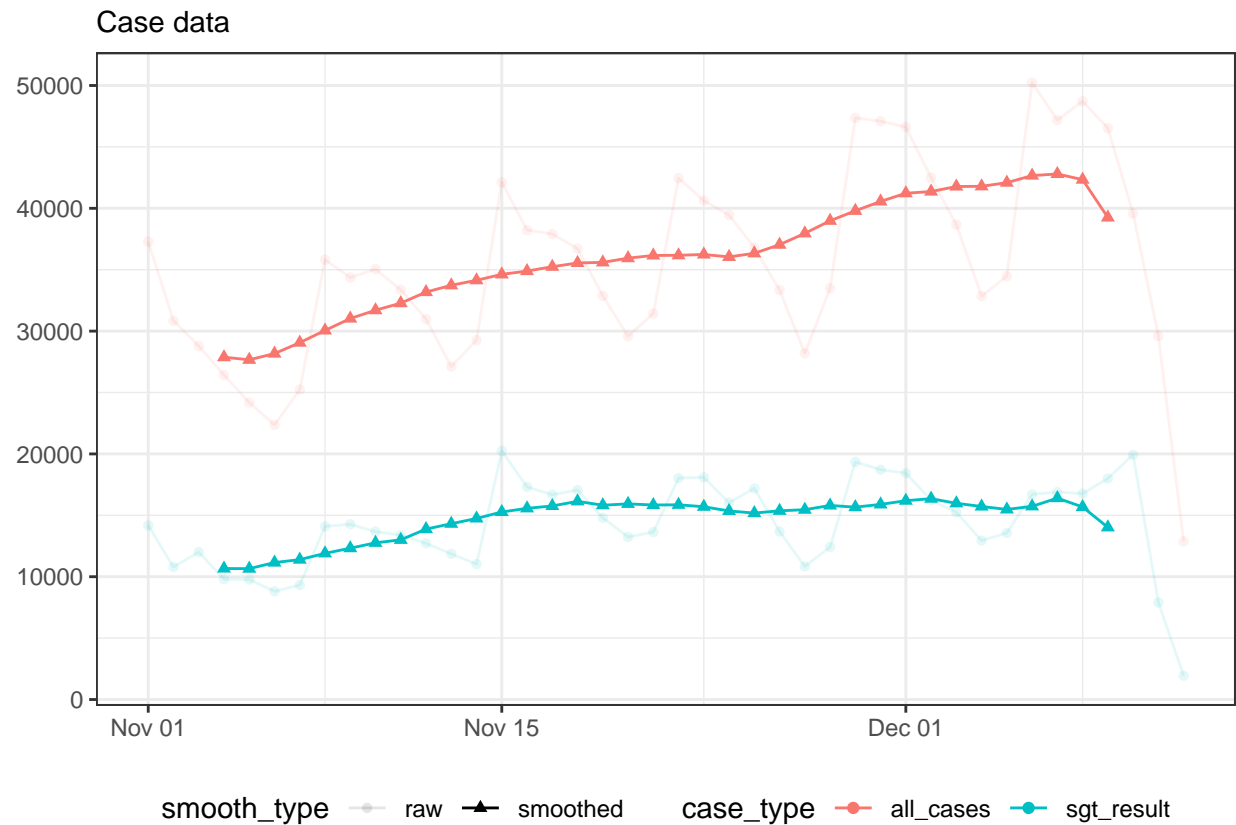
Table 1: Modelled transmission advantage by data pre-processing

data_type	median	q5	q95
raw	1.01	0.99	1.03
smooth	0.98	0.98	0.99

Table 2: Model scores by data pre-processing

data_type	interval_score	sharpness	underprediction	overprediction	coverage_deviation	bias
raw	3540	1640	180	1720	-0.09	0.32
smooth	5620	264	2170	3190	-0.66	-0.05