Analysis of transmissibility based on genomics

PHE, Imperial College London, University of Edinburgh

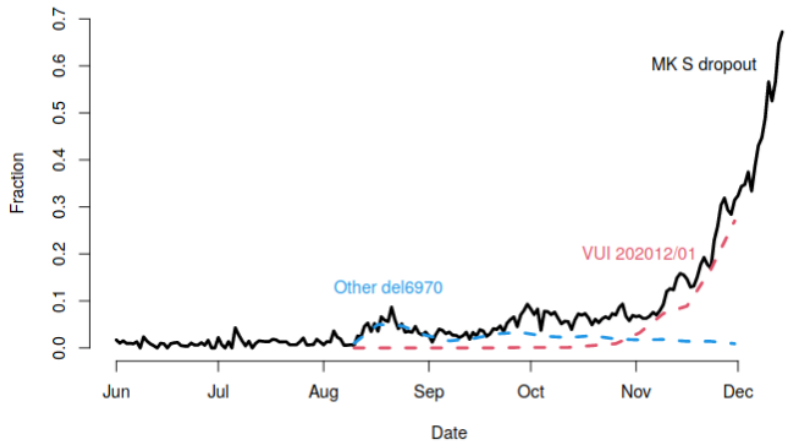
Data correct as of 15/12/2020

Analysis of cases with S gene target failure (SGTF) in diagnostic assays

- S-gene target failure (SGTF) in an assay commonly used in the LL has been shown through sequencing to be associated with deletions at position 69 and 70 (University of Birmingham)
- Deletion at 69/70 is found at the VUI but also in other lineages
- As of 16 Nov, of those SGTFs sequenced from the lighthouse laboratories, 87% were the VUI, rising to 97% in the latest genomic data in early December.
- These proportions will continue to be monitored

SGTF will be used as a proxy indicator for surveillance and trend analysis of the VUI from Monday 21 December (and retrospectively from 16 November onwards)

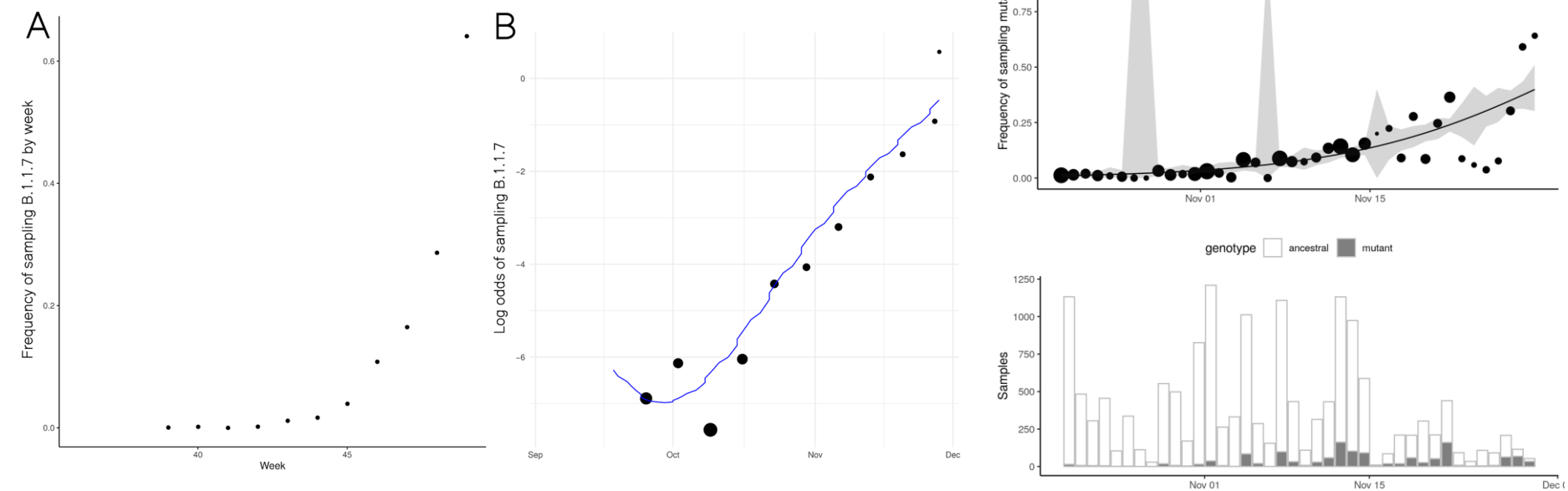
S gene target failure (provided by the Wellcome Sanger Institute)



Fraction of all Milton Keynes Lighthouse Laboratory which are SGTF, and fraction of all sequenced samples which are the VUI lineage, or other lineages including the same deletion.

The proportion of cases that are SGTF at the Milton Keynes Lighthouse Laboratory has increased sharply. Of all pillar 2 samples that are sequenced, the proportion that are the VUI has shown the same trajectory, whereas other lineages with this deletion have stayed constant frequency.

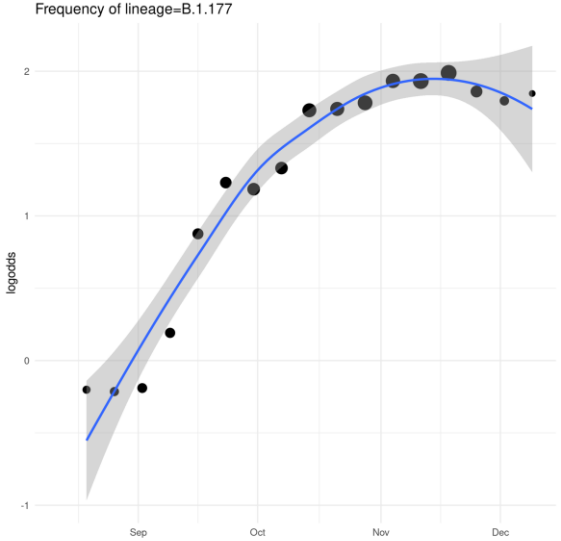
Growth in sample frequency of VUI



- Frequency of VUI in sequences from Pillar 2 sampling has increased exponentially since late November
- Change in frequency consistent with but not indicative of a constant selective advantage of VUI
- Logistic growth model indicates VUI grows +71% (95%CI: 67%-75%) faster per generation (6.5 days)
    - Limitations: Sample frequency is noisy & overdispersed in ways not captured by this model

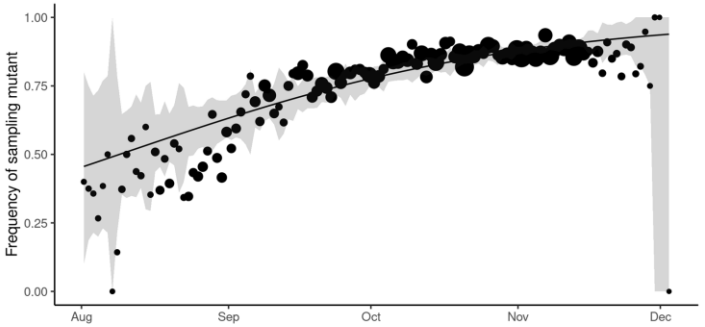
Limitations: Genetic variants can achieve high frequency even if selectively neutral

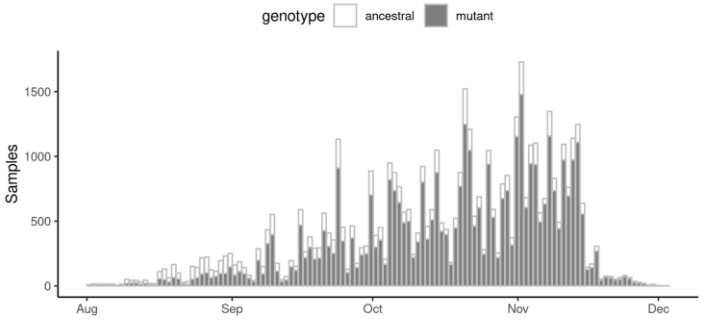
Recent Example: Frequency of B.1.177 lineage in UK with A222V variant, Multiple introductions to UK in August-October 2020

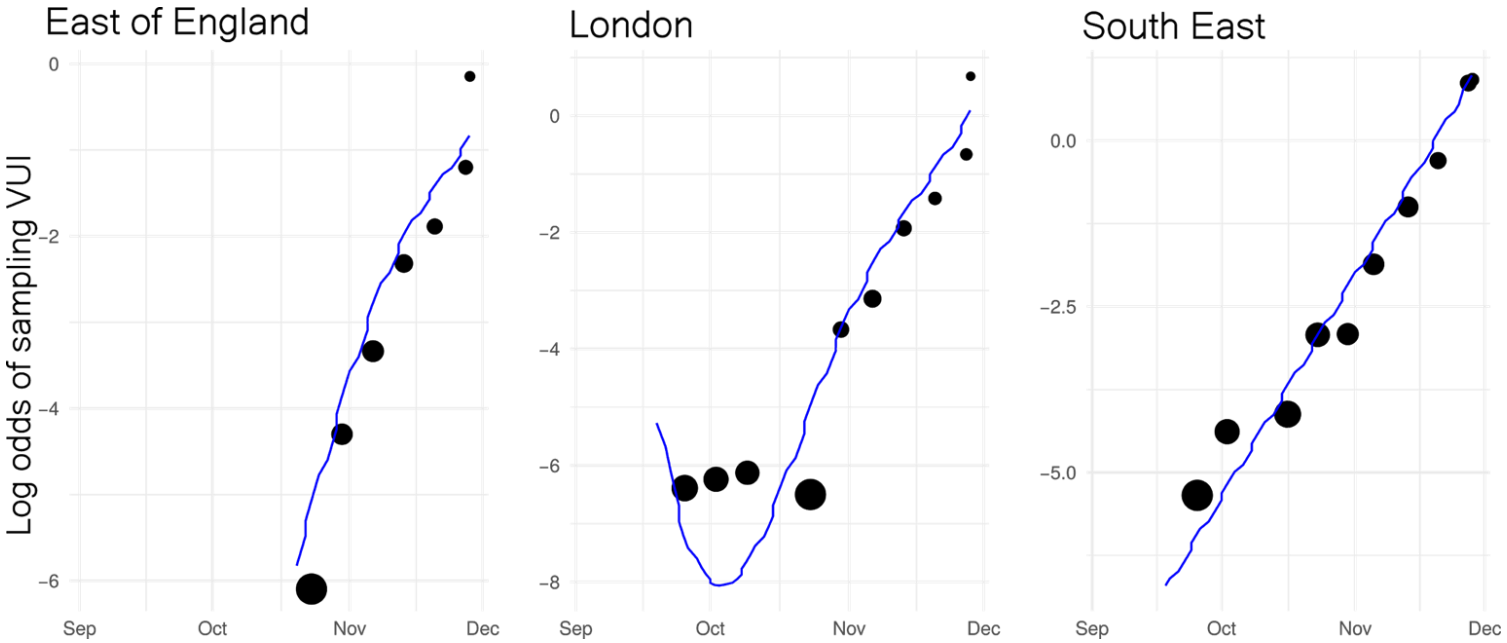


Currently >70% frequency in UK

Initial growth fuelled by holiday travel in Europe. Growth has declined with reduced travel.





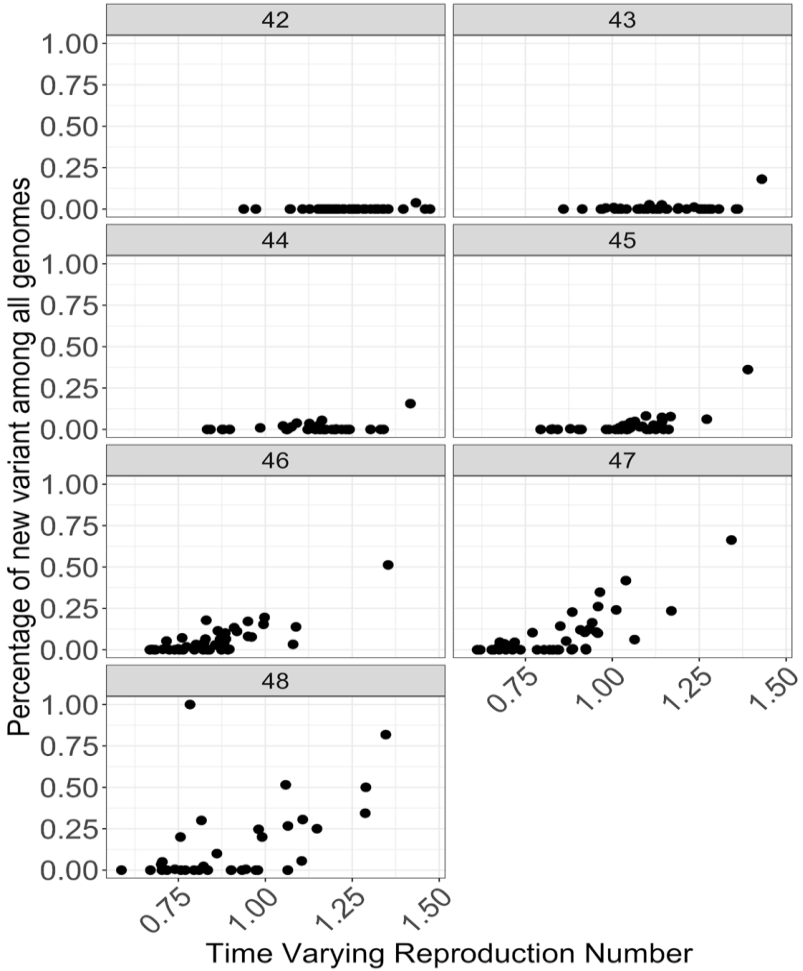


Similar rates of growth observed in different regions. Relative difference in growth rate between B.1.1.7 and other lineages:

- East of England: +72% (95%CI: 62%-82%)
- London: +86% (95%CI: 78%-94%)
- South East: +71% (95%CI: 65%-78%)

Relationship with transmission

- Time varying reproduction number[1] is correlated with the increase in fraction of new variant at many places
- Figure shows relationship between fraction of new variant among all genomes plotted against the time varying reproduction number for each week. Each datapoint is an STP area.



Coalescent phylogenetic estimate

Data
1007 genomes from London and Kent sampled by Pillar 2 from 20-Sep to 30-Nov. A second analysis performed with samples up until 21-Nov to remove potential biases from lag in sequencing and non-representative sampling towards the present.
Analysis
Analysis using BEAST v1.10.4, exponential growth coalescent model, strict molecular clock.

Results - Samples from 20-Sep to 30-Nov:
Growth rate (per year): 31.96 [95% credible interval: 25.53, 38.90]
Doubling time (days): 7.9 [6.5, 9.9]
R: 1.57 [1.45, 1.69]

Results - Samples from 20-Sep to 21-Nov:
Growth rate (per year): 40.43 [95% credible interval: 30.66, 53.21]
Doubling time (days): 6.3 [4.8, 8.3]
R: 1.72 [1.55, 1.95]

Caveats
Lag in sequencing from pillar 2 results in a drop off of sequences towards the end of November -
If this is non-random then this may cause an underestimation of the growth rate.
R estimate assumes a serial interval of 6.5 days

Regression analysis: Data

- Sequence of S:N501Y and deletion at 69-70 were used as a proxy for membership in lineage B.1.1.7
- 1451 unique pillar 2 samples collected from Sep 2nd to Nov 29 2020 across 163 local authorities areas in England
- Pillar 2 cases, deaths and new hospital admissions taken from UK dashboard
- Data aggregated by STP regions and week

Methods

- R_t for each STP per week modelled as a weekly random walk process and estimated using a semi-mechanistic Bayesian model from case and death data
  (Mishra, et al, medRxiv 2020.11.24.20236661; doi: https://doi.org/10.1101/2020.11.24.20236661)

- Then regress R_t for each STP against the fraction of the new variant, with categorical variables for each STP area and for each week to account for spatiotemporal effects (two variants – unweighted and weighted)

  ➔ Additive model : estimate the exact amount of increase or decrease in R_t by using R_t as response in the linear model

  ➔ Multiplicative model : estimate the relative increase or decrease in R_t by using $\log(R_t)$ as the response variable in the linear model

Results

- Additive model (unweighted): increase in R_t of 0.39 [0.24-0.55]

  ➔ For example, under the additive assumption, an area with an R_t of 0.8 without the new variant would have an R_t of 1.19 [1.04-1.35] if only N501Y was present

- ~~Multiplicative model: relative increase in R_t of 48% [27%-74%]~~

Limitations and assumptions

- Frequency may be underestimated from genomic data
- Confidence intervals assume independence of the observations, homoscedasticity and normality of the observations
- Spatial correlation has not be taken into consideration
- Population is considered homogeneous and all age bands are considered equally
- No causal relationship established. Only associative effects are estimated

References

[1] A COVID-19 Model for Local Authorities of the United Kingdom
Swapnil Mishra, Jamie Scott, Harrison Zhu, Neil M. Ferguson, Samir Bhatt, Seth Flaxman, Axel Gandy
medRxiv 2020.11.24.20236661; doi: https://doi.org/10.1101/2020.11.24.20236661

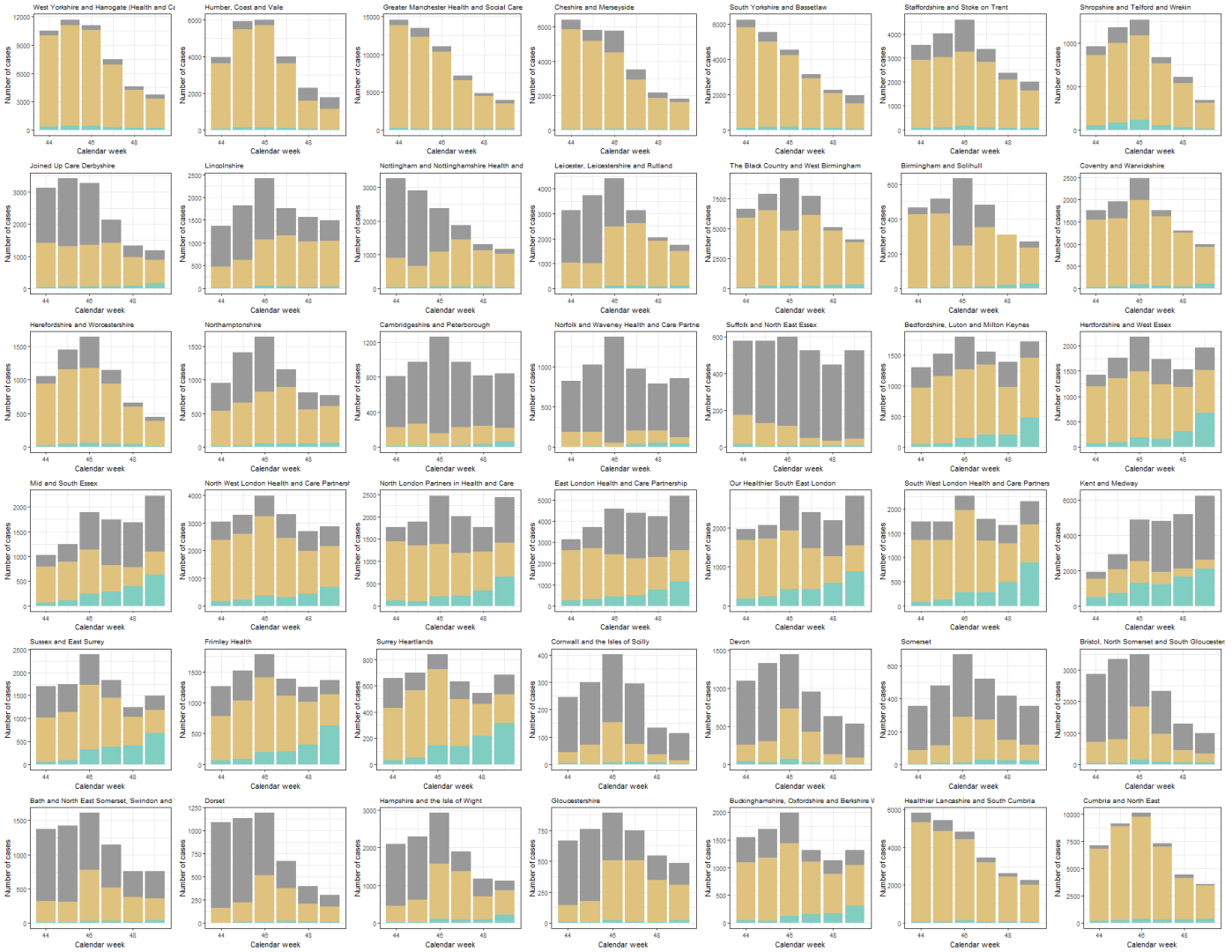
Analysis of transmissibility based on SGTF

S-gene drop-out data

- Currently, >97% of pillar 2 PCR tests which test negative on S but positive on other targets are due to the new variant
- So can use frequency of S-gene negatives among PCR positives as a proxy for frequency of the variant
- This allows us to reconstruct incidence trends for variant and non variant strains
- Valid for epiweeks 44-49 (25th October-5th December)
- Poorer proxy going further back in time
- Data issues after 5th (being worked on)
- Additional caveat: not all pillar 2 cases have S gene results – have to assume these are distributed like the ones that do

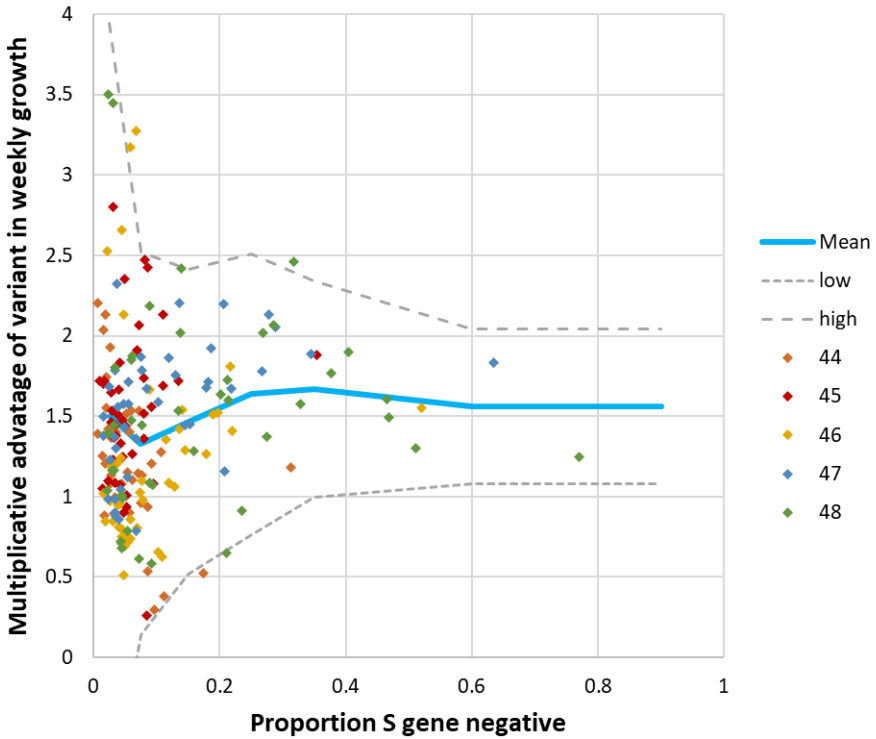
S-gene pillar 2 data

- Case incidence by STP region for variant and non-variant



Variant and non-variant growth rate trends from S-gene data

- Look at weekly growth factors: (cases next week)/(cases this week) - for variant and non-variant
- Calculate ratio of growth factors for each STP area (50 in England and week), and plot against variant frequency in that week
- Mean ratio of growth factors (corrected by power of 6.5/7 to give reproduction number scale) = 1.47 (95% CI: 1.34-1.59)
- Slightly higher for areas/weeks with higher frequencies: 1.56 (1.50-1.61) for >0.1



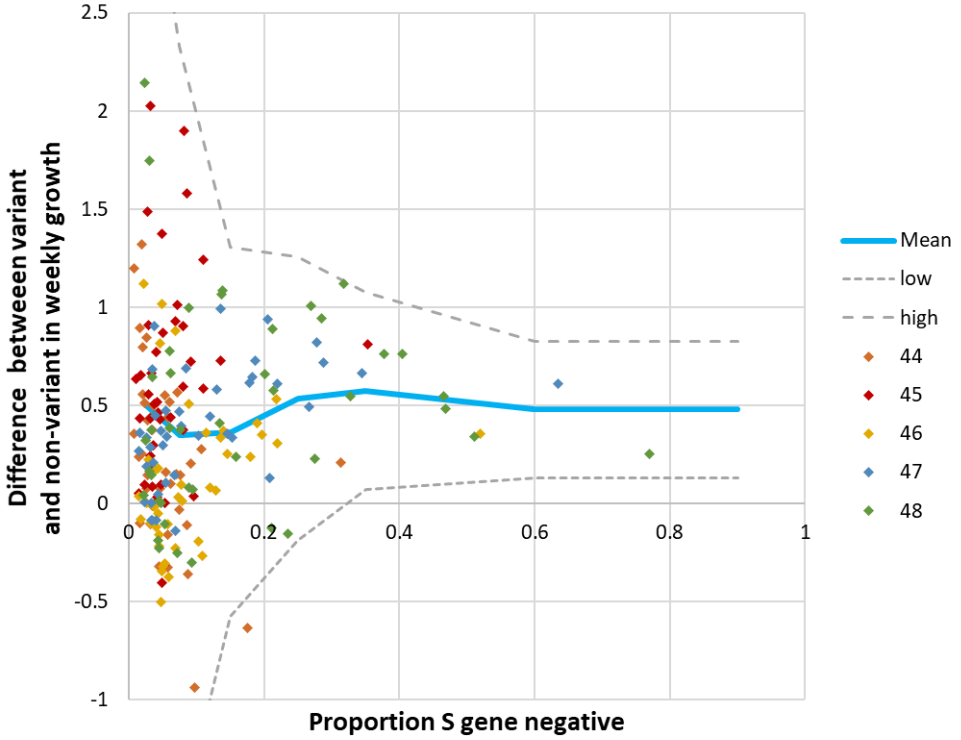
Change in R or r?

- Treat STP weekly growth trends for variant and non variant as case-control pairs
- 42x5=210 such pairs spanning weekly growth factors in weeks 44-48 (case data from 44-49)
- Higher proportion of pairs where S+ grows than S- ($p<1e-4$)
- Fisher exact test two-sided $p<1e-4$, coefficient of association =0.31 – indicating correlation between S+ and S- growth (unsurprising given geographic variation)
- McNemar or exact binomial test (comparing off-diagonal elements) $p<1e-4$ – indicating S- has more likely to have growth>0, controlling for week and STP variation
- Limited specificity of S-gene (plus low counts) in early weeks likely reduces effect size

	S+ growth>0	S+ growth<=0
S- growth>0	64	70
S- growth<=0	13	63

Variant and non-variant growth rate trends from S-gene data

- Can also look at differences in R calculated from weekly growth factors of variant and non-variant
- Mean R difference across all STPs and weeks 44-48 = 0.45 (95% CI: 0.28-0.60)



Regression analysis

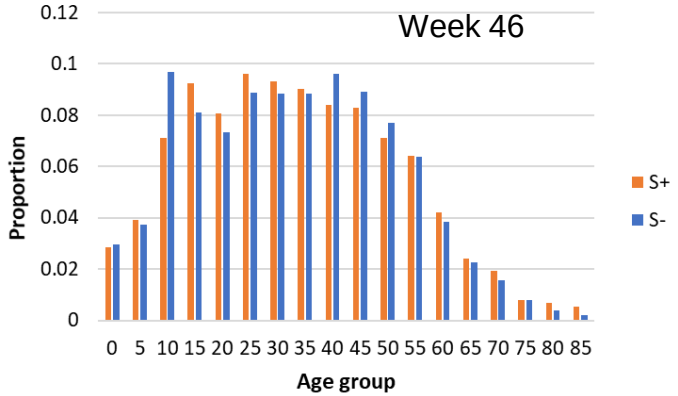
- As analysis based on genomic frequencies previously described, but using S-gene dropout frequency data
- Regress weekly R estimated from all COVID-19 cases against frequency of variant across all LTLAs
- Include week and LTLA fixed effects
- Only sensible for additive model (arithmetic difference in R)
- Variant has 0.40 (0.33-0.53) higher R across weeks 44-49

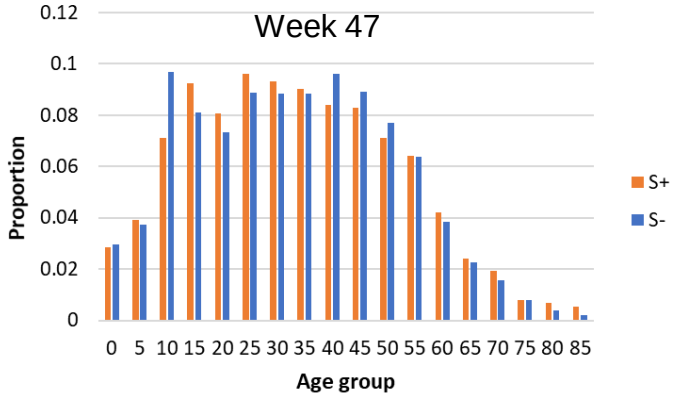
LTLA analysis

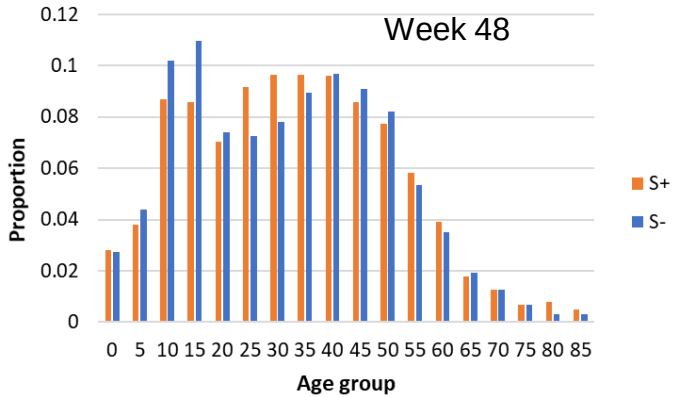
- Use *Epidemia* to estimate R over time by LTLA separately for variant and non-variant
- Gives very similar estimates

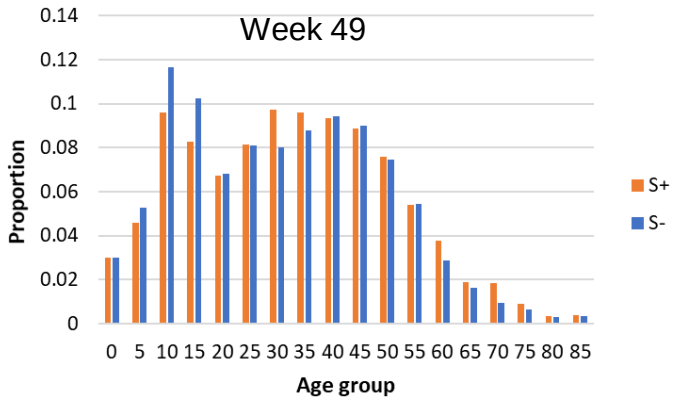
Age distribution of cases

- Weekly age distribution of S+/S- pillar 2 cases
- S+ cases adjusted to match distribution of S- cases across STPs each week
- Age shift towards 10-20 year-olds becomes clearer as frequency increases
- Note: lockdown in effect in these weeks









Phenotype of changed transmissibility

- Both multiplicative and additive models are likely overly simplistic
- Transmissibility changes may be focussed on particular population groups or transmission contexts
- Hint that new variant infects children more readily – comparison controlling for area and week indicates a higher proportion of cases in <15s