B.1.617.2 transmission in England: risk factors and transmission advantage

Neil M Ferguson, MRC Centre for Global Infectious Disease Analaysis, Imperial College London

June 1st 2021

Abstract

We analyse trends in B.1.617.2 transmission since April, using genomic and S-gene target failure data for symptomatic pillar 2 cases. Direct estimation of the reproduction numbers for these two variants suggest B.1.617.2 has a 50-100% transmissibilty advantage over B.1.1.7 and a current doubling time of approximately 9 days in England. Logistic regression predicting B.1.617.2 infection among B.1.1.7 and B.1.617.2 cases suggests comparable growth rates over time in all ethnic groups and age classes, albeit some variation in initial seeding levels by ethnicity. Log odd frequency growth rates still show statistically significant differences by region, but these are narrowing. Statistically signficant reductions in 1 and 2 dose vaccine efficacy are seen.

Analysis conducted with SPI-M data dump of 31/5/2021

1. Analysis of genomically confirmed B.1.617.2 frequency and incidence trends

We consider trends in genomically-confirmed B.1.617.2 absolute incidence and frequency among Pillar 2 cases of either B.1.617.2 or B.1.1.7. For the frequency analysis we consider epidemiological weeks 17-20 (2021-04-26 to 2021-05-23). We exclude other variants from this analysis since our focus is to compare those two variants alone. Our analysis is restricted to symptomatic cases and those who did not first receive a lateral flow test to reduce biases caused by asymptomatic surge testing in hotspot areas, though we would note that results are very similar if no filtering by symptoms is applied.

Of the 9808 B.1.617.2 cases detected in that time interval, 9538 were detected in pillar 2, 4953 were symptomatic (of which 925 had received at least one vaccine dose 28 or more days prior to their COVID-19 test date). There were 6499 symptomatic pillar 2 B.1.1.7 cases in the same time interval (of which 1073 had received at least one vaccine dose 28 or more days prior to their COVID-19 test date).

We note that 88% of genomically confirmed B.1.617.2 cases have a definitive positive S-gene signal in pillar 2 testing, while 83% of B.1.1.7 cases test S-gene negative. The vast majority of genomically-confirmed B.1.617.2 and B.1.1.7 cases which do not test S-gene positive or negative, respectively, have an unclassified S-gene result. Fewer than 0.2% of B.1.617.2 and B.1.1.7 cases give a false S-gene negative or positive result, respectively.

We first examine overall incidence trends in B.1.1.7 and B.1.617.2 cases. We smooth daily symptomatic pillar 2 case incidence with a 7 day running average calculated on a log scale to reduce day of week effects. We then estimate B.1.617.2 and B.1.1.7 incidence by multiplying smoothed pillar 2 case incidence by the frequency of the respective variant among all sequences with that specimen date (note that frequency is not smoothed). Figure 1 shows incidence trends for the three weeks to week 19 (2021-05-16), since delays in cause sequencing data to be more limited after that date. Note that very low case incidence in the South West gives high uncertainty in estimates.

We then used EpiEstim (assuming a generation time distribution with mean 5.4 days and standard deviation of 1.5 days) to estimate the effective reproduction number, R, over time (Figure 2). Prior to May 10th,

estimates of R for B.1.617.2, which are highly uncertain, are highly likely to be inflated for by imported infections in travellers. Due to the large proportion of missingness in data on traveller status, we did not attempt to adjust estimates for reported importations. We note that these R estimates imply a doubling time for B.1.617.2 of 12.8 (95% CI:10.8-16) days in England on 2021-05-16.

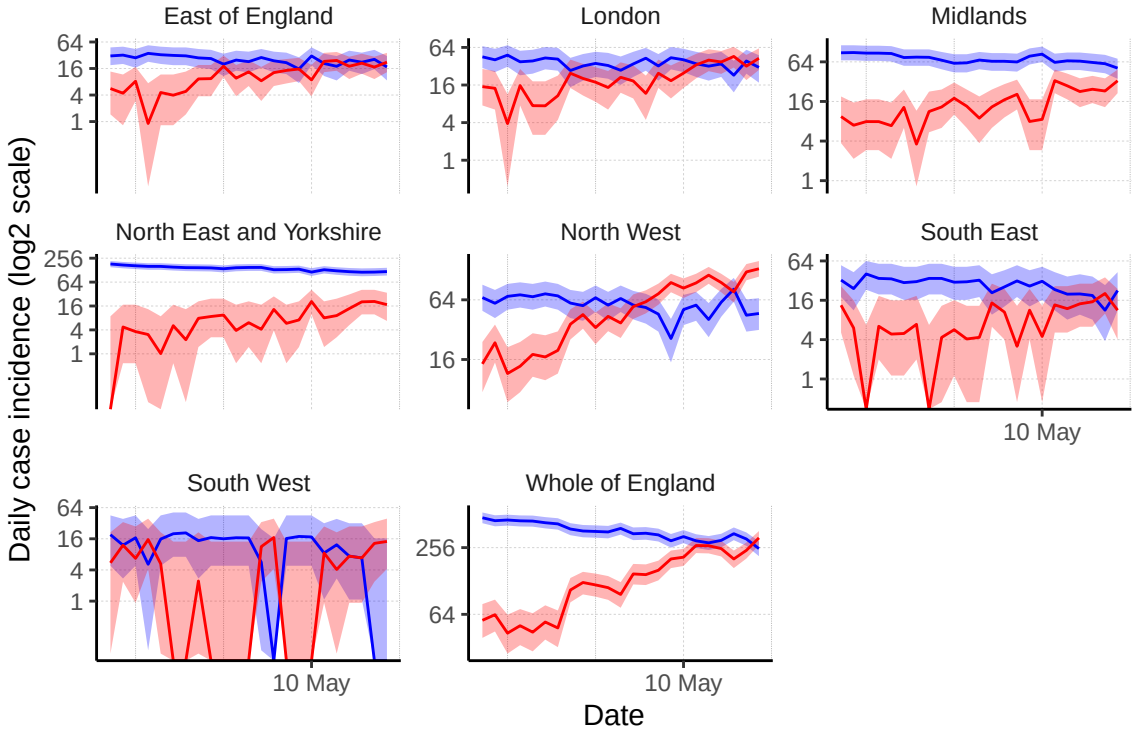

Figure 1: Daily incidence of B.1.617.2 (red) and B.1.1.7 (blue) cases in recent weeks, estimated from genomic data. Ribbons show 95% confidence bounds, which are dominated by numbers of sequences available. Log2 scale used to allow doubling times to be inferred. Date axis uses 1 week grid lines.

Last, we estimated the transmission advantage of B.1.617.2 relative to B.1.1.7 as the ratio of the reproduction number estimates of the two variants (Figure 3). As for the R estimates, these are likely to have been inflated by imported B.1.617.2 infections in earlier weeks, but given the travel ban from India on 23rd April, are expected to be more reliable from 10th May onwards.

We note that estimates are, allowing for data limitations, broadly consistent between NHS regions, albeit the larger case numbers in the Nort West dominate the national estimates. Overall, the results are consistent with a transmission advantage of B.1.617.2 over B.1.1.7 in the 50-100% range.

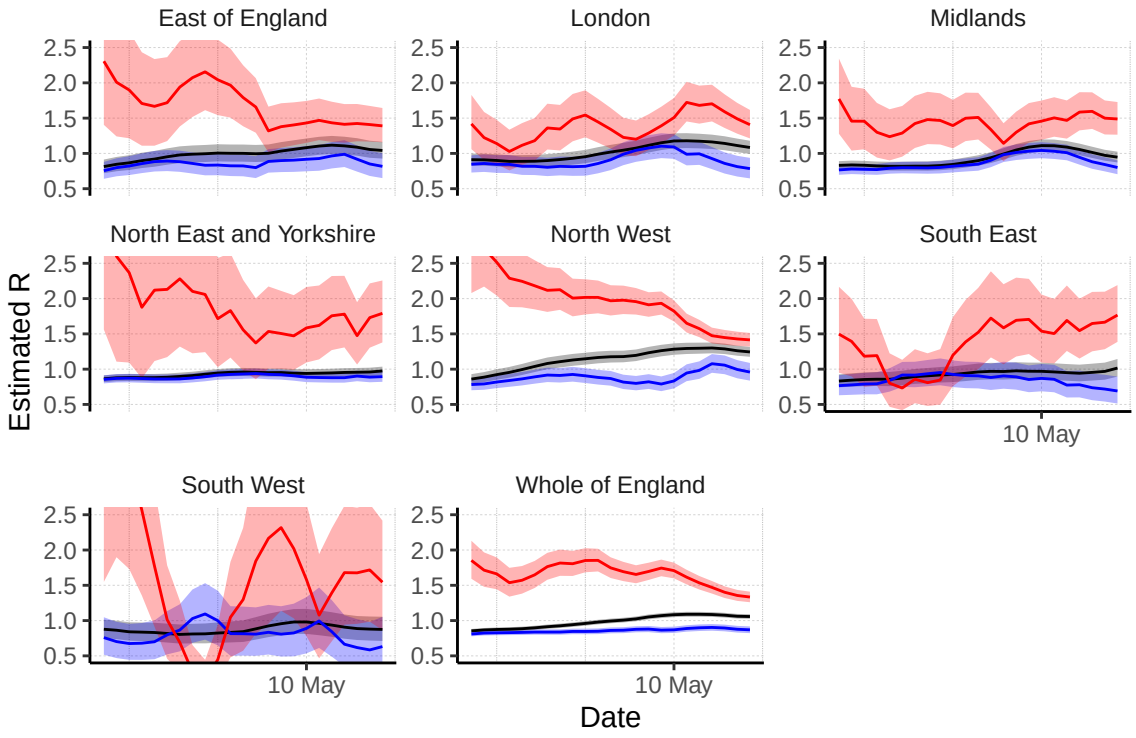

Figure 2: Estimated daily instantaneous reproduction number of all symptomatic pillar 2 cases (black), B.1.617.2 cases (red) and B.1.1.7 cases (blue) in recent weeks, estimated from genomic data. Ribbons show 95% confidence bounds. The analysis included two weeks of data prior to that shown on Figure 1, to allow estimates to stabilise. Date axis uses 1 week grid lines.

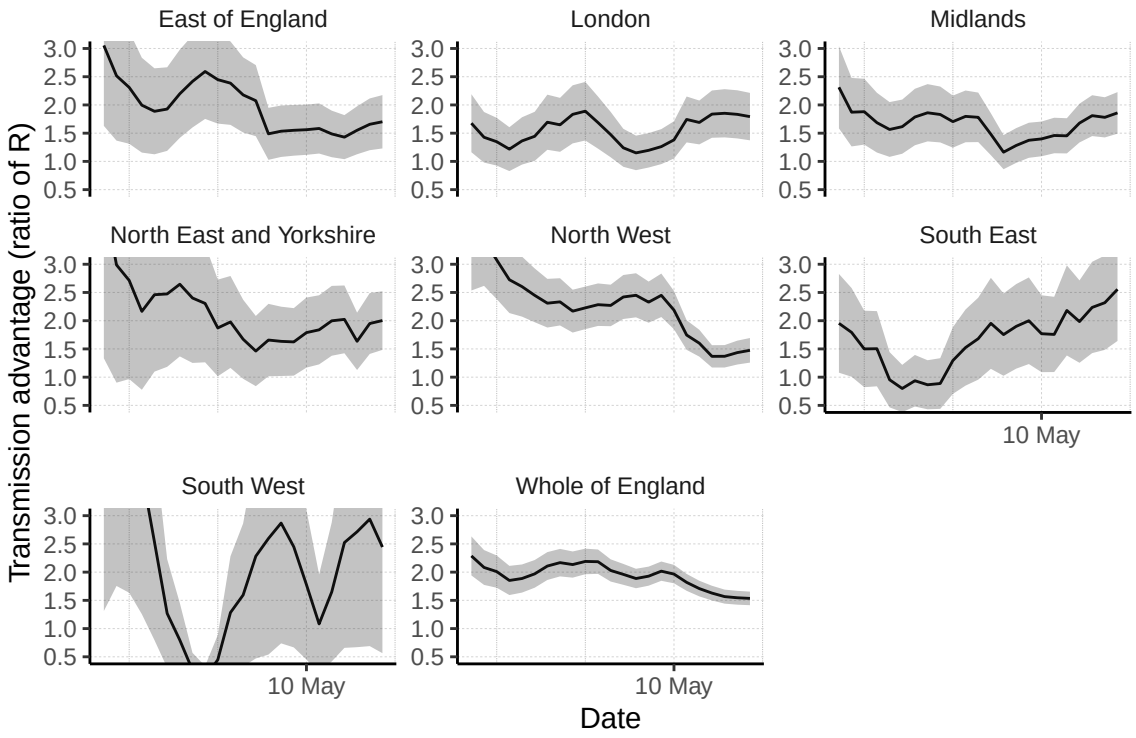

Figure 3: Estimated transmission advantage of B.1.617.2 relative to B.1.1.7 in recent weeks, estimated as the ratio of reproduction numbers shown in previous figure. Ribbons show 95% confidence bounds estimated using the delta method and assuming no correlation between B.1.617.2 and B.1.1.7 reproduction number estimates. Date axis uses 1 week grid lines.

We then used logistic regression to model the probability that each B.1.1.7 or B.1.617.2 case was B.1.617.2. The following categorical predictors were tested:

- Area - NHS region, NHS STP area and LTLA were explored. LTLA minimised model AIC, but STP gave the optimal out-of-sample AUC.
- Ethnicity group, grouped into five classes, based on previous analyses.
- Age class - five age classes defined: 0-20, 21-35, 36-50, 51-69, 70+. This updated analysis retains cases of all ages, making this variable signficant.
- Traveller status - case labelled as a traveller. This was not significant for the analysis of genomically confirmed cases, but was retained as it was significant when the regression was repeated for S-gene positivity.
- Vaccination status - we stratify by vaccine type (AstraZeneca or Pfizer, where all non-AZ vaccines are classed as Pfizer). Number of doses received was defined as having had a first dose of vaccination at least 28 days before the specimen date or a second dose at least 14 days before the specimen date. Model AIC still slightly favours only including number of doses, rather than vaccine type, but results for the more stratified model are compatible with the simpler model and give insight into potential differences between vaccines.
- Travel ban - an indicator variable defined to be 1 if the specimen date occured on or before 23rd April, and 0 otherwise. Inclusion of this variable did not significantly improve model fit.

In addition, we used specimen date as a predictor, but transformed this onto a "generation time" scale by dividing days since April 19th by 5.4. This allows the resulting odds ratios to be approximately interpreted as a measure of transmissibility advantage on a reproduction number scale. Model AIC favoured allowing the specimen date coefficients (i.e. transmission advantage) to vary by NHS region, but there was insufficient data to use a finer spatial scale (except for the hotspot local authority areas).

Figure 4 shows trends in B.1.617.2 frequency in recent weeks and model fit to the data. Note that the while model was fitted to individual data, aggregated results are shown in the figure.

Table 1 shows model estimates (area-specific intercepts not listed). Indian and unknown ethnicity were significantly associated with a substantially higher odds of being a B.1.617.2 case (odds ratio>15). Pakistani, "Any other Asian background" and "Unknown" ethnicity groups also had odds ratios significantly over 1.

Having received one vaccine dose 28 or more days before the specimen date was also associated with being a B.1.617.2 case, with an odds ratio of 1.3 (95% CI: 1.1-1.54) for AstraZeneca and 1.03 (95% CI: 0.75-1.42) for Pfizer. Having received a second dose at least 14 days before the specimen date was also associated with being a B.1.617.2 case, with an odds ratio of 1.36 (95% CI: 0.88-2.12) for AstraZeneca and 2.4 (95% CI: 1.35-4.39) for Pfizer. Note that the higher odds ratio for the second dose of Pfizer still only implies a limited drop in post-dose two efficacy. Assuming two-dose vaccine efficacy for B.1.1.7 is 95% for Pfizer, it would be 100 - (100 - 95) * 2.4 = 88% against B.1.617.2. Conversely, assuming two-dose vaccine efficacy for B.1.1.7 is 66% for AstraZeneca, our estimates imply efficacy of 100 - (100-two_dose_eff_AZ) * 1.36 = 53.76% against B.1.617.2. Assuming 60% one dose efficacy against B.1.1.7 for both Pfizer and AstraZeneca vaccines, our estimates imply, respectively, 58.8% and 48% efficacy against B.1.617.2. It should be noted that if B.1.617.2 is still typically circulating in communities with lower vaccine coverage than those in which B.1.1.7 is circulating, these odds ratios may be underestimated.

Odds ratio estimates for age classes in 21+ year-olds are now generally close to 1, albeit still significantly less than 1 for 21-25 year olds. Being labelled as a traveller in the dataset was associated with a very high central odds ratio estimate of being infected with B.1.617.2 (not shown in Table), albeit this was not statistically significant.

Estimates of B.1.617.2 transmission advantage were consistent with those estimated from direct estimation of R (Figure 2) and varied significantly by NHS region, with central estimates ranging from 1.7 to 2.8, with 2.0 being the central estimate if no regional variation in advantage was fitted.

Figure 5 demonstrates that growth rates of B.1.617.2 vary little by age band or ethnicity group, though seeding was initially higher in people of Indian and other non-White ethnicity. The very similar slopes of the

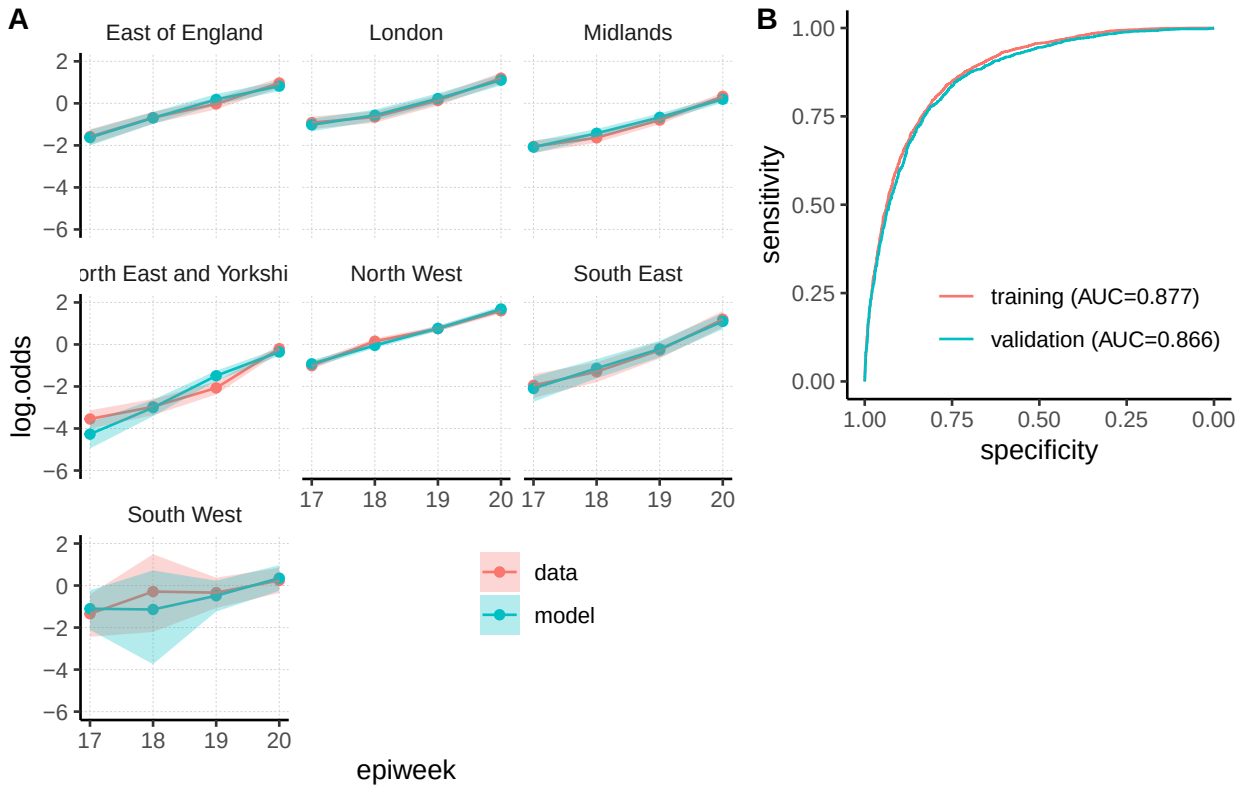

Figure 4: A. Log-odds of B.1.617.2 frequency in symptomatic pillar 2 cases. Data (red) and fit of logistic regression model shown. B. ROC curve for logistic regression for a 70% training set and 30% validation set.

Table 1: Estimates from logistic regression model fitted to B.1.617.2 and B.1.1.7 symptomatic cases. Odds ratios for ethnicity, vaccination status and NHS region-specific transmission advantage shown. NHS STP area-specific intercepts not listed. Traveller status related odds ratio also not listed (»1 but non significant).

Characteristic	OR	95% CI	p-value
Ethnicity group			
white	—	—	
all_other	1.62	1.39, 1.89	<0.001
other_asian	2.58	2.26, 2.93	<0.001
indian	8.71	7.11, 10.7	<0.001
unknown	2.32	1.60, 3.39	<0.001
Number and type of vaccine doses received			
None	—	—	
AZ1	1.30	1.10, 1.54	0.002
AZ2	1.36	0.88, 2.12	0.2
PF1	1.03	0.75, 1.42	0.8
PF2	2.40	1.35, 4.39	0.004
Age band			
0-20	—	—	
21-35	0.82	0.73, 0.93	0.002
36-50	0.87	0.76, 1.00	0.045
51-69	1.11	0.90, 1.37	0.3
70+	0.90	0.52, 1.58	0.7
Region specific transmission advantage			
Tg * East of England	1.73	1.54, 1.95	<0.001
Tg * London	1.90	1.73, 2.11	<0.001
Tg * Midlands	1.85	1.70, 2.01	<0.001
Tg * North East and Yorkshire	2.83	2.51, 3.23	<0.001
Tg * North West	1.98	1.88, 2.09	<0.001
Tg * South East	2.52	2.13, 3.03	<0.001
Tg * South West	1.82	1.27, 2.77	0.002

OR = Odds Ratio, CI = Confidence Interval

log odds frequency plots by age and ethnicity in the last 2 weeks suggests the infection is now well-mixed in the general population, albeit at a higher frequency in Indian and other Asian ethnic groups due to that seeding-driven founder effect. This gives greater confidence in transmission advantage estimates. However, given there are still statistically significant differences between growth advantage estimates for different regions, so it remains difficult to derive a definitive overall estimate. Values below 1.5 (a 50% advantage) are now very unlikely though.

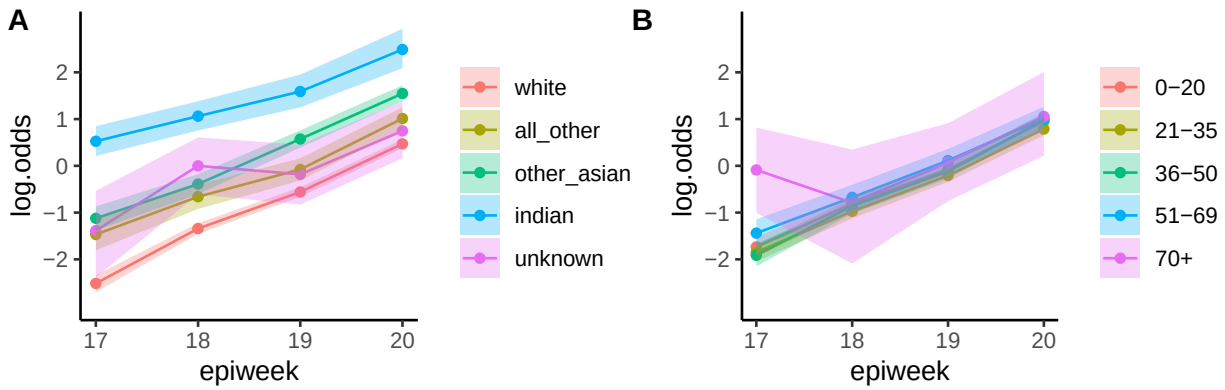

Figure 5: Log-odds of B.1.617.2 frequency in symptomatic pillar 2 cases stratified by A. ethnicity group; B. by age band.

2. Analysis of positive S-gene frequency

We repeated the above analyses above but modelling B.1.617.2 frequency with the proxy of testing definitely S-gene positive or negative in TaqPath pillar 2 testing. We were able to include an additional week of data in this analysis, given S-gene results are available sooner than sequencing data. Hence we analysed frequency of S-gene positivity for epidemiological weeks 17-21 (2021-04-26 to 2021-05-30). A total of 9202 S-gene positive cases were included (of which 1741 had received at least one vaccine dose 28 or more days prior to their COVID-19 test date), and 7394 S-gene negative cases (of which 1239 had received at least one vaccine dose 28 or more days prior to their COVID-19 test date).

Figures 6, 7 and 8 show incidence timeseries, R estimates and transmission advantage estimates, respectively, using SGTF data. We also plot the emergence of B.1.1.7 between November 2020 and January 2021 for comparison. Estimates of both R and transmission advantage from SGTF data were very similar to those using genetic data, with current estimates of the transmission advantage being close to 100% for England. The R estimates imply a current doubling time for B.1.617.2 of 9.2 (95% CI:8.4-10.1) days in England.

Logistic regression modelling of S-gene positive infection (among all pillar 2 symptomatic cases) gave very similar results (Figure 9 and Table 2) to those from the analysis of genomically-confirmed cases, though given the stratification of vaccination status by both dose and vaccine type, only the Pfizer post-dose 2 odd-ratio was significant. Unlike the genomic dataset, traveller status was a significant predictor ($p<0.001$) for the S-gene positivity analysis.

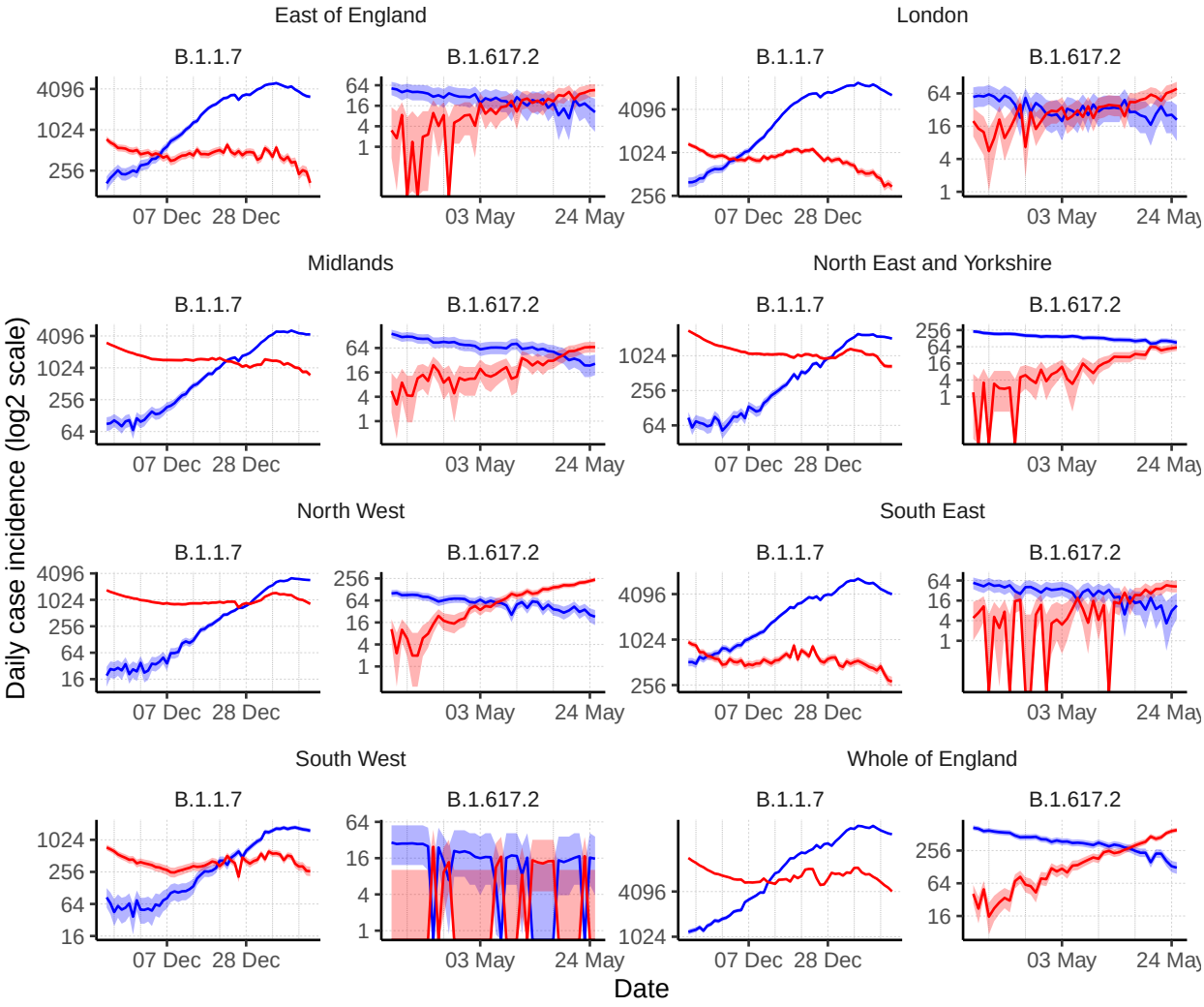

Figure 6: Daily incidence of S-gene positive (red) and S-gene negative (blue) cases in two different time periods (November 2020-January 2021 and April-May 2021). Ribbons show 95% confidence bounds, which are dominated by numbers of SGTF results available (notably low in the South West). Log2 scale used to allow doubling times to be inferred.

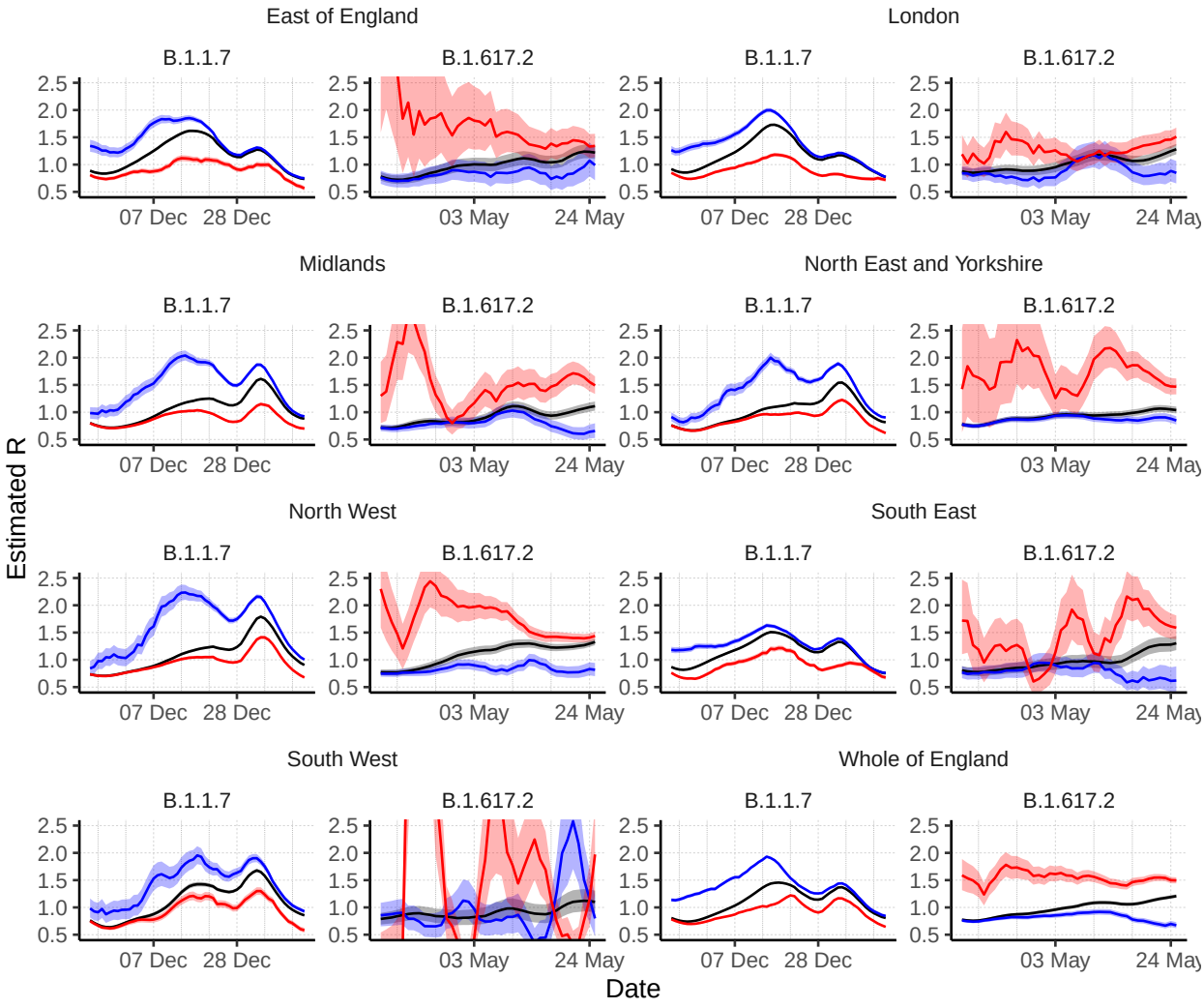

Figure 7: Estimated daily instantaneous reproduction number of all symptomatic pillar 2 cases (black), S-gene positive cases (red) and S-gene negative cases (blue) for the time periods covering B.1.1.7 emergence (left panels) and B.1.617.2 emergence (right panels). Ribbons show 95% confidence bounds. The analyses included two weeks of data prior to that shown on the previous Figure, to allow estimates to stabilise.

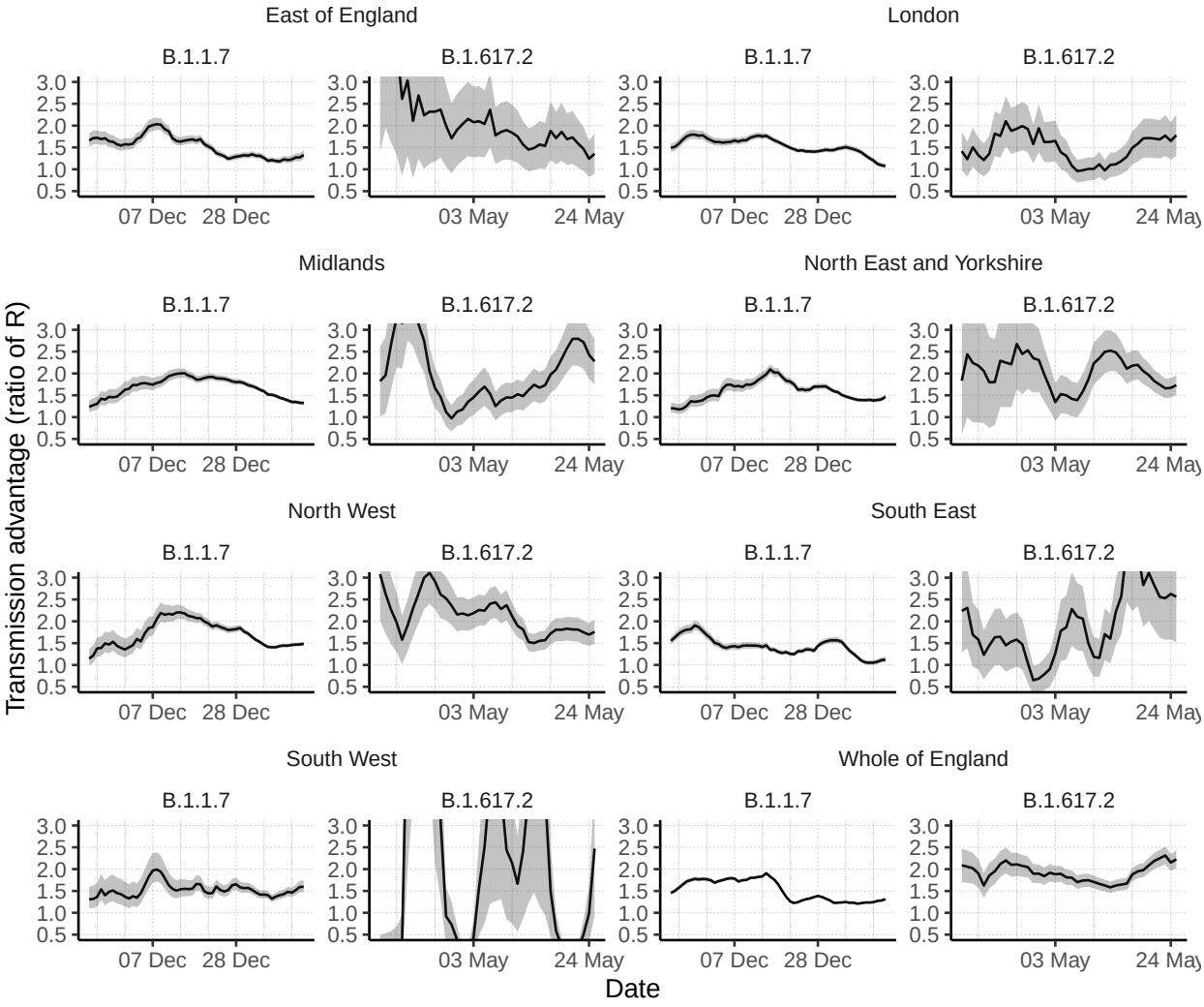

Figure 8: Estimated transmission advantage of B.1.1.7 relative to prior variants between November 2020 and January 2021 (left panels) and B.1.617.2 relative to B.1.1.7 in recent weeks (right panels) derived from SGTF data. Transmission advantage was estimated as the ratio of reproduction numbers shown in previous figure (S-/S+ for B.1.1.7 and S+/S- for B.1.617.2). Ribbons show 95% confidence bounds estimated using the delta method and assuming no correlation between B.1.617.2 and B.1.1.7 reproduction number estimates.

Table 2: Estimates from logistic regression model fitted S-gene positive and negative symptomatic cases. Odds ratios for ethnicity, vaccination status and NHS region-specific transmission advantage shown. NHS STP area-specific intercepts not listed.

Characteristic	OR	95% CI	p-value
Ethnicity group			
white	—	—	
all_other	1.58	1.38, 1.82	<0.001
other_asian	2.36	2.10, 2.65	<0.001
indian	7.35	6.11, 8.88	<0.001
unknown	2.48	1.81, 3.44	<0.001
Number and type of vaccine doses received			
None	—	—	
AZ1	1.17	1.01, 1.35	0.031
AZ2	1.37	0.97, 1.94	0.076
PF1	1.14	0.85, 1.53	0.4
PF2	2.24	1.40, 3.64	<0.001
Case labelled as traveller			
0	—	—	
1	13.6	4.18, 63.4	<0.001
Age band			
0-20	—	—	
21-35	0.88	0.80, 0.98	0.018
36-50	0.85	0.76, 0.96	0.007
51-69	1.07	0.90, 1.28	0.4
70+	0.97	0.58, 1.63	0.9
Region specific transmission advantage			
Tg * East of England	1.78	1.63, 1.95	<0.001
Tg * London	1.56	1.43, 1.70	<0.001
Tg * Midlands	1.98	1.87, 2.10	<0.001
Tg * North East and Yorkshire	2.02	1.88, 2.18	<0.001
Tg * North West	2.00	1.93, 2.08	<0.001
Tg * South East	2.64	2.31, 3.04	<0.001
Tg * South West	1.66	1.13, 2.64	0.016

OR = Odds Ratio, CI = Confidence Interval

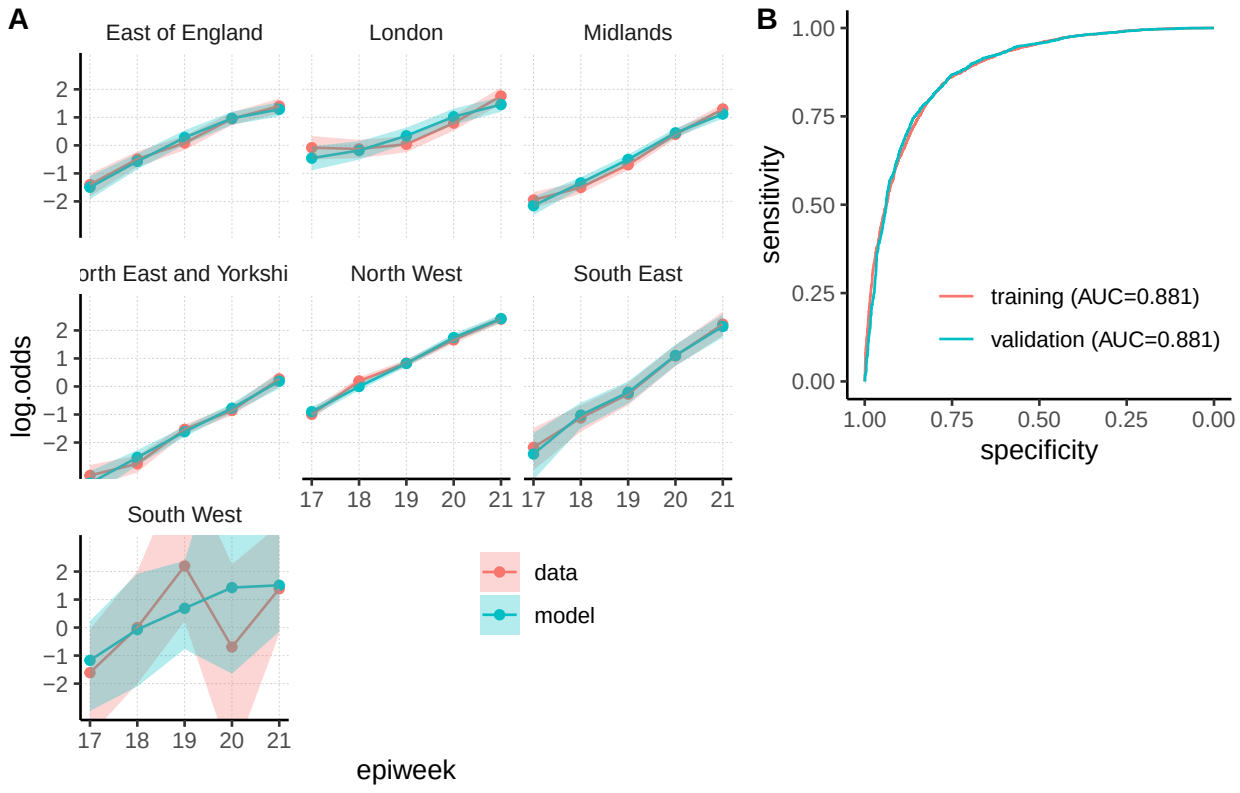

Figure 9: A. Log-odds of S-gene positive frequency in symptomatic pillar 2 cases. Data (red) and fit of logistic regression model shown. B. ROC curve for logistic regression for a 70% training set and 30% validation set.

Conclusions

This analysis provides more convincing evidence that the transmissibility of B.1.617.2 is at least 50% greater than that of B.1.1.7. In addition, it supports the results of matched cohort studies of the impact of B.1.617.2 on vaccine efficacy. While the variant has a head start in Indian and some other ethnicity groups, growth rates of log-odds frequency compared with B.1.1.7 are similar in all ethnic groups and age classes.

Comparing B.1617.2 emergence with that of B.1.1.7 is informative. Firstly, incidence levels are still over 10-fold lower than they were in November 2020 (6). Secondly, while R estimates for S-postive and S-negative mirrored each other during B.1.1.7's emergence, a similar relationship is less clear during the recent emergence of B.1.617.2 (Figure 7). This is likely to be caused by a number of factors: (a) lower case incidence, giving greater uncertainty in estimating R, (b) the effects of imported B.1.617.2 cases biasing R estimates until the last two weeks, and (c) circulation of B.1.617.2 still being more focussed in different communities than that of B.1.1.7. We would also note that the transmission advantage B.1.1.7 exhibited over prior lineages following its emergence varied over time (Figure 8), likely as a result of the intensification of social distancing and reimposition of lockdown in January 2021. Thus while the evidence is now strong for B.1.617.2 having at a minimum of a 50% transmission advantage over B.1.1.7, it perhaps remains too early to derive more precise estimates.

However, irrespective of the precise transmission advantage of B.1.617.2 over B.1.1.7, the fact that B.1.617.2 currently has a reproduction number of 1.5 (assuming a mean serial interval of 5.4 days - higher in a longer serial interval is assumed) and that case incidence is currently doubling approximately every 9 days is of considerable concern. Indeed, this doubling time might be expected to reduce further as the effects of the May 17th Step 3 relaxation feed through into case numbers.

Our findings about vaccine efficacy support matched cohort analyses and suggest a moderate but significant drop in efficacy for both Pfizer and AstraZeneca vaccines after both one and two doses. We would also note that the odds ratio estimates for vaccination do not substantially change if one examines a different time interval between vaccination and specimen date (e.g. 21 days from second dose).