BMJ Open

Modelling of the growth and reproduction number of SARS-CoV-2 novel Variants of Concern (VOC) in the United Kingdom. Analysis of the factors affecting the spread of these VOCs.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-056636
Article Type:	Original research
Date Submitted by the Author:	25-Aug-2021
Complete List of Authors:	Ward, Thomas; Public Health England, Department of Health and Social Care Glaser, Alex; United Kingdom Department of Health and Social Care Johnsen, Alexander; United Kingdom Department of Health and Social Care Xu, Feng; The University of Manchester Hall, Ian; University of Manchester, Department of Mathematics Pellis, Lorenzo; The University of Manchester
Keywords:	Epidemiology < INFECTIOUS DISEASES, COVID-19, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Modelling of the growth and reproduction number of SARS-CoV-2 novel Variants of Concern (VOC) in the United Kingdom. Analysis of the factors affecting the spread of these VOCs.

Thomas Ward[1*], Alex Glaser[2], Alexander Johnsen[2], Feng Xu[3], Ian Hall[3], Lorenzo Pellis[3]

1. Public Health England, London UK
2. Department of Health and Social Care, London UK
3. University of Manchester, Manchester UK

*Corresponding Author: Tom.Ward@dhsc.gov.uk

Abstract

Objectives: Importations of novel variants of concern (VOC), particularly B.1.617.2, have become the impetus behind recent outbreaks of SARS-CoV-2. Concerns around vaccine efficacy, the impact on transmission and severity are now driving the public health response to these variants. This paper analyses the patterns of growth in hospitalisations and confirmed cases for novel variants of concern by age groups, geography, and ethnicity in the context of changing behaviour, non-pharmaceutical interventions (NPIs) and the UK vaccination programme. We seek to highlight where strategies have been effective and periods that have facilitated the establishment of new variants.

Methods and Design: We have algorithmically linked the most complete testing and hospitalisation data in England to create a dataset of confirmed infections and hospitalisations by SARS-CoV-2 genomic variant. We have used this fully sequenced genomic data and targeted gene sequencing to analyse geographic and demographic distinctions. To assess the instantaneous growth in variants of concern we have developed a Generalised Additive Model (GAM) fit to multiple splines and a varying day of the week fixed effect. We have further modelled the instantaneous reproduction number R_t for B.1.1.7 and B.1.617.2 variants using a doubly interval censored model to temporally adjust the confirmed variant cases.

Results: We can observe a clear replacement of the predominant B.1.1.7 by the B.1.617.2 variant without observing sustained exponential growth in other novel variants. Modelled growth of triple positive variants was initially detected in the youngest age groups, although we now observe across all ages a very short doubling times of 10.7 (CI: 9.1, 13.2) days and 8 (CI: 6.9, 9.1) days for cases and hospitalisations respectively , including age groups that have been largely doubly vaccinated. We observe that growth in triple positive variants was first detected in the Indian ethnicity group in late February, with a peak of 0.6 (CI: 0.07, 0.05) in the instantaneous growth rate, but is now maintained by the white ethnicity groups, observing a doubling time of 6.8 (CI: 4.9, 11) days. R_t analysis indicates a reproduction number advantage of 0.45 for B.1.617.2 relative to B.1.1.7, with the triple positive variant's R_t value peaking at 1.85.

Conclusions: Our results illustrate a clear transmission advantage for B.1.617.2 and the growth in hospitalisations illustrates that this variant is able to transmit and cause serious illness within age groups that are largely doubly vaccinated. There are concerning signs of intermittent growth in the B.1.351 variant, reaching 28 day doubling time peak in March 2021, although this is presently not showing any evidence of a transmission advantage over B.1.617.2.

Strengths and limitations of this study

- Geographic sequencing bias is evident in England with the North West having the highest coverage, which leads to larger confidence intervals for other areas.
- There is limited fully sequenced data so targeted sequencing was used for regions, lower tier local authority, and age groups. Fully sequenced data was used to complement the targeted sequenced data at a national geography and to illustrate the dominance of B.1.617.2.
- We did not include specific information on the vaccination status of each individual involved as this would subset the data further and reduce the feasibility of the analysis. Therefore, age groups were used a proxy for vaccination status.
- To calculate R_t we assumed a consistent ascertainment bias in the testing data.
- The study calculates the growth and R_t in clinical and sequenced testing data therefore, conclusions are inferential and descriptive with regards to the impacts of NPIs and vaccination
- The analysis illustrates the impact of demographics, geography and behaviour on the introduction and growth of novel variants

- There is currently no study published that looks at the link between the growth in hospitalisations by variants of concern (VOC).

Introduction

The emergence of SARS-CoV-2 has had an unprecedented impact and global reach since the first officially confirmed case in December 2019 (WHO, 2020). Periods of high global prevalence of the virus has driven novel mutations through antigenic drift, with evidence this is largely a reaction to the host immune response (Wang, et al., 2020) and we may now begin to see selective mutation in response to natural and vaccine induced immunity. The rate of mutation for coronaviruses have been poorly understood; however, evidence from seasonal human coronaviruses HCoV-229E and HCoV-OC43 illustrate that the evolution of SARs-CoV-2 may have parallels with the influenza A virus (IAV), including more concerning adaptive changes to the receptor binding domains (RBD) (Jo, et al., 2021). Viruses akin to SARS-CoV-2, that are RNA based, tend to show high rates of mutation, which are likely to be related to insufficient proofreading abilities (Yin, 2020). Imports of novel variants of COVID-19 are now of great concern as they become the impetus behind localised outbreaks in the United Kingdom (Department of Health and Social Care, 2021). A recent UK government modelling report highlighted the significance of importations (Kucharski, et al., 2021) finding that individuals that had recently travelled had a higher relative reproductive number.

The vaccination campaign began in England on 8th December 2020 with care home residents, the most clinically vulnerable, and hospital staff. This was followed by an age stratified structure that began with the over 80s on the 17th January 2021 and reach the 21-30 age group on the 16th June 2021 (Department of Health and Social Care, 2021). The vaccination campaign began with Pfizer/BioNTech and AstraZeneca with first doses prioritised. The age groups over 40 were primarily administered with AstraZeneca; Pfizer/BioNTech and Moderna used largely for the younger age groups in response to concerns over haemostatic side effects (Pottegård, et al., 2021). The chief concerns around importations of novel variants have been driven by immunological escape. A recent trial in South Africa (Madhi, et al., 2021) found that the AstraZeneca vaccine had a two-dose efficacy of 10% against B.1.351 at preventing mild to moderate disease, albeit this study utilised very limited data. Moreover, the B.1.617 variant, that was first detected in October 2020 in India, carries two mutations on the RBD and preliminary results indicate this may have an impact upon vaccine efficacy (Hoffmann, et al., 2021) (Ferreira, et al., 2021). However, it is B.1.617.2, a sub-lineage of B.1.617, that is now causing global concern. A recent study (Planas, et al., 2021) that analysed the sera of patients infected with B.1.617.2 found that it has 9 spike mutations on the N terminal domain (NTD) and the RBD. This study observed that B.1.617.2 is resistant to neutralisation with the efficacy of the Pfizer vaccine around 3 to 6 times less than observed with B.1.1.7. Nonetheless, there is some evidence that doubly vaccinated individuals may still possess robust neutralisation titres against B.1.617.2 and there is still relatively high efficacy against symptomatic disease (Bernal, et al., 2021). However, these results do not take into account that symptomatic status is poorly recorded for PCR tests in England and that fully sequenced B.1.617.2 variant cases were limited at the time.

The evidence of substantial viral epitopic mutation has necessitated a risk categorisation for novel mutant strains in the United Kingdom. Variants that display epidemiologically and immunologically characteristics of concern are defined as initially a Variant Under Investigation (VUI) (Public Health England, 2021) and after committee evaluation may be escalated to a Variant of Concern (VOC). As of the 12th May 2021 there are eight VUIs and five variants defined as VOCs: B.1.1.7 (VOC-20DEC-01) *Alpha*, B.1.351 (VOC-20DEC-02) *Beta*, P.1 (VOC-21JAN-02) *Gamma*, B.1.1.7 with E484K (VOC-21FEB-02), and B.1.617.2 (VOC-21APR-02) *Delta* (Public Health England, 2021). The most concerning VOCs presently are B.1.351 and B.1.617.2 due to evidence of diminished vaccine efficacy, particularly in the former. There is also growing evidence that B.1.617.2 acquired mutations that increased the viral fitness improving the transmissibility of this lineage.

In this paper we have utilised gene targeted and fully sequenced confirmed tests for COVID-19 that have been algorithmically linked to hospitalisation datasets. We assess the temporal variability in the growth of VOCs relative to the predominant B.1.1.7 variant across the geography of the United Kingdom. We further assess how the instantaneous growth rate has changed across ages, ethnicity, and the response to the easing of nonpharmaceutical interventions (NPIs). Finally, we assess the relative difference in the reproductive number between novel triple positive VOCs and the established B.1.1.7 variant.

Methods

Epidemiological and Clinical Data

Accident and Emergency (A&E) dataset is from the SUS suite of datasets. This data was linked with the PCR targets from the Second Generation Surveillance System (SGSS). The linkage allows the patient's pathway to be followed and provides additional information beyond what is obtainable from the standalone datasets. The linkage algorithm has evolved from research and development undertaken on the 2 datasets. The algorithm is primarily based on fields that:

- identify the patient, using a pseudo identifying number

- report the outcome of the A&E attendance, and

- report the method of admission.

Further, the basic principles behind the linkage method are where:

i) the unique patient pseudo identifier is the same in A&E and SGSS data, and
ii) the SGSS Specimen date is between 6 days before or 14 days after the A&E admissions date.

For multi-episode spells the admission date used for the linkage comes from the first episode in a spell. Linkage was conducted in a secure research environment and with full anonymisation of the data. The linked hospitalisation data was subset nationally by age, region and LTLA.

PCR testing for COVID-19 involves a detection of three genes OR, N, and S. The S-gene mutation in B.1.1.7 results in a dropout of S-gene detection, providing an easier prevalence indicator for this variant where targeted sequencing has been conducted. Due operational and logistical limitations fully sequenced viral genome data was limited and therefore we employed S-dropout to identify B.1.1.7 and triple positive (OR, N, and S gene positive) was used as a proxy for the identification of VOC B.1.351, B.1.617, P1, B.1.617.2, and P2. We assessed S-dropout and triple positive confirmed positive cases from the Public Health England NPEX data which was subset by travel status, ethnicity, age, region, and Lower Tier Local Authority (LTLA). Further fully sequenced data was acquired through SGSS and suspected variants from the reflex assays. The P.1, B.1.617.2, B.1.351, and B.1.1.7 variants were included in this analysis and other variants were excluded due to low numbers. The last analysis was only conducted at a national spatial resolution.

Instantaneous Growth Rate and Doubling Times

The method for the estimation of the time varying growth rates and doubling times is adapted from a Generalised Additive Model (GAM) with a canonical link (Wood, 2018) (Wood, 2017). We allow for a varying day of the week fixed effect: no day, weekend, or weekday effect. We further fit to cubic regression splines (Wood, 2006), P-splines (Eilers & Marx, 1996), thin-plate splines (a low rank isotropic smoother) (Wood, 2003), Duchon splines (allowing for lower orders of the derivative in the penalty relative to the thin plate splines) (Duchon, 1977) and Gaussian process smoothers. The model assumes the number of cases $\mathrm{y(t)}$ is proportional to $\exp\left(s(t)\right)$ for some smoother $\mathrm{s(t)}$ (Pellis, et al., 2021). The over-dispersed noise inherent in both disease dynamics and surveillance data motivates the use of a negative binomial error structure. The instantaneous growth rate is obtained as the time derivative of the smoother, $r_s = \dot{s}(\mathrm{t})$, and the instantaneous doubling time is calculated as $t_D = \log(2)/\dot{s}$. Asymptotic confidence intervals (CIs) on r_s are only indicative of uncertainty on t_D, especially when the variance grows as r_s approaches zero. The number of knots used by the spline is fixed as one twentieth the length of the time-series (for time-series shorter than 200 days the default number of knots is used) to avoid over smoothing the data or loosing signal in the noise. The model for each group, fit to each spline and day of the week effect, is assessed by the leave-one-out (LOO) and the Akaike information criterion (AIC) metrics to select the best model fit. Included in each plot is the date of the first confirmed case for B.1.1.7, B.1.617.2, B.1.351, and P.1 as they were considered of most concern at this time due to overall volume and phenotypic characteristics. In addition, each step of the lockdown easing (Cabinet Office, 2021) has been included:

- Step 1a – Schools and universities are to re-open, care homes allow visitors and recreation within households and support bubbles are allowed.
- Step 1b – The 'stay at home' rule will end and outdoor sports to resume. Furthermore, the rule of 6 begins and two household can meet outdoors.
- Step 2 – Non-essential retail, gyms and outdoor hospitality will reopen.
- Step 3 – A lifting of most legal restrictions on mixing outdoors, events of up to 30 persons can be held, and indoor hospitality can recommence.

Instantaneous Reproduction Number R_t

This model utilised the targeted sequencing data for confirmed positive tests of the S dropout and triple positive targets as a proxy for infections by those variants. We calculate the instantaneous reproduction number (Cori, et al., 2013) that corresponds

to the average quantity of secondary cases that develop from the primary cases infected at a time period we call t, if conditions remained constant. Defined as:

$$R_t = \frac{I_t}{\sum_{s=1}^{t} I_{t-s} w_s}$$

Here I_t is defined as the quantity of incidence on day t and w_s is the discretised generation time distribution.

This approach was utilised as it is a reflection of the instantaneous transmissibility at a given point in time of the SARS-CoV-2 variant. The limitation of this approach is that it assumes there is a reasonable consistency in the ascertainment bias in the testing data and those tests included for targeted sequencing. Individuals that had a flag for recent travel were removed from the dataset prior to analysis.

The serial interval pertains to the duration of time from the onset symptoms of a primary case to the onset of symptoms for a secondary case that was infected by the primary case. To account for uncertainty in the serial interval we utilised a Monte Carlo simulation model of a meta-analysis (Reed, et al., 2021) that included studies which have published on the mean and standard deviations of the serial Interval for SARS-CoV-2.

The most complete available testing data for England is recorded at specimen date of the test. To calculate the instantaneous reproduction number, we would optimally utilise the symptom onset date of positive cases at time t. Therefore, to adjust for this temporal discrepancy we have adapted the time delay Bayesian Monte Carlo Markov Chain (MCMC) doubly interval censored model from Ward & Johnsen, (2021) to calculate temporal changes over time for the lag from symptom onset to specimen date fitting to a Weibull distribution.

Ethical Summary

The data employed in this study were fully anonymised prior to use and linkage was conducted through a strict a non-identifiable process.

Patient and Public Involvement Statement

Patients were not involved in the development of the research question and study design.

Results

Across all age groups in England we can observe that the decay rate for confirmed positive triple positive cases peaked at the start of February in *Fig 1*, with the model fit illustrated in *Fig 2* and *Fig 3*. The exponential decay of triple positive cases has since slowed and there has been in exponential growth in these variants from the 25th March, which is 10 days after the first confirmed case of B.1.617.2. Conversely, we observe exponential decay in the S-dropout cases from the 7th January. We can observe that Step 1a occurred shortly before the exponential increase in growth of triple positive cases and a steeper gradient in the line can be seen after Step 1b and Step 2.

Fig 1 Instantaneous growth rate and doubling times for confirmed positive tests in England by triple positive and S dropout variants

Fig 2 The maximum likelihood model fit to the S dropout confirmed positive testing data

Fig 3 The maximum likelihood model fit to the triple positive confirmed positive testing data

We can observe a sharp reduction in the rate of the exponential decay for triple positive hospital admissions from the 14th March in *Fig 4*, which was 1 day before the first confirmed case of B.1.617.2. The data fit to the model can be seen in *Fig 5* and *Fig 6* for S dropout and triple positive variants respectively. We observe exponential growth in triple positive hospitalisations from the 8th April which is 13 days later than this growth was detected in the testing data. Exponential decay in hospitalisations can be observed from the 16th January for S-dropout admissions after a peak in mid December, which proceded the second

wave of SARS-CoV-2 hospitalisations in England. Wider confidence intervals of S-dropout hsopitalisations in mid June are a result of the low numbers now observed.

Fig 4 Instantaneous growth rate and doubling times of hospitalisations in England by variants

Fig 5 The maximum likelihood model fit to the S dropout confirmed hospitalisation data

Fig 6 The maximum likelihood model fit to the triple positive confirmed hospitalisation data

Targeted Sequencing - Confirmed SARS-CoV-2 Positive Tests

Age Group

Analysis of testing data across ages illustrates that the earliest reduction in the speed of the decay rate was observed for triple positive variants was in the youngest age groups (in *Fig 7* and *Fig 8*) from the end of January. This was followed very shortly by exponential growth in triple positive variants in all ages as can be seen in *Appendix A*. We observe slightly wider confidence intervals in the over 75-year-old age group, seen in *Fig 10* and *Fig 11*, due to smaller numbers, at this time, producing greater uncertainty. In these age groups we also observe smaller numbers leading to larger confidence intervals in s dropout cases that overlap into positive growth.

Fig 7 Instantaneous growth rate and doubling times for confirmed positive tests for ages 0-24

Fig 8 Instantaneous growth rate and doubling times for confirmed positive tests for ages 25-34

Fig 9 Instantaneous growth rate and doubling times for confirmed positive tests for ages 65-74

Fig 10 Instantaneous growth rate and doubling times for confirmed positive tests for ages 75-84

Fig 11 Instantaneous growth rate and doubling times for confirmed positive tests for ages over 85

Region

In the regional testing data we can observe that the North East *(Fig 12)*, North West *(Fig 13)* and Yorkshire and the Humber *(Fig 14)* have seen the largest exponential growth in the triple positive variants, with the full results in *Appendix B*. We do, however, observe small growth rate in s dropout cases in these regions from the start of May and slightly larger in the South West. Nonetheless, we observe large confidence intervals and the South West is known to have poor targeted gene sequencing coverage, which will lead to greater uncertainty in the growth estimates.

Fig 12 Instantaneous growth rate and doubling times of confirmed positive tests for the North East region

Fig 13 Instantaneous growth rate and doubling times of confirmed positive tests for the North West region

Fig 14 Instantaneous growth rate and doubling times of confirmed positive tests for Yorkshire and the Humber region

Fig 15 Instantaneous growth rate and doubling times of confirmed positive tests for the South West region

Ethnicity

We can observe in *Fig 16* that exponential growth in triple positive variants began initially in the Indian ethnicity group from late February, which is now in exponential decay. Moreover, we observe very steep growth in the Pakistani ethnicity group, in *Fig 17*, from 10th April that coincided with religious festival of Ramadan. It is evident that the growth of the triple positive variants since

the 7th April has been strongly sustained by the white ethnicity groups *Fig 18* with a doubling time of approximately 7 days and the full results can be found in *Appendix C.*

Fig 16 Instantaneous growth rate and doubling times of confirmed positive tests for the Indian ethnicity group

Fig 17 Instantaneous growth rate and doubling times of confirmed positive tests for the Pakistani ethnicity group

Fig 18 Instantaneous growth rate and doubling times of confirmed positive tests for the White ethnicity group

Lower Tier Local Authority

From the subset of LTLAs analysed, full results can be seen in *Appendix D*, we observe the largest growth rates for triple positive variants at the time of this study are in Blackpool 3.9 (CI: 3.2, 4.8) days and Kirklees 5.6 (CI: 4.5, 7.4) days. Bolton saw the earliest exponential growth in triple positive variants before interventions reaching 0.16 (CI: 0.15, 0.18) days, that included surge testing, increased vaccination and public health awareness campaigns (Public Health England, 2021), appeared to slow growth substantially. It is striking in LTLAs in the North West like Manchester the rate of exponential decay began to reduce from the end of January, which was over a month before the relaxation in NPIs began with Step 1a. This is not observed for s dropout variants and this illustrates that triple positive variant transmission was able to increase despite a strict national lockdown.

Targeted Sequencing - Confirmed SARS-CoV-2 Positive Hospitalisations

Age Group

Akin to the results in terms of positive cases we observe corresponding growth in the youngest age groups (*Fig 19, Fig 20, Fig 21*) also when looking at hospitalisations (full results can be seen in *Appendix E*). This will be indicative of the ages that had the largest concentration of infections. The oldest age groups similarly have a slightly more heterogenous picture, although we are still observing strong growth in triple positive variant hospitalisations in *Fig 22, Fig 23,* and *Fig 24* for those 65 and over.

Fig 19 Instantaneous growth rate and doubling times of hospitalisations for the ages 0-24

Fig 20 Instantaneous growth rate and doubling times of hospitalisations for the ages 25-34

Fig 21 Instantaneous growth rate and doubling times of hospitalisations for the ages 35-44

Fig 22 Instantaneous growth rate and doubling times of hospitalisations for the ages 65-74

Fig 23 Instantaneous growth rate and doubling times of hospitalisations for the ages 75-84

Fig 24 Instantaneous growth rate and doubling times of hospitalisations for the ages over 85

Region

Corresponding to the analysis of the testing data, the regional hospitalisation analysis shows the most substantial growth in hospitalisations with the tightest confidence intervals in the North West (*Fig 25*) and the South East (*Fig 26*), with the full results in *Appendix F*. The highest central estimate is in the North East (*Fig 27*) of 3.4 days, but with large confidence intervals that may be related to poor CT value coverage in this area.

Fig 25 Instantaneous growth rate and doubling times of hospitalisations for the region the North West

Fig 26 Instantaneous growth rate and doubling times of hospitalisations for the region the South East

Fig 27 Instantaneous growth rate and doubling times of hospitalisations for the region the North East

Lower Tier Local Authority

The strongest growth with the tightest confidence intervals we observe in the LTLAs of concern are in Salford and Trafford with doubling times of 3.3 (CI: 2.4, 5.8) and 4 (CI: 2.5, 10.7) days respectively, with full LTLA analysis in *Appendix G*. However, the tighter confidence intervals in the North West are due to sequencing geographic bias. We also observe in an LTLA in Yorkshire and the Humber, Kirklees, a comparably short doubling time of 4.2 (CI: 2.5, 12.4) days but with larger confidence intervals that are a by-product of the poor targeted sequencing coverage.

The Instantaneous Reproduction Number R_t

To parametrise the R_t model we have calculated the minimum, maximum, standard deviation and mean of the values from Reed, et al., (2021) meta-analysis in *Appendix H* to create a MCMC simulation of the distribution. For this model we further calculated the time lag from symptom onset date to specimen date in England that can be seen in *Appendix I*. There has been a marked reduction from the first wave of SARS-CoV-2 in England, from January to May 2020, which may be related to an improved public health message and more effective contact tracing.

Analysis of the instantaneous reproduction in *Fig* 28 number illustrates the from mid-March R_t has begun to grow for triple positive variants with a steep increase by mid-April reaching 1.85. B.1.1.7 has been below 1 since January apart from a brief period of growth in March. The R_t value implies greater transmissibility of the triple positive dominant variant B.1.617.2 and from the time of the first confirmed case we observe an average reproduction number advantage of 0.445.

Fig 28 The instantaneous reproduction number for s dropout and triple positive variants

Fully Sequenced and Reflex Assays - Confirmed SARS-CoV-2 Positive Cases

Analysis of the fully sequenced genomic data in *Fig 29* illustrates that since the first detected case of B.1.617.2 in England there has been sustained exponential growth with a doubling time of 7 days now observed. It is apparent that P.1 has not managed to gain much traction and been in a steady decline from the time of first detection and importation. B.1.1.7, as observed in the S-dropout results, has been in negative growth since February after a period of high prevalence contributing to the second wave in the United Kingdom. B.1.351 conversely has seen periods of growth and decay but without a substantial period of sustained growth that would allow this variant to become prevalent and more dominant. However, we observe consistent and concerning signs that B.1.351 continues to grow despite early fluctuations. It is not clear that any of the NPI easing had an impact on B.1.1.7 growth however, it is likely to be related to competition with B.1.617.2 rather than transmission blocking effects of vaccination campaigns.

Fig 29 Instantaneous growth rate and doubling times of confirmed positive cases by the highest priority variants of interest

Discussion

There has been a reduction in the exponential decay of triple positive cases since February in England and exponential growth since April. It is evident from *Fig 35* that B.1.617.2 is the dominant triple positive variant and that other imported variants have thus far failed to compete effectively and have been largely in exponential decay. The results suggest that importations were initially concentrated in the North West of England, particularly Bolton, before appearing in Manchester, Trafford and Salford. It is evident the variant has spread across the country with areas in Yorkshire and the Humber like Kirklees now seeing some of the most significant growth. S-dropout cases, the proxy for B.1.1.7, have conversely been largely in exponential decay since the January national lockdown in the UK with isolated areas of growth. Further research should focus on whether we can observe greater growth in B.1.1.7 in the areas that have greater proportion of unvaccinated individuals.

A limitation of this study is that it does not directly include the vaccination status of the infections in the analysis and therefore, the analysis employs age as a proxy for vaccination status. This is due to the limited amount of sequencing conducted, which after linkage with vaccination status preclude meaningful analysis on the growth and reproduction number of the groups included. The very high rates of vaccination in the oldest age groups seen in *Table 2* and *Table 3* illustrate a clear stratification between ages: by the end of June over 91.1% of those over 60 had received two doses of the vaccination and that most individuals under the age of 50 have not received their second dose of the vaccination by July

The arrival of a VOC can result in higher rate of sequencing for specific LTLA geographic locations. However, this targeted approach was not conducted to a considerable extent for B.1.617.2 due to how quickly widespread transmission became in England. However, the sporadic growth we observe in B.1.351, which has failed to maintain growth in the absence of importations, has been influenced by surge testing and enhanced contact tracing of locations where this variant has been found.

The primary limiting factor for the sequencing data is geographic location with the North West having the largest amount of sequencing conducted and some of the lowest levels seen in the South West and Yorkshire. Although, this geographic bias is temporally variable as changes in laboratory capabilities evolve over the pandemic.

Month	18-29	30-39	40-49	50-59	60+
Dec	0.5	0.8	1.2	1.4	5.3
Jan	4.6	6.0	8.1	10.2	40.0
Feb	9.2	12.6	18.7	28.8	79.3
Mar	15.9	21.9	36.5	81.5	91.8
Apr	17.3	24.9	55.5	84.4	92.4
May	21.8	47.8	74.2	85.9	92.9
Jun	50.2	63.1	76.9	86.7	93.2

Table 2 Proportion of each age group who have recived their first vaccination by the end of each month.

Month	18-29	30-39	40-49	50-59	60+
Dec	0.0	0.0	0.0	0.0	0.0
Jan	0.1	0.2	0.3	0.4	2.6
Feb	0.3	0.5	0.7	0.8	2.9
Mar	2.3	3.3	4.5	5.6	15.7
Apr	6.3	8.4	12.0	17.3	62.0
May	11.9	16.9	27.3	54.7	87.4
Jun	16.7	24.9	50.1	81.9	91.1

Table 3 Proportion of each age group who have recived their second vaccination by the end of each month.

We observe in *Fig 16* that triple positive exponential growth was initially observed in February within the Indian ethnicity group, which was due to the importations of the B.1.617.2 variant. Growth in this group has subsequently declined and we can see from early April in *Fig 18* that the B.1.617.2 has now been largely sustained in the white and the black British ethnicity group, which illustrates from April the variant was no longer dependent upon importations to maintain exponential growth in England. Interestingly, we observe exponential growth in the Pakistani ethnicity group around the holiday of Ramadan and this illustrates the significance of public and religious events in driving strong growth of SARS-CoV-2, exemplified by the Christmas period in the UK when we observed similar growth in the B.1.1.7 variant. We observe that the first phase for the relaxation of the national lockdown step 1a (Cabinet Office, 2021), the reopening of schools, appeared to have the earliest impact on London for triple positive variant growth and exponential growth across the regions was already very strong before the final easing of restrictions in step 3. The East Midlands that had the slowest prior growth of s dropout cases in December had a subsequent wave of exponential growth in March that was not present or weakly observed in other regions. The high level of prevalence for SARS-CoV-2 now observed in England is facilitating sporadic growth of s dropout cases that can be seen in the North West, North East, South West and Yorkshire and the Humber where it is growing from a very small baseline.

The implications of the strong growth in triple positive cases followed by similar patterns in the hospitalisations is very significant for the implications of vaccine efficacy. We observe the most significant growth in the younger age groups that have a much lower infection hospitalisation rate (IHR) to SARS-CoV-2 infection (Birrell, et al., 2021) and have only largely received one dose of a vaccine at the point of this study (Public Health England, 2021). Nonetheless, growth within the hospitalisations will be largely indicative of the demographics where most of the infections are concentrated at that time. Significantly, we observe trailing trends in the, largely doubly vaccinated, over 65 groups particularly pronounced in the 75-84 group where we can observe a doubling time of almost 4 days. The regions that are seeing the most concerning growth in triple positive hospitalisations are the North West, North East and South East, which is very much in line with where we observed the initial growth in positive cases, although it is now evident that the variant is in exponential growth throughout England.

We can observe in *Fig 34* that the R_t number showed some growth in B.1.1.7 in April when overall cases began to initially surge across England. However, since this time it has hovered around 0.8 and we observe largely exponential decay across the country, but with some sporadic growth as can be seen in *Fig 20*. If R_t continues to be < 1 then transmission of this variant within England is likely to decay to insignificance. There is believed to be an increased risk of within household transmissibility of 60% for B.1.617.2 (Public Health England, 2021) relative to B.1.1.7. We find a similar transmission advantage with the mean difference for R_t found in this study to be 0.445.

Conclusion

To conclude, the sustained exponential growth in cases of fully sequenced B.1.617.2 and the exponential decay of other triple positive variants illustrates that this variant drives almost all of the triple positive transmission. The confirmed triple positive cases indicate that B.1.617.2 appeared earlier than the first confirmed case in March and that the relaxation in NPIs coincided with exponential growth in this variant. We have seen that growth initially began in the north of England, particularly the North West and pockets of Yorkshire and the Humber. However, this has now spread across the country and the South West has one of the smallest doubling times for triple positive hospitalisations. There is a substantial transmission advantage for the B.1.617.2 variant relative to B.1.1.7 that we estimate is around 0.45. There have been small indications of growth in B.1.1.7 with R_t above 1 in March in line with increases in B.1.617.2 but it is now clear that there has been a replacement of the predominant B.1.1.7 variant. We have observed some worrying trends in B.1.351 although it has failed to gain traction and a sustained enough period of growth for this variant to become a substantial public health concern.

Contributorship Statement

TW conceived the idea of the article. TW wrote the article. TW, LP, IH, FX, AJ, and AG developed the model methodology. AG and TW created the graphical representations. LP, AJ, TW, and IH reviewed the final draft.

Conflict of Interest

The authors have declared that no competing interests exist. The authors were employed by the Department of Health and Social Care but received no specific funding for this study.

Data Availability Statement

To access the data used for this study, an application can be made to Public Health England, Department of Health and Social Care. Data requests can be made to the Office for Data Release (https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data) and contacting odr@phe.gov.uk. All requests to access data are reviewed by the ODR and are subject to strict confidentiality provisions in line with the requirements of:

- the common law duty of confidentiality
- data protection legislation (including the General Data Protection Regulation)
- 8 Caldicott principles
- the Information Commissioner's statutory data sharing code of practice
- the national data opt-out programme

References

Bernal, J. L. et al., 2021. Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. *medRxiV,* Issue doi.org/10.1101/2021.05.22.21257658.

Birrell, P. et al., 2021. *COVID-19: nowcast and forecast.* [Online]
Available at: https://www.mrc-bsu.cam.ac.uk/now-casting/nowcasting-and-forecasting-15th-july-2021/

Cabinet Office, 2021. *COVID-19 Response - Spring 2021 (Summary).* [Online]
Available at: https://www.gov.uk/government/publications/covid-19-response-spring-2021/covid-19-response-spring-2021-summary#step-1---8-and-29-march

Cori, A., Ferguson, N., Fraser, C. & Cauchemez, S., 2013. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *Am J Epidemiology,* 178(9), pp. 1505-1512.

Department of Health and Social Care, 2021. *Further measures in additional areas to tackle Delta (B1.617.2) variant.* [Online]
Available at: https://www.gov.uk/government/news/further-measures-in-additional-areas-to-tackle-delta-b16172-variant

Department of Health and Social Care, 2021. *UK COVID-19 vaccines delivery plan.* [Online]
Available at: https://www.gov.uk/government/publications/uk-covid-19-vaccines-delivery-plan/uk-covid-19-vaccines-delivery-plan#prioritisation-1

Duchon, J., 1977. Splines minimizing rotation-invariant semi-norms in Solobev spaces. In: *Construction theory of functions of several variables.* Berlin: Springer, pp. 85-100.

Eilers, P. & Marx, B., 1996. Flexible Smoothing with B-splines and Penalties. *Statistical Science,* 11(2), pp. 89-121.

Ferreira, I. et al., 2021. SARS-CoV-2 B.1.617 emergence and sensitivity to vaccine-elicited antibodies. *BioRxiv ,* Issue 2021.05.08.443253.

Hoffmann, M. et al., 2021. SARS-CoV-2 variant B.1.617 is resistant to Bamlanivimab and evades antibodies induced by infection and vaccination. *bioRxiv,* Issue 2021.05.04.442663..

Jo, W., Drosten, C. & Drexler, J., 2021. The evolutionary dynamics of endemic human coronaviruses. *Virus Evol ,* 7(1).

Kucharski, A., Davies, N., Eggo, R. & Funk, S., 2021. *Dynamics of B.1.617.2 in England NHS regions from importations, traveller-linked and non-traveller-linked transmission.* [Online]
Available at:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993232/S1272_LSHTM_Modelling_Paper_B.1.617.2.pdf

Madhi, S. et al., 2021. Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. *The New England Journal of Medicine,* 384(DOI: 10.1056/NEJMoa2102214), pp. 1885-1898.

Pellis, L. et al., 2021. Challenges in control of Covid-19: short doubling time and long delay to effect of interventions. *Philosophical Transactions of the Royal Society.*

Planas, D. et al., 2021. Reduced sensitivity of infectious SARS-CoV-2 variant B.1.617.2 to monoclonal antibodies and sera from convalescent and vaccinated individuals. *biorxiv,* Issue doi.org/10.1101/2021.05.26.445838.

Pottegård, A. et al., 2021. Arterial events, venous thromboembolism, thrombocytopenia, and bleeding after vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: population based cohort study. *BMJ,* Volume 373.

Public Health England, 2021. *Confirmed cases of COVID-19 variants identified in UK.* [Online]
Available at: https://www.gov.uk/government/news/confirmed-cases-of-covid-19-variants-identified-in-uk
[Accessed 15 06 2021].

Public Health England, 2021. *Confirmed cases of COVID-19 variants identified in UK.* [Online]
Available at: https://www.gov.uk/government/news/confirmed-cases-of-covid-19-variants-identified-in-uk

Public Health England, 2021. *COVID-19 vaccination first phase priority groups.* [Online]
Available at: https://www.gov.uk/government/publications/covid-19-vaccination-care-home-and-healthcare-settings-posters/covid-19-vaccination-first-phase-priority-groups

Public Health England, 2021. *SARS-CoV-2 variants of concern and variants under investigation in England.* [Online]
Available at:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984274/Variants_of_Concern_VOC_Technical_Briefing_10_England.pdf

Public Health England, 2021. *Variants: distribution of cases data.* [Online]
Available at: https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-cases-data

Reed, I., Walker, E. & Landguth, E., 2021. SARS-CoV-2 Serial Interval Variation, Montana, USA, March 1–July 31, 2020. *Emerging Infectious Disease,* 27(5), p. 1486–1491.

Wang, R. et al., 2020. Host Immune Response Driving SARS-CoV-2 Evolution. *Viruses,* 12(10), p. 1095.

Ward, T. & Johnsen, A., 2021. Understanding an Evolving Pandemic: An Analysis of the Clinical Time Delay Distributions of COVID-19 in the United Kingdom Plos One. *Plos One.*

WHO, 2020. *Archived: WHO Timeline - COVID-19,* Geneva: s.n.

Wood, S., 2003. Thin plate regression splines. *J.R.Statist.Soc.B ,* 65(1), pp. 95-114.

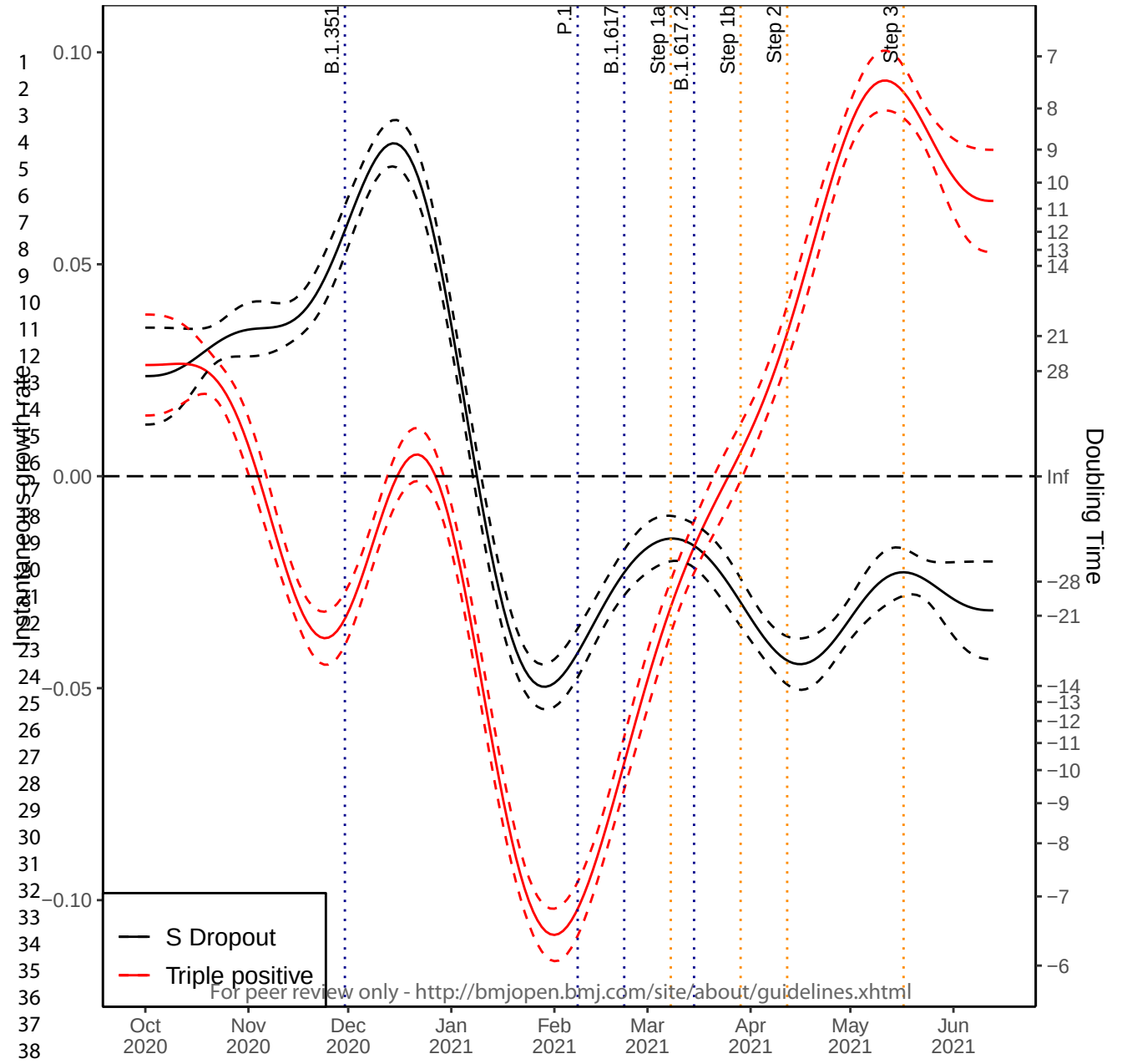
Wood, S., 2006. Generalized Additive Models: an introduction with R. *CRC.*

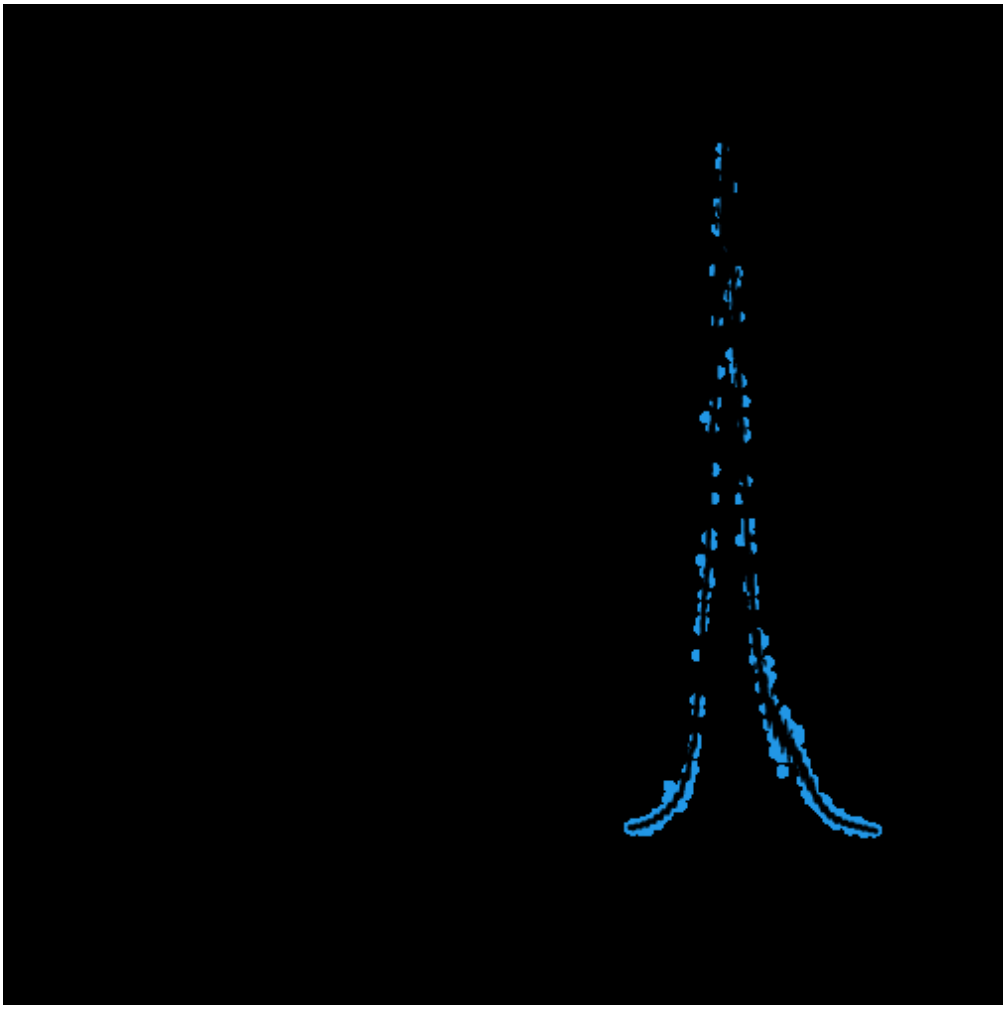
Wood, S., 2017. *Generalized Additive Models.* New York: Chapman and Hall/CRC.

Wood, S., 2018. Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL. *R package version https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/mgcv-package.html,* pp. 1-8.

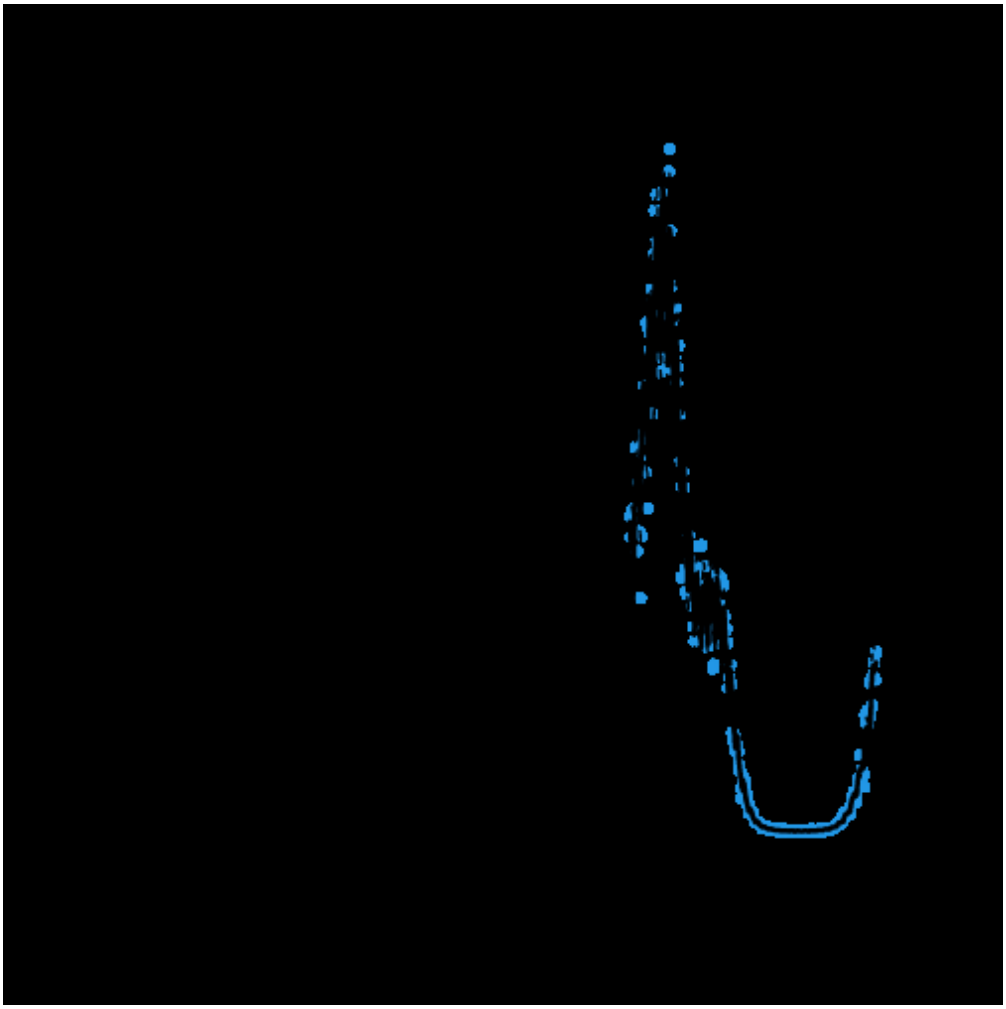
Yin, C., 2020. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics,* 112(5), pp. 3588-3596.

SARS−CoV−2 Confirmed Positive Cases : England

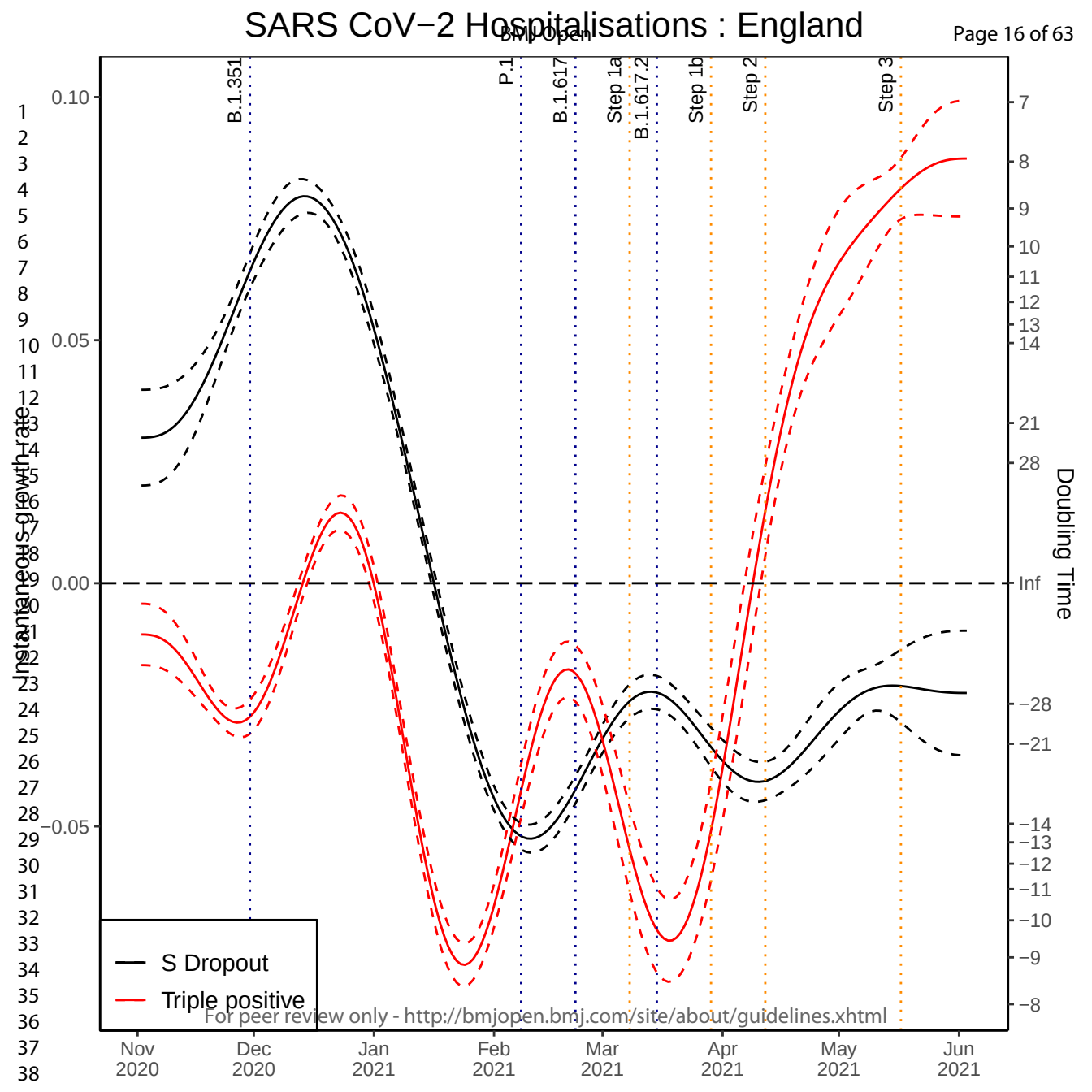


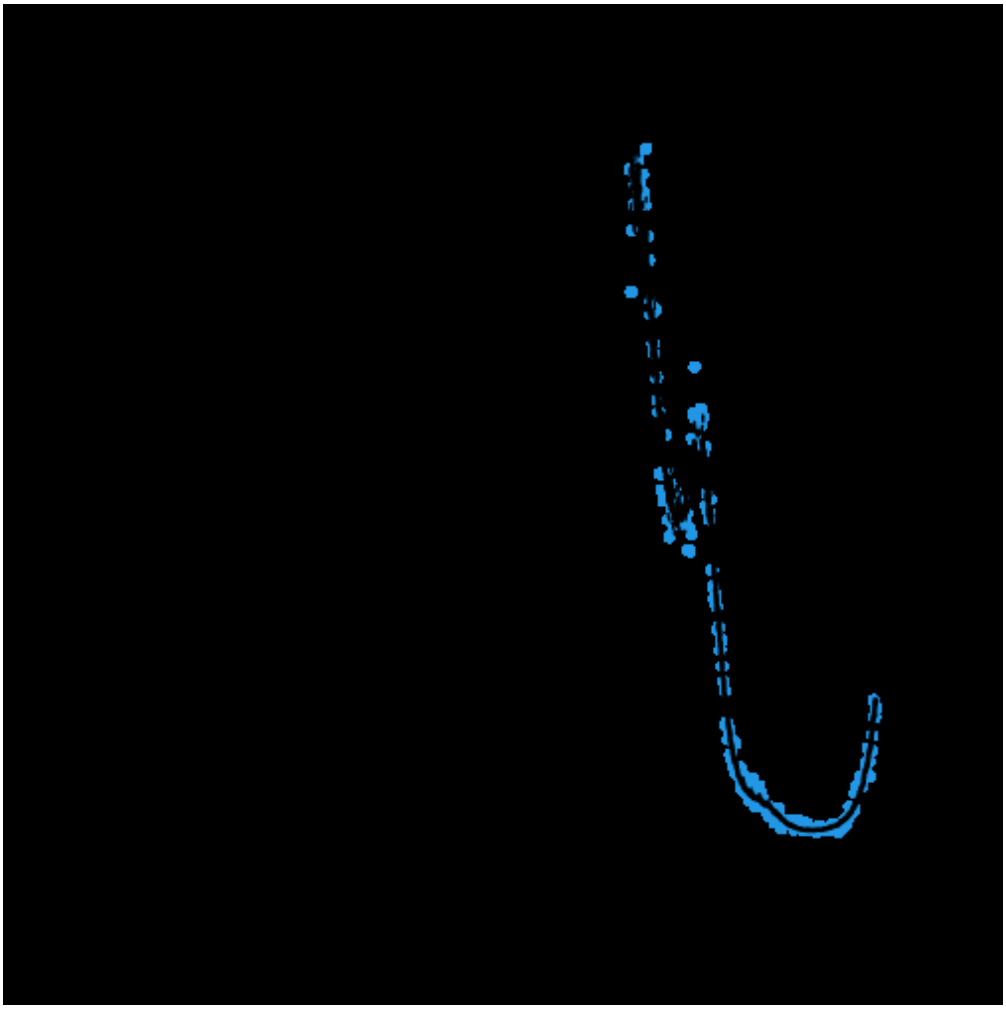


336x336mm (38 x 38 DPI)

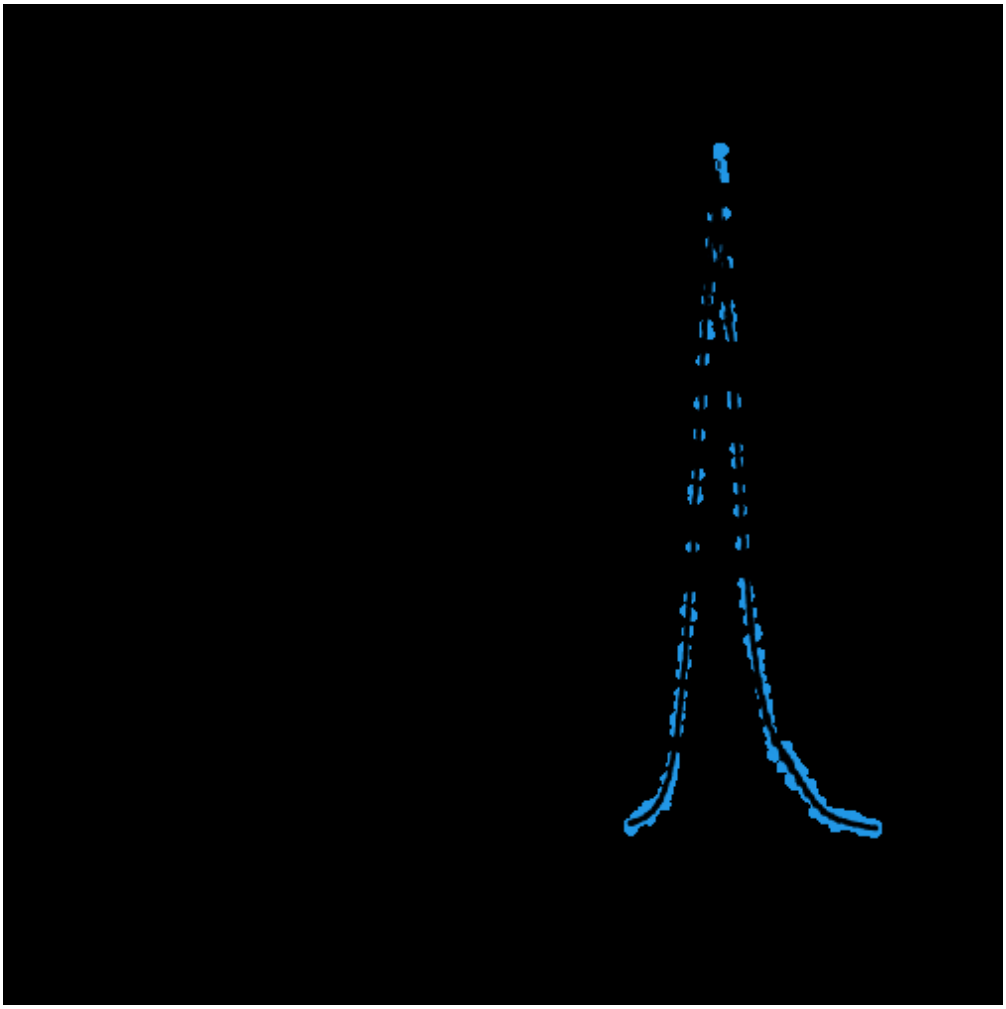


336x336mm (38 x 38 DPI)



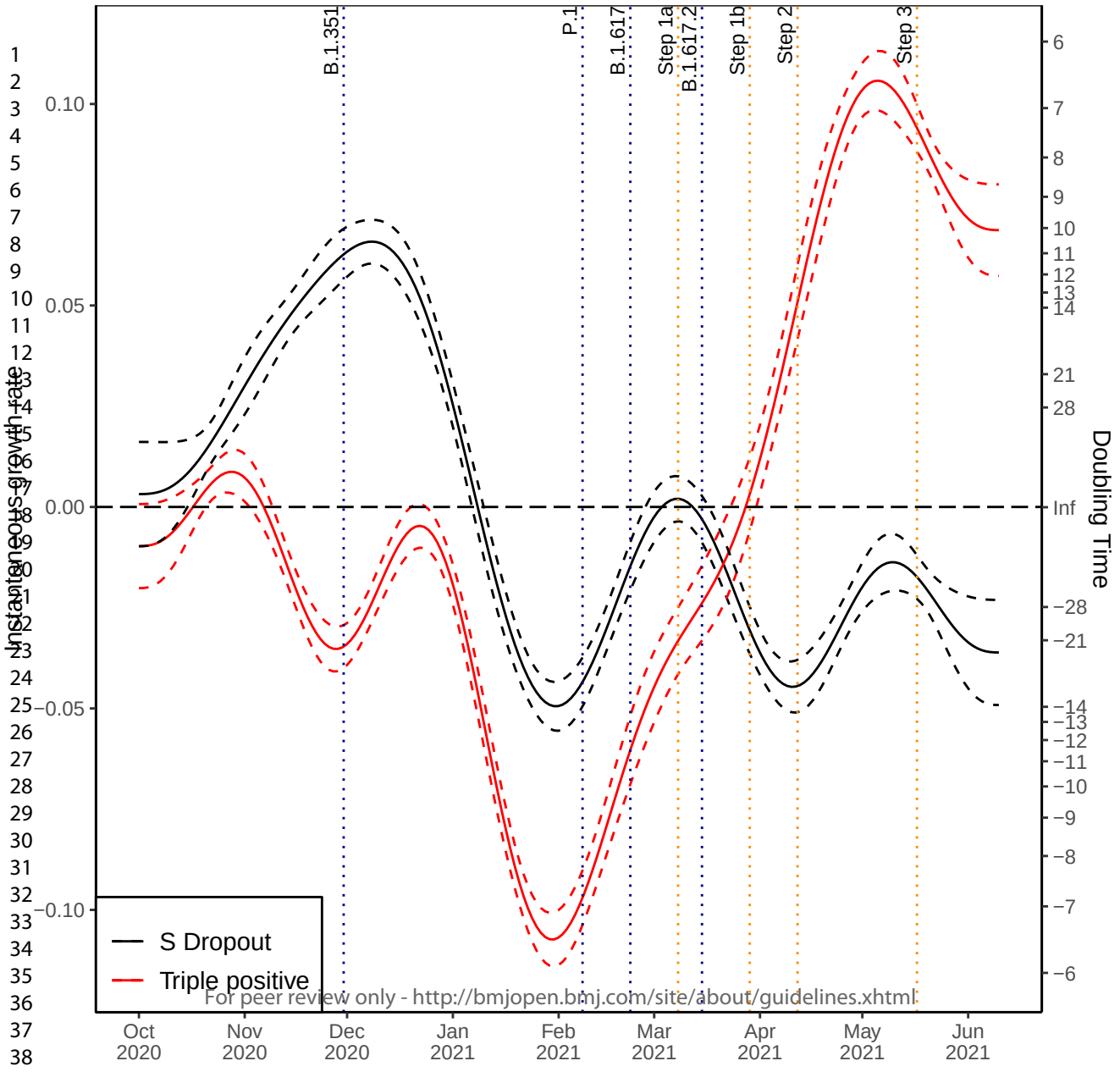


336x336mm (38 x 38 DPI)

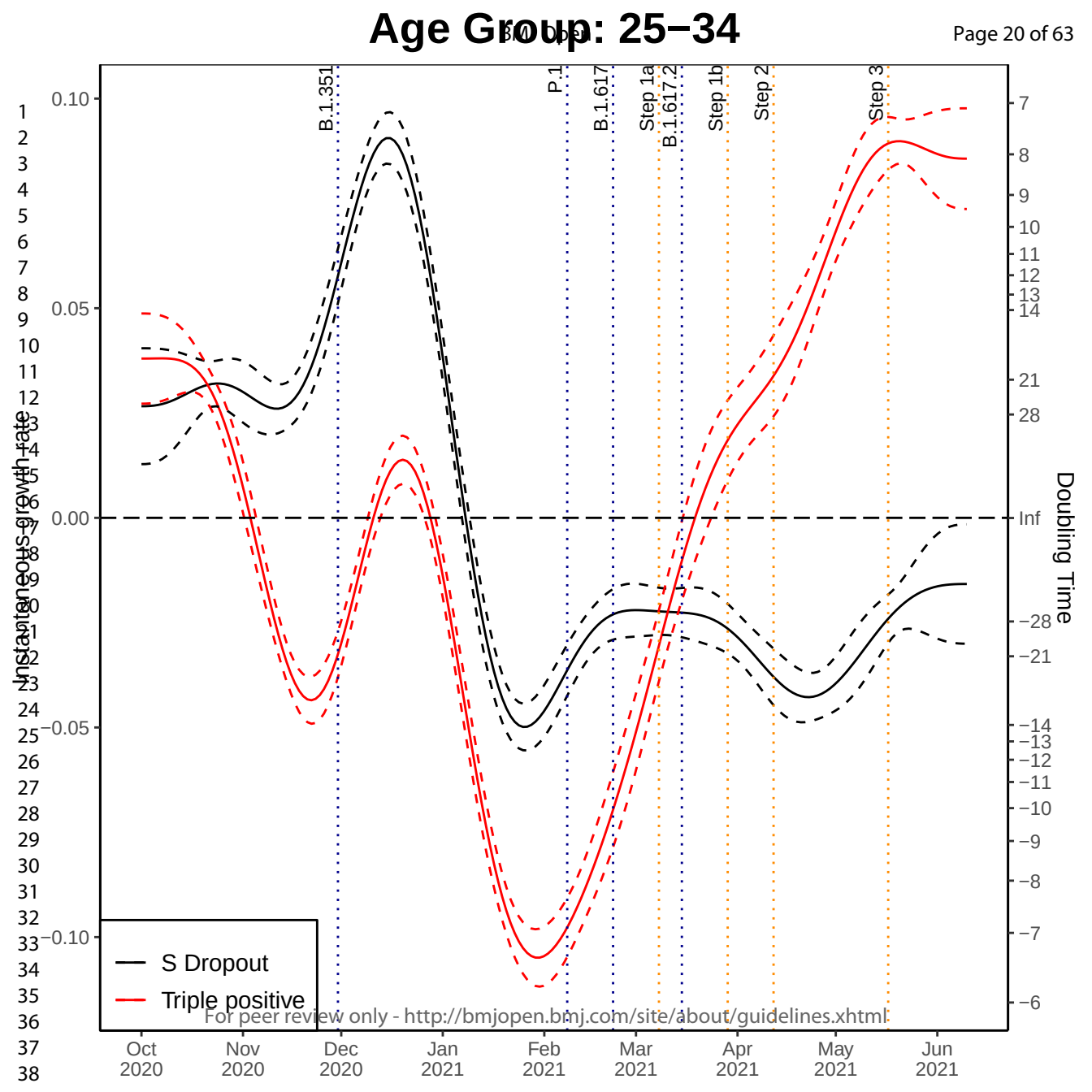


336x336mm (38 x 38 DPI)

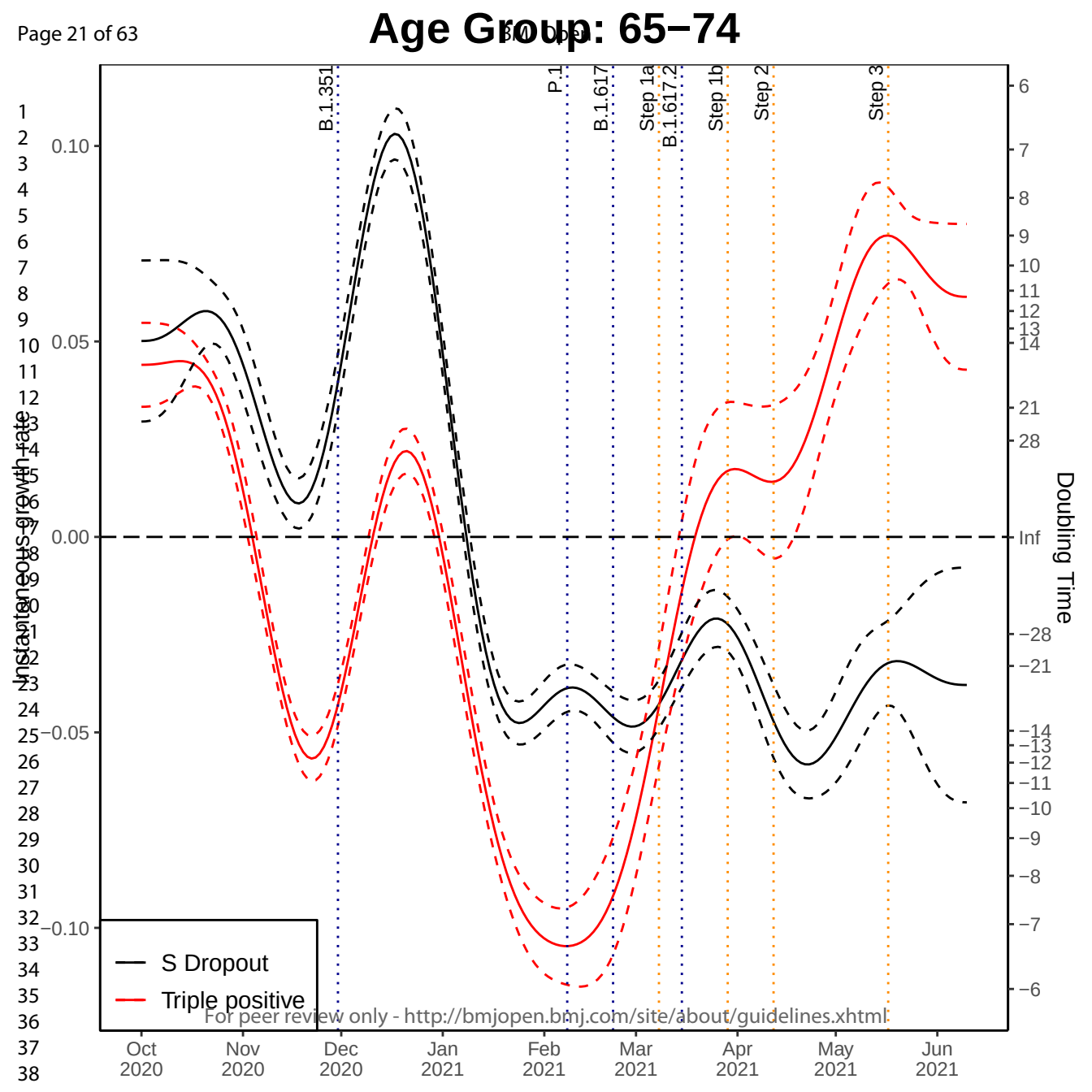
Age Group: 0–24



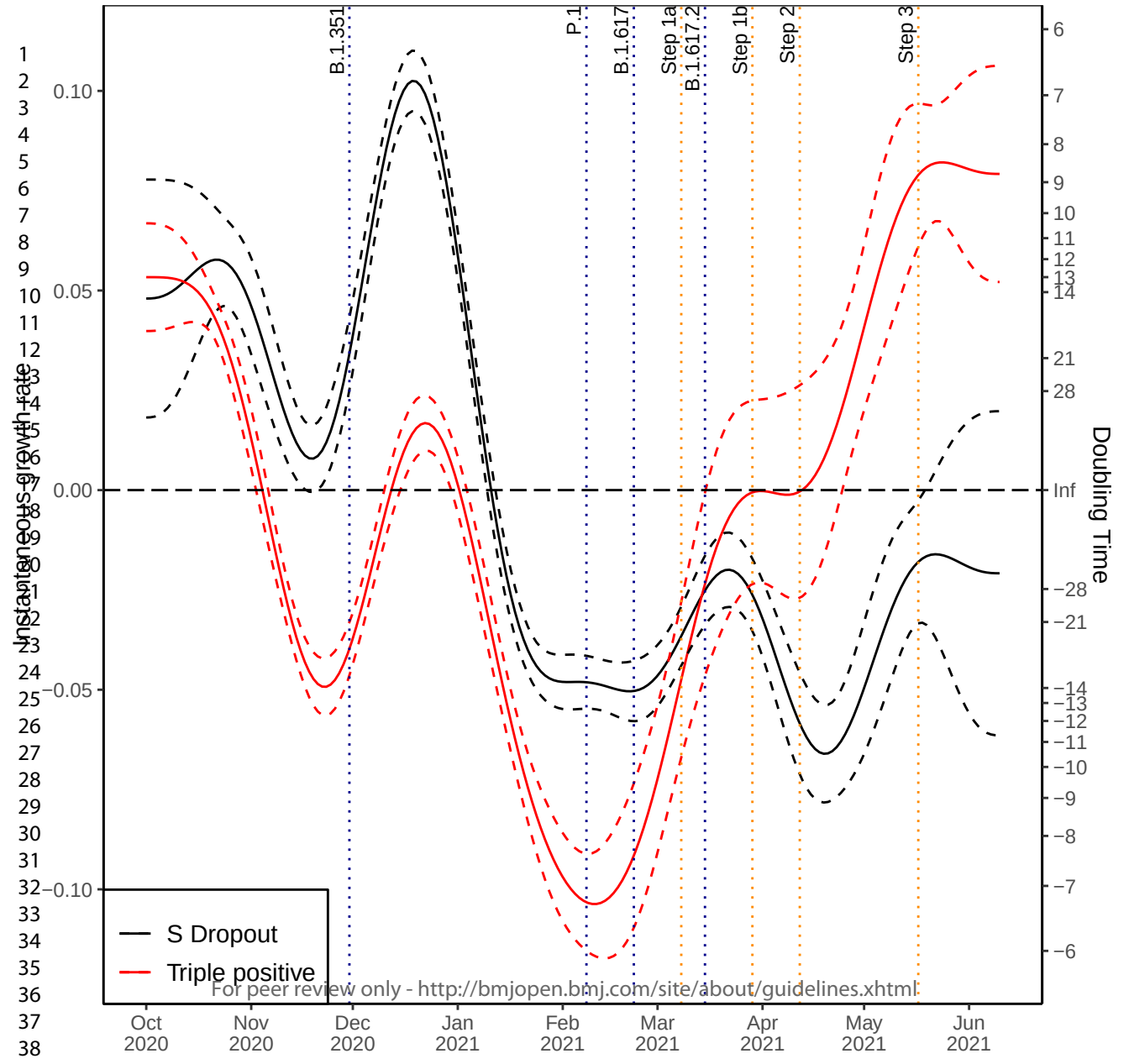
Age Group: 25−34



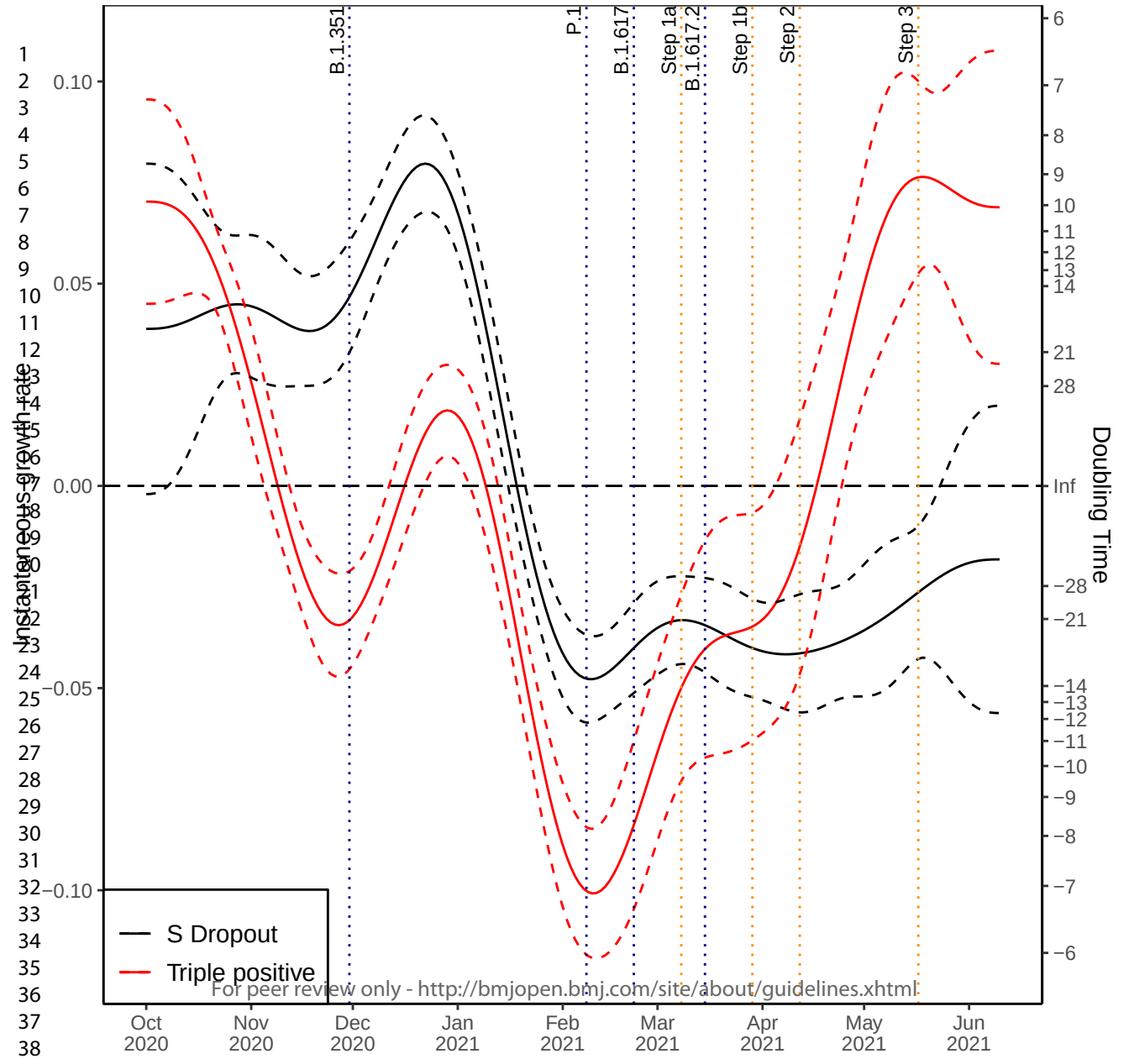
Age Group: 65–74



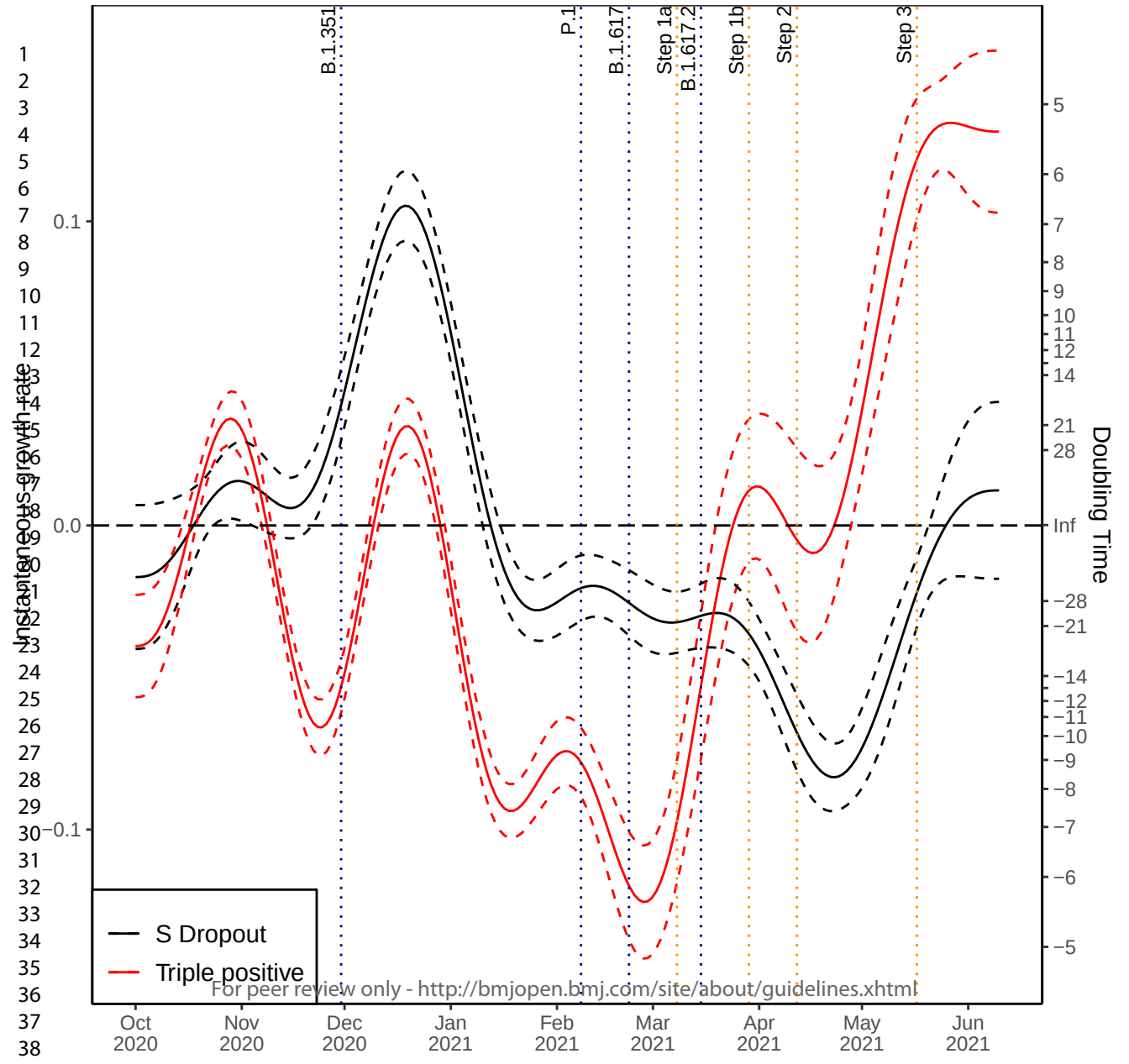
Age Group: 75−84



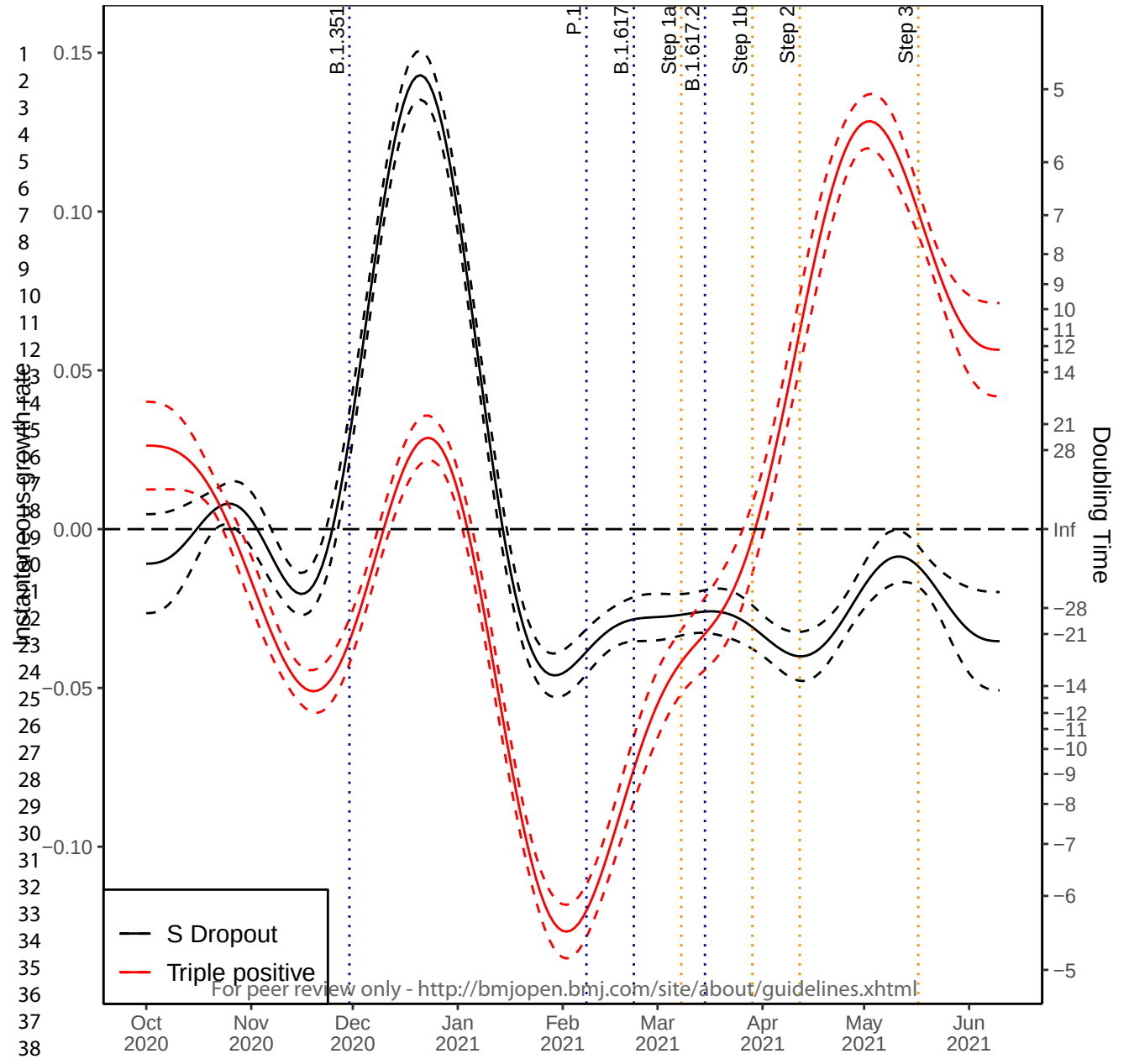
Age Group: 85_plus



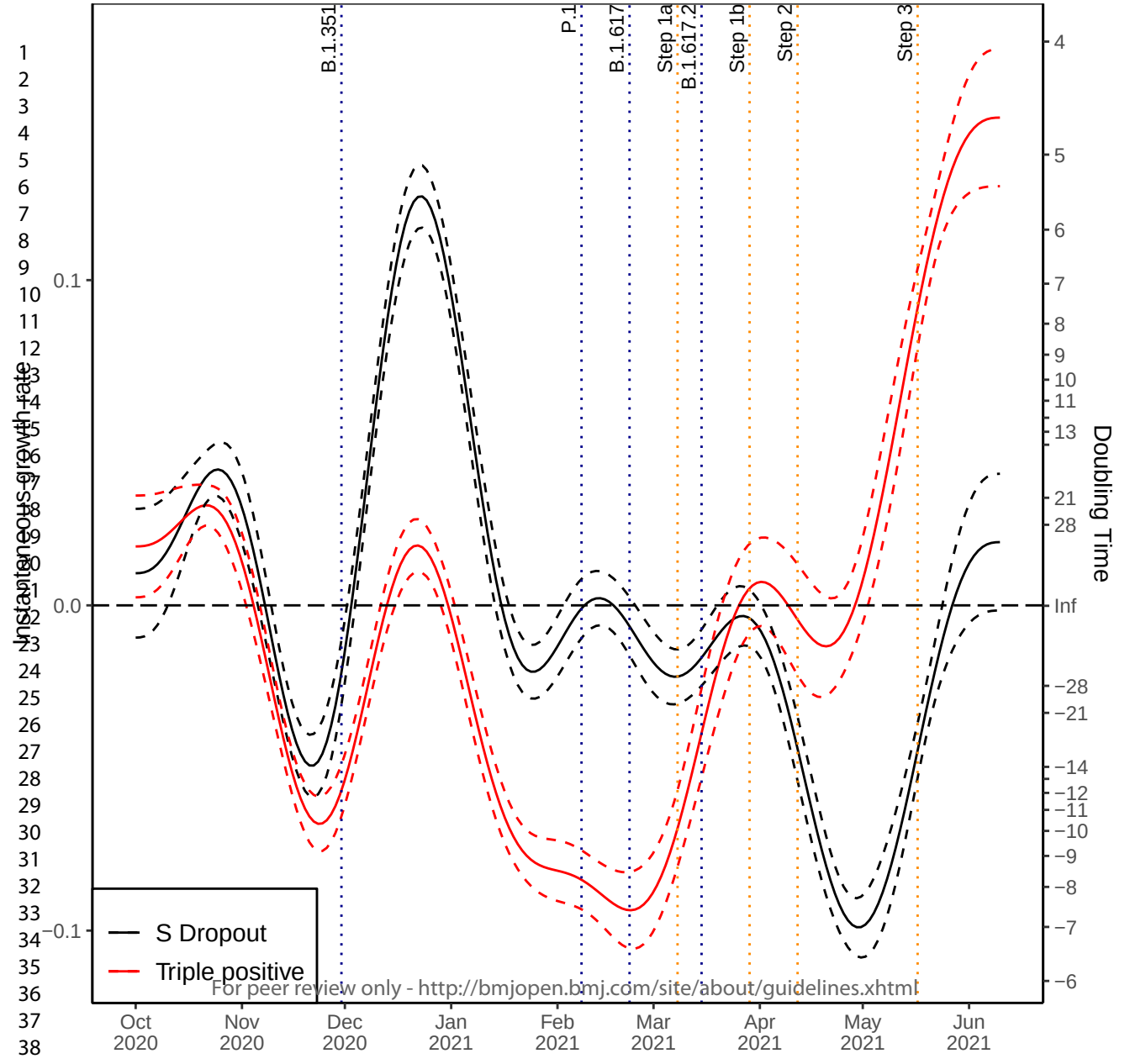
Region: North_East



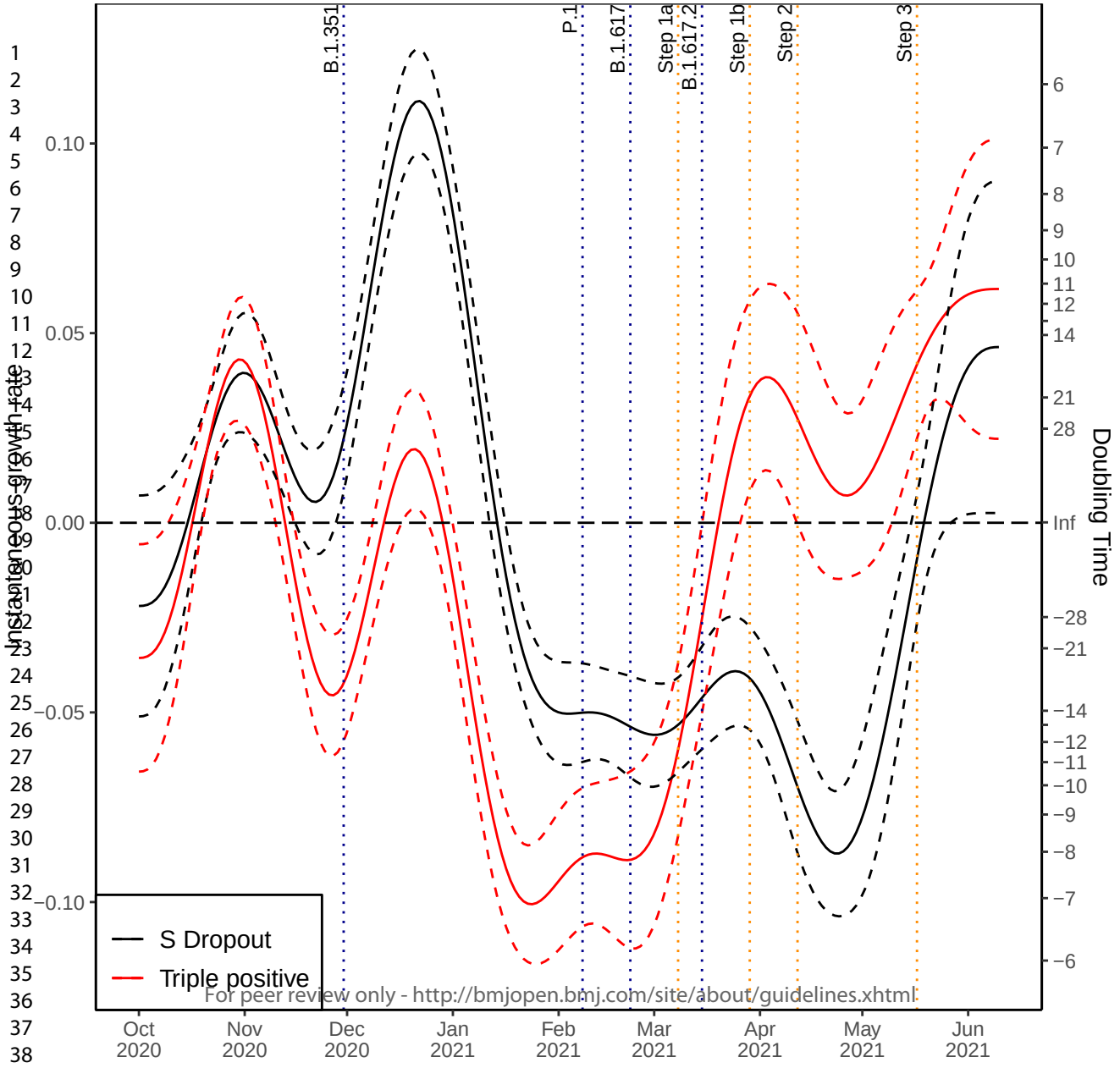
Region: North_West



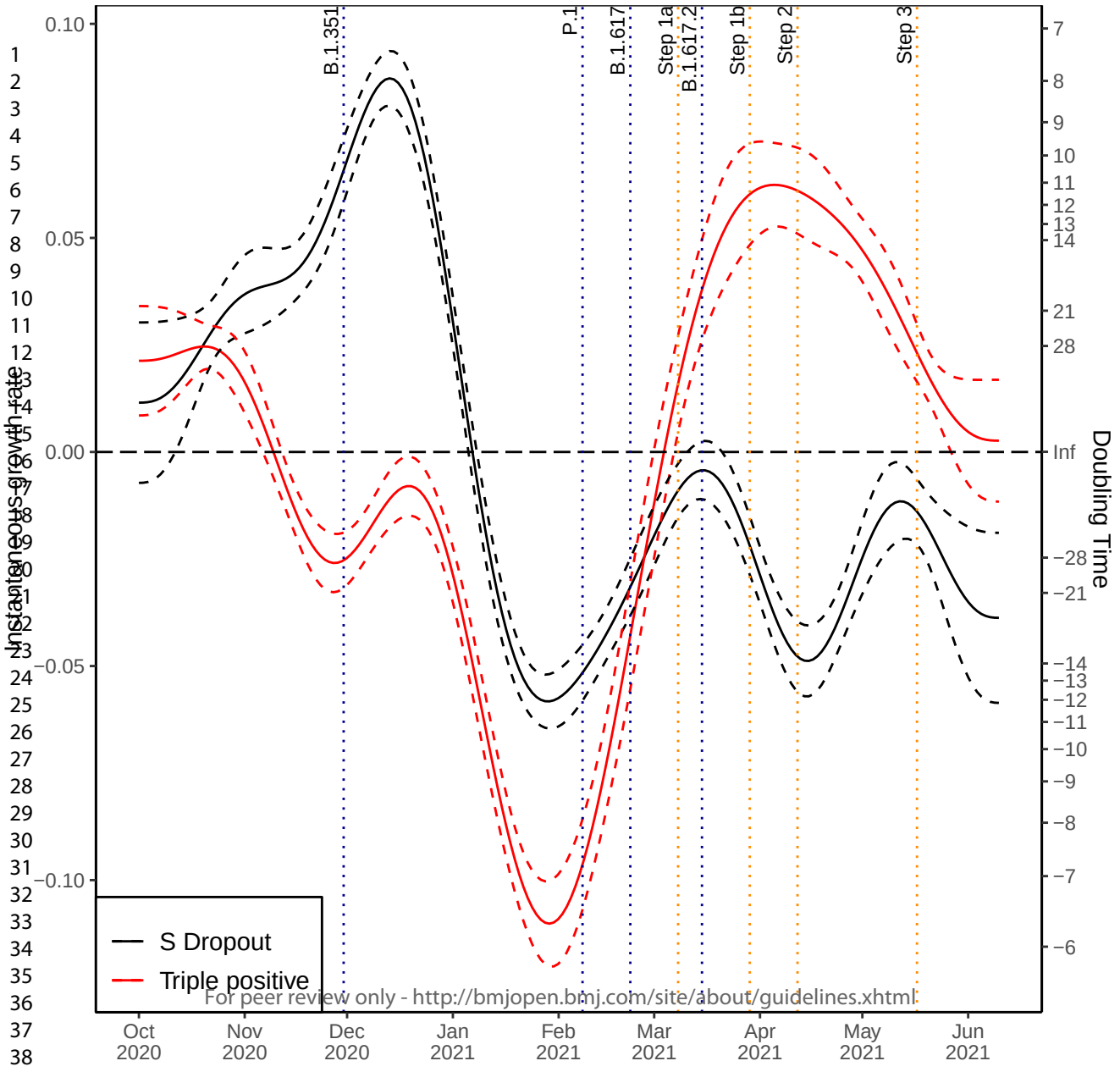
Region: Yorkshire_and_The_Humber



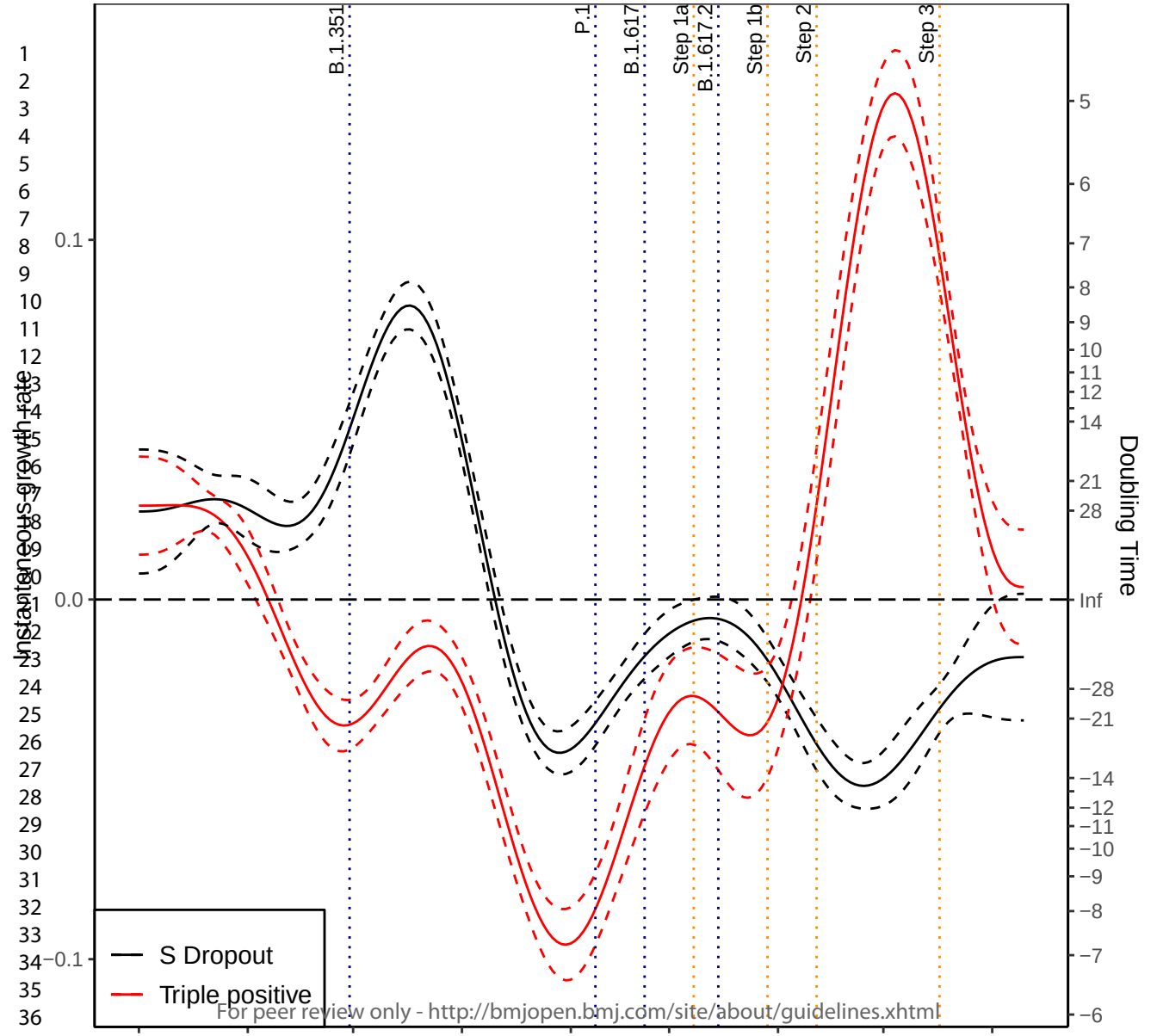
Region: South_West



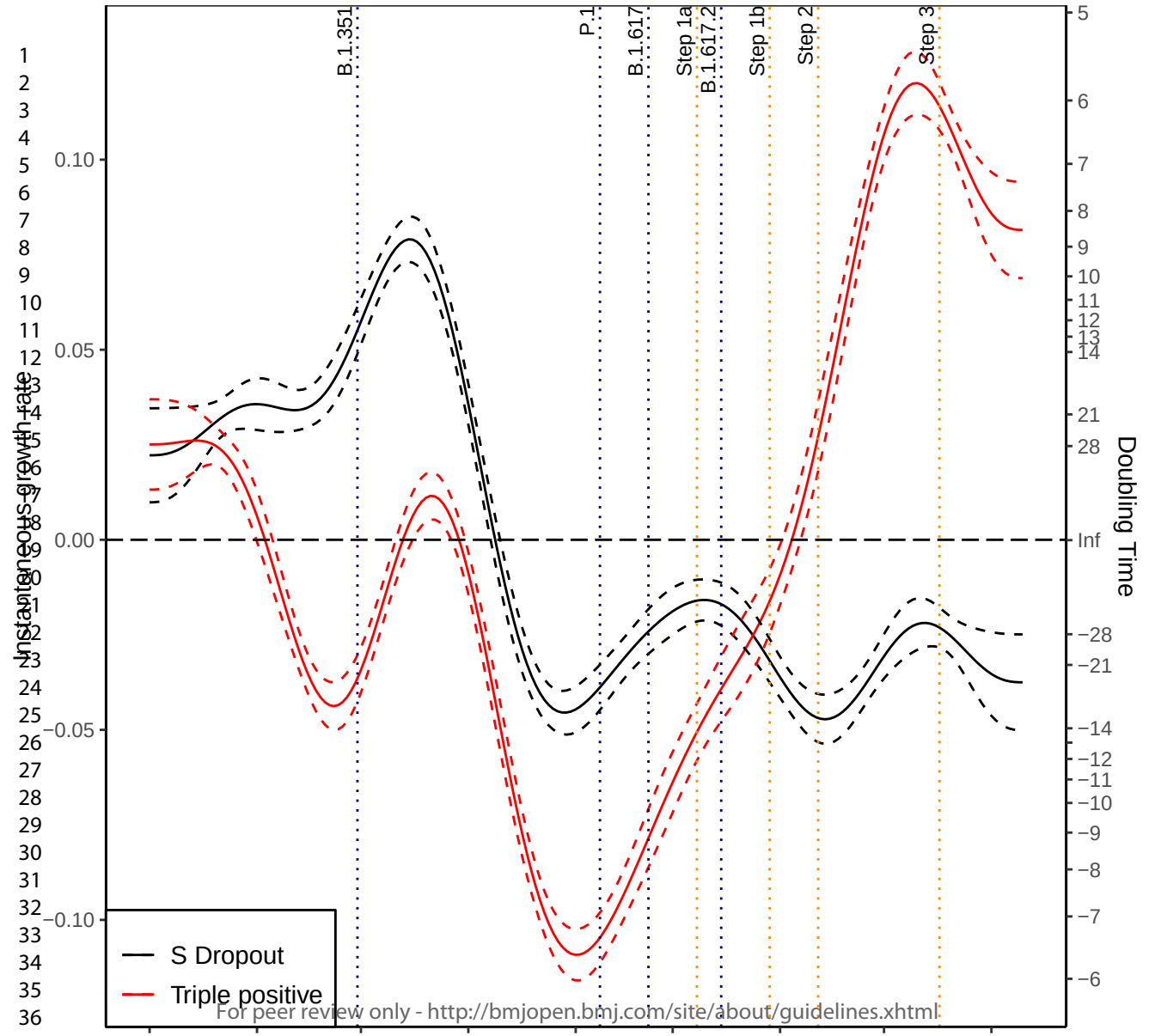
Ethnicity: Indian



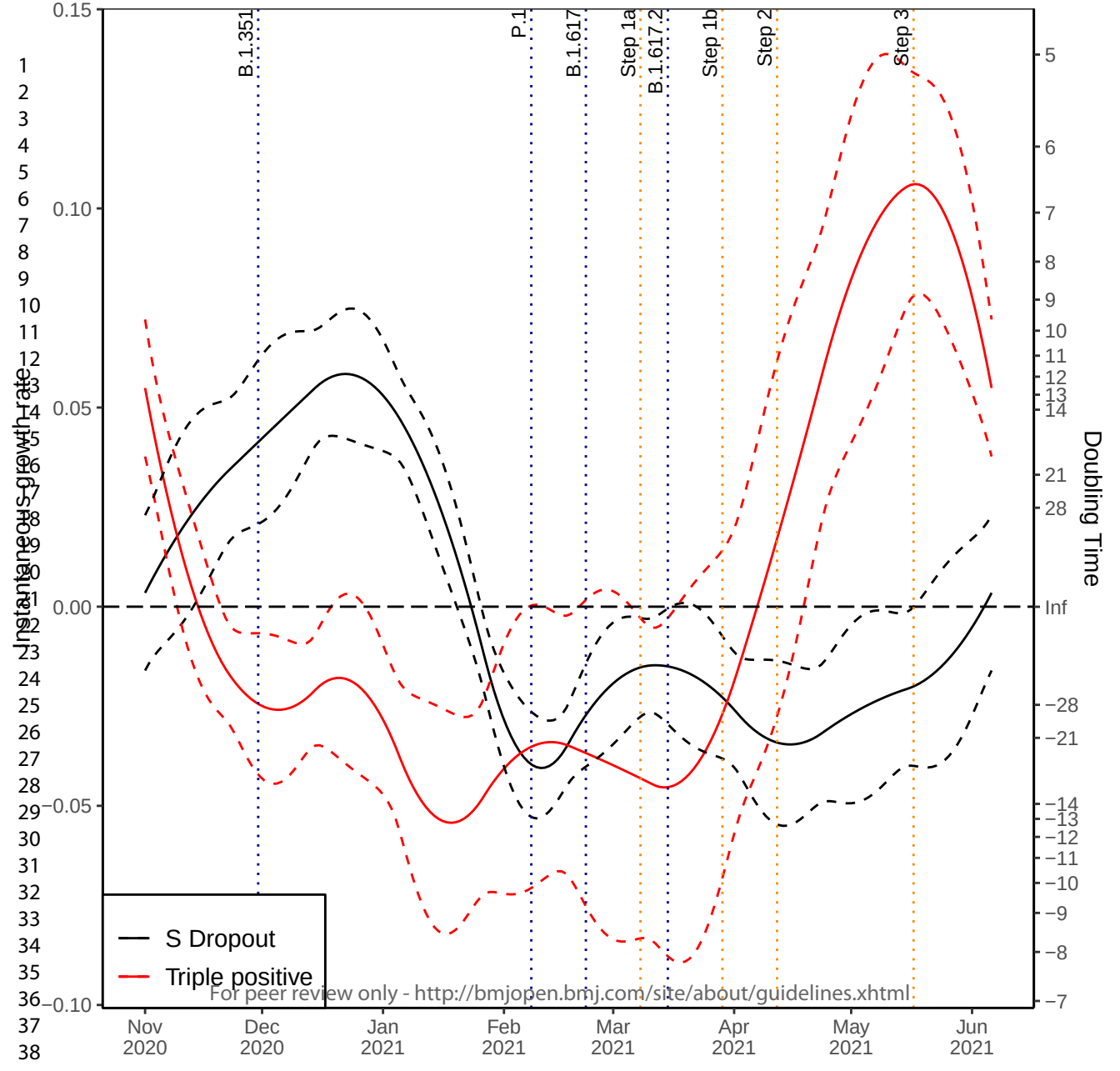
Ethnicity: Pakistani



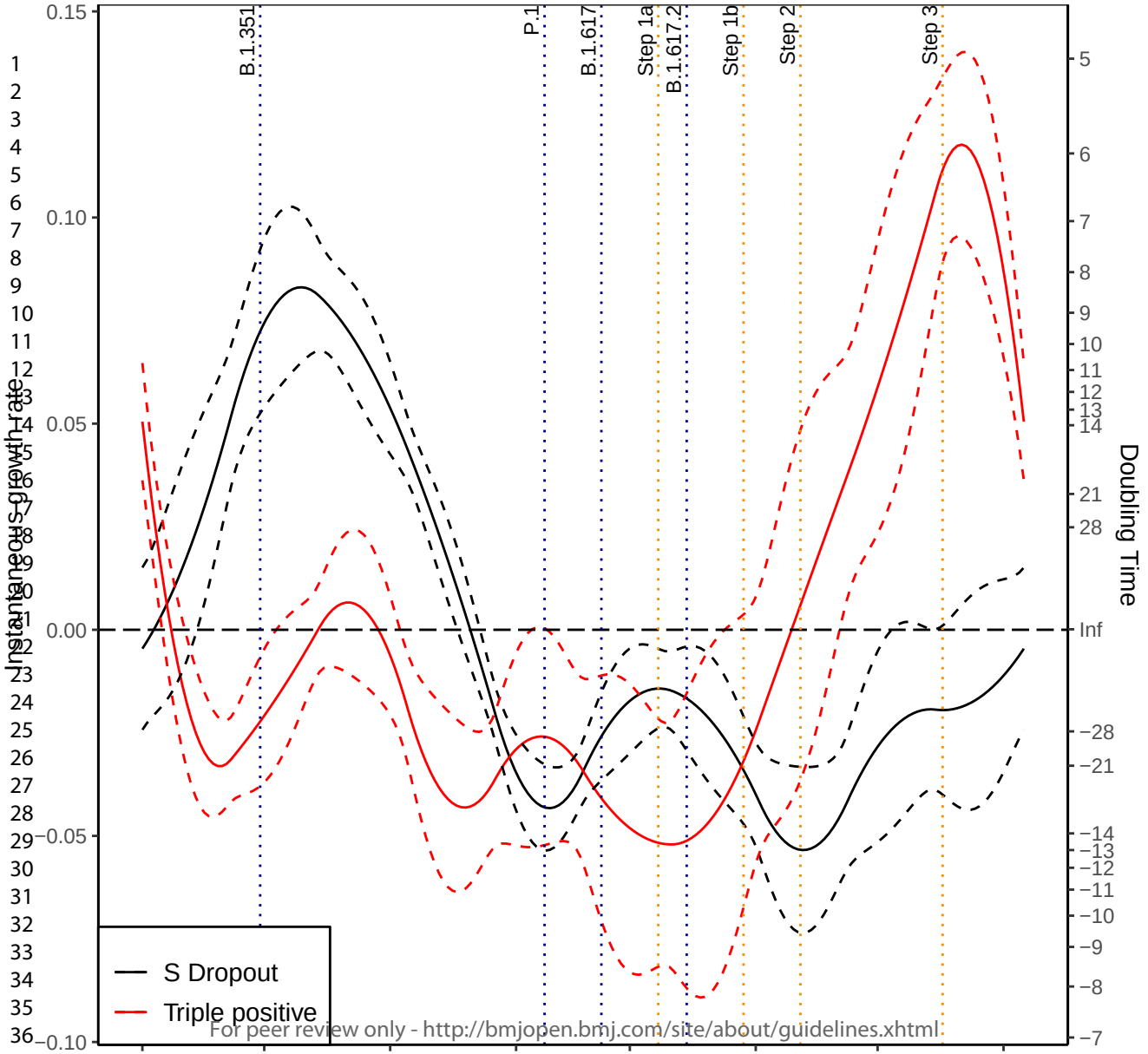
Ethnicity: White



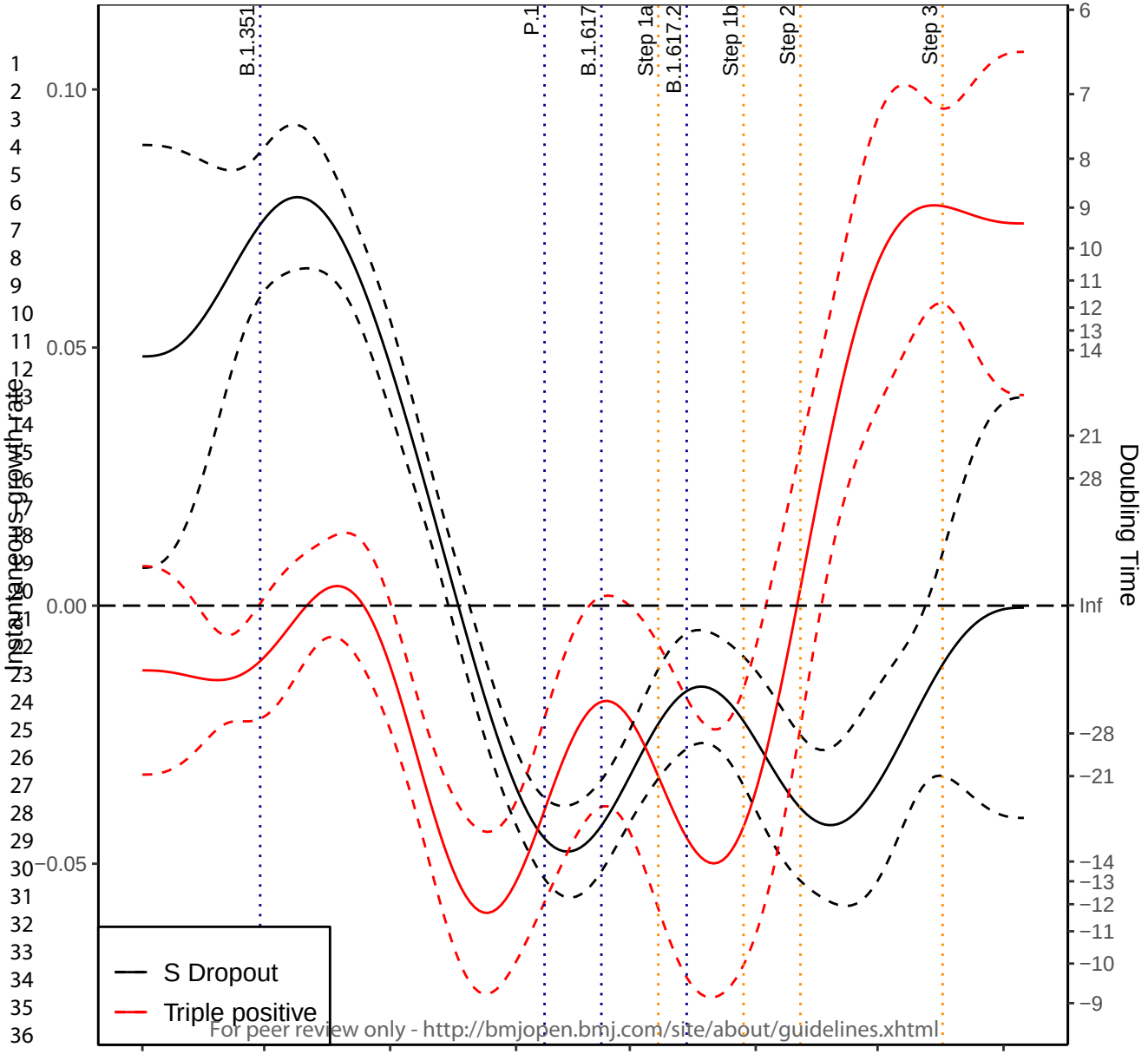
BMJ Open



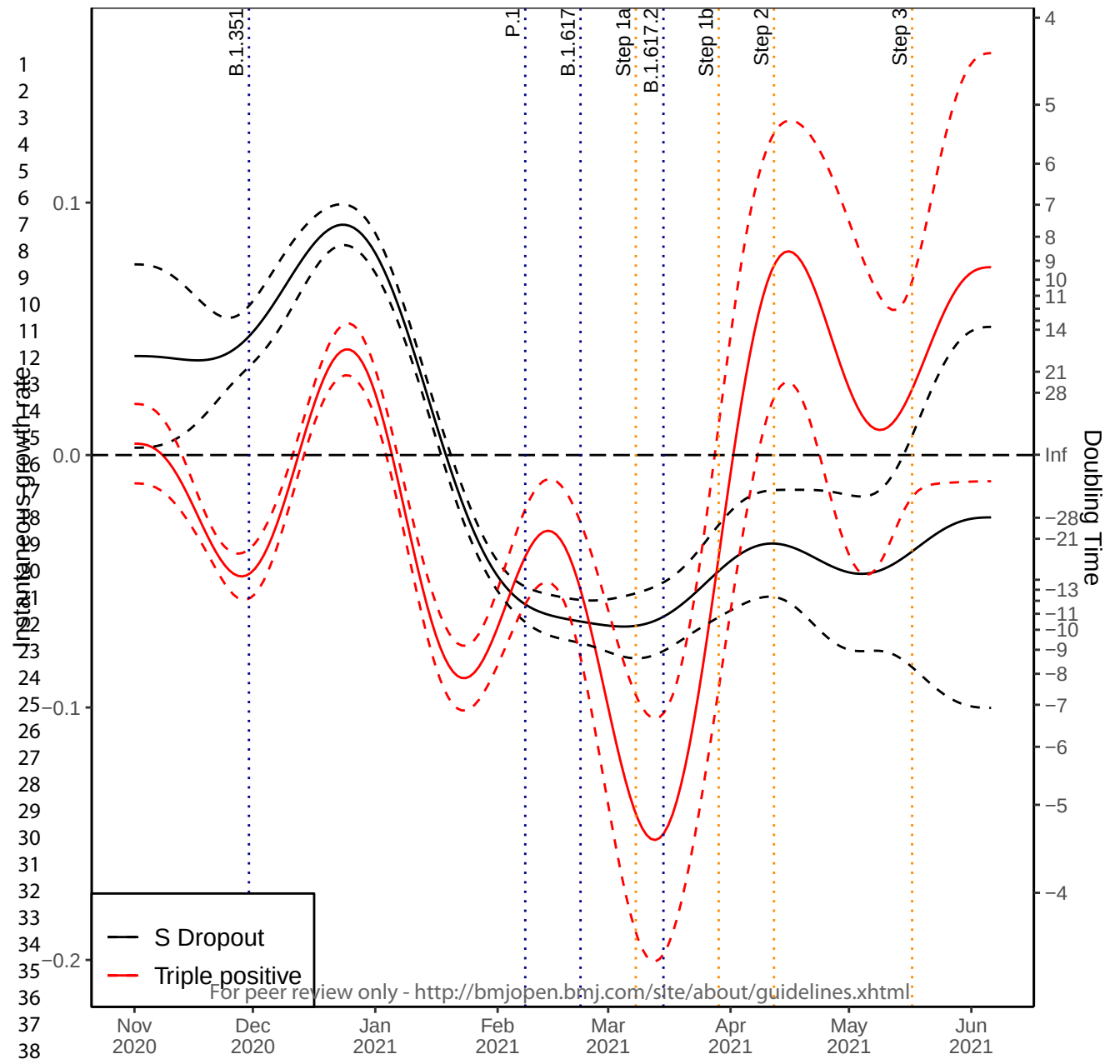
Hospitalisations Age Group: 25−34



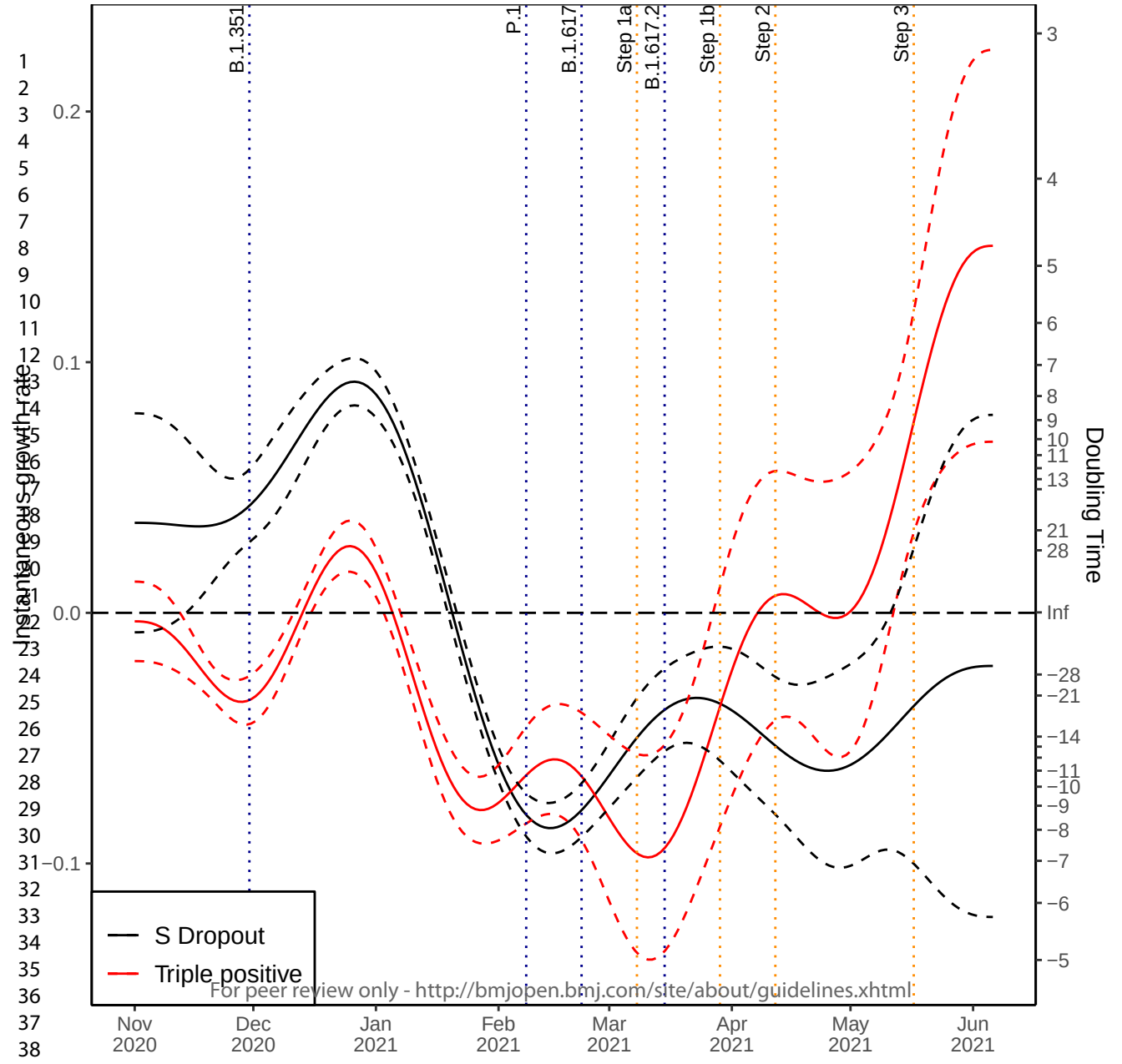
Hospitalisations Age Group: 35−44

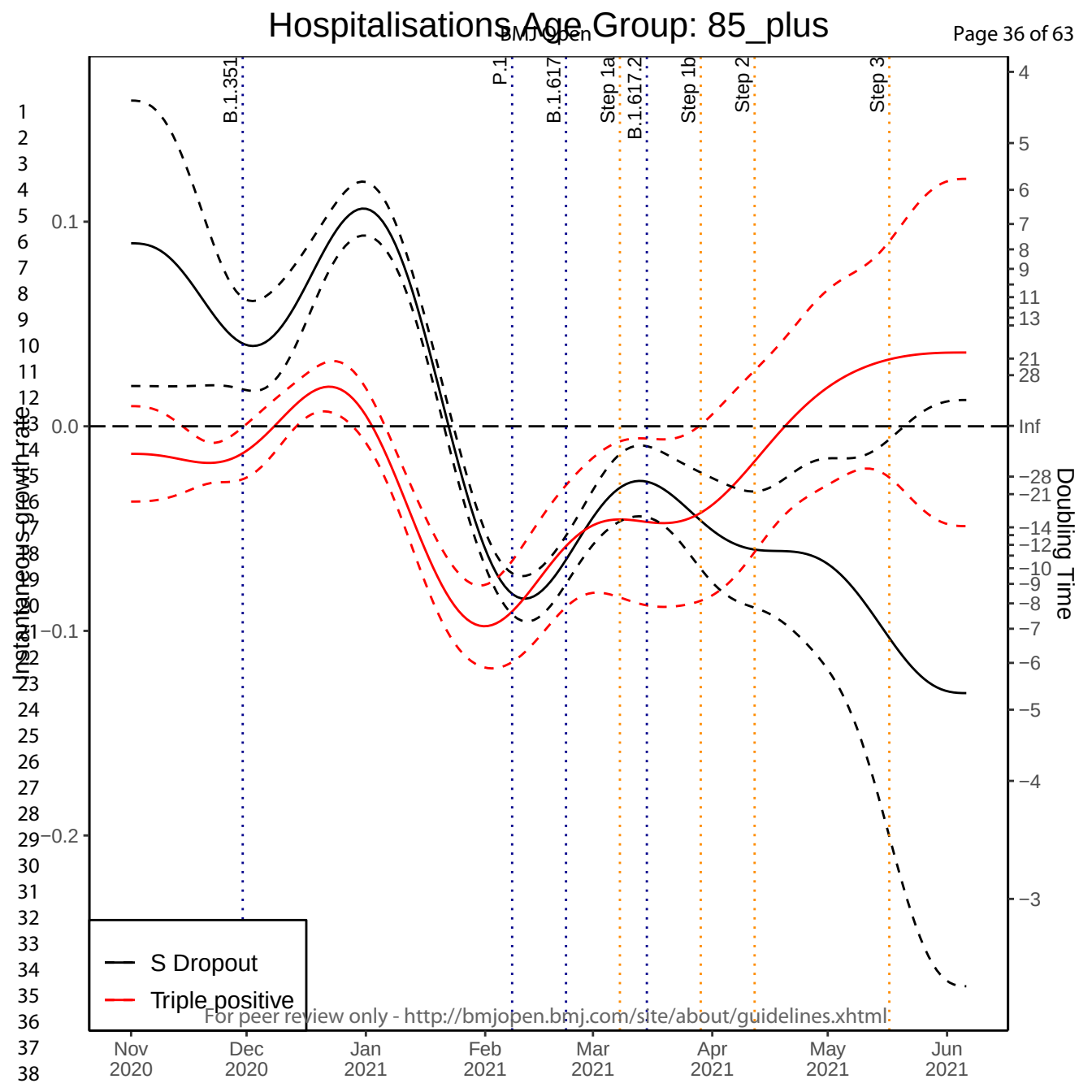


Hospitalisations Age Group: 65−74

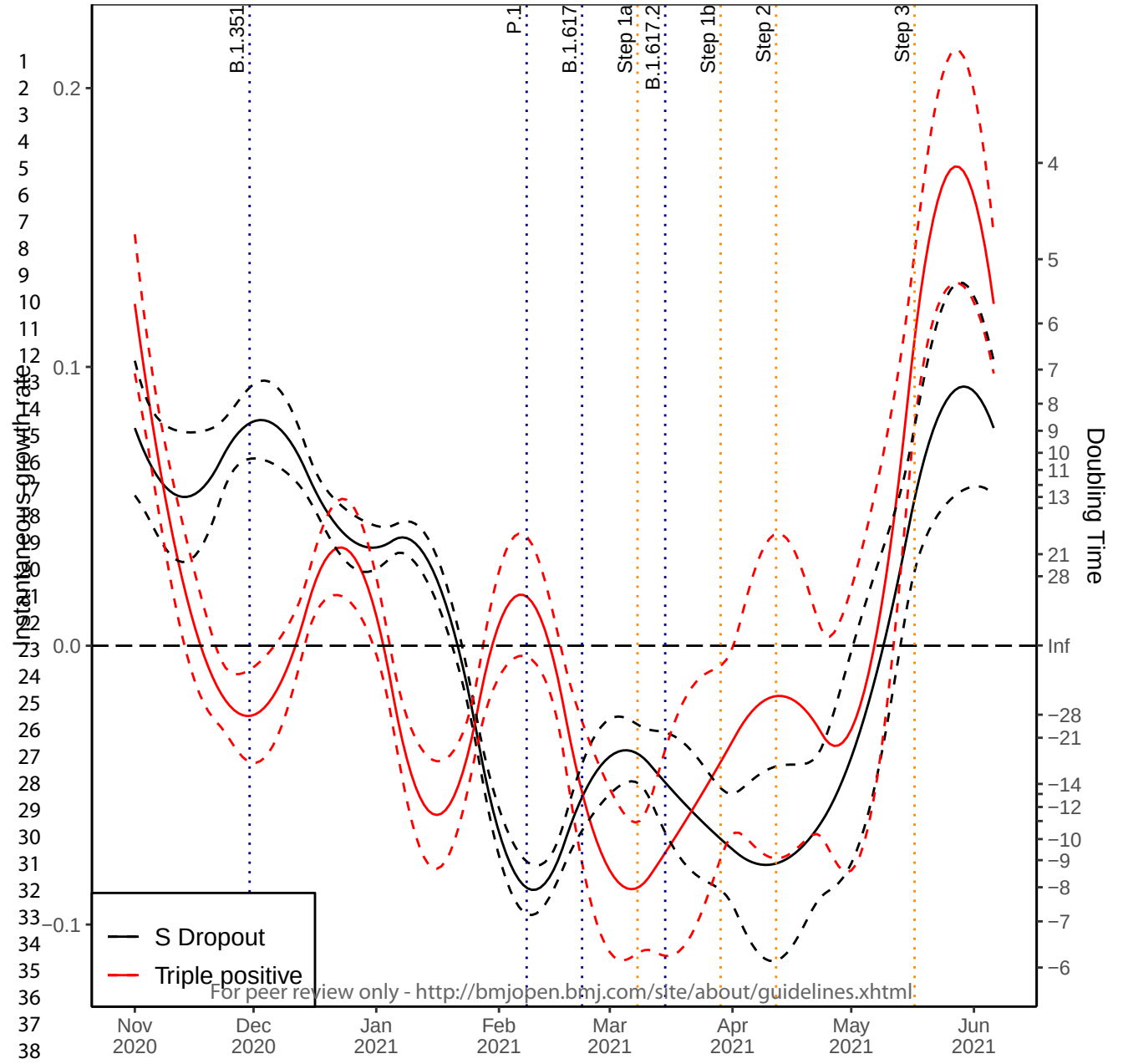


Hospitalisations Age Group: 75−84

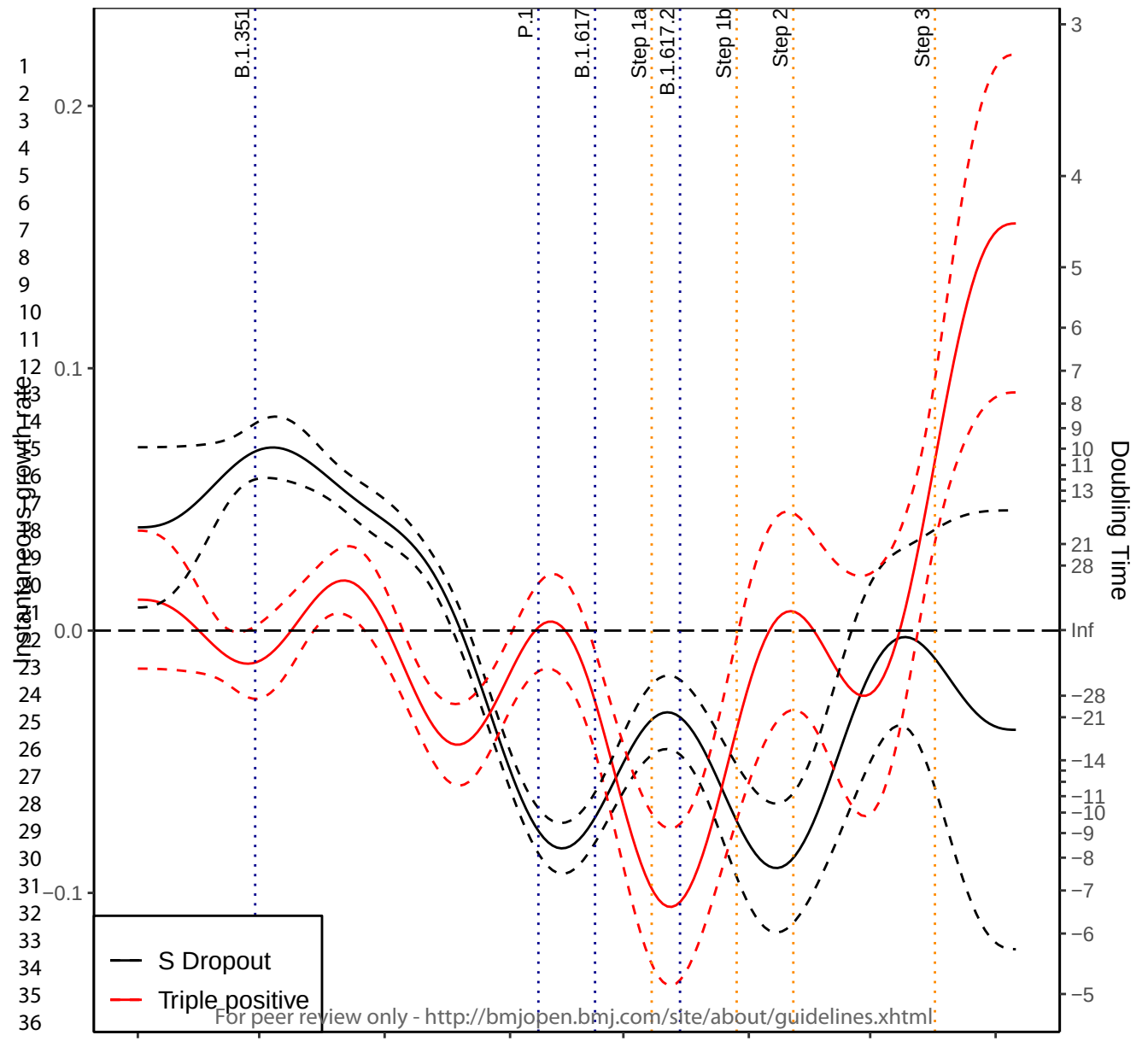




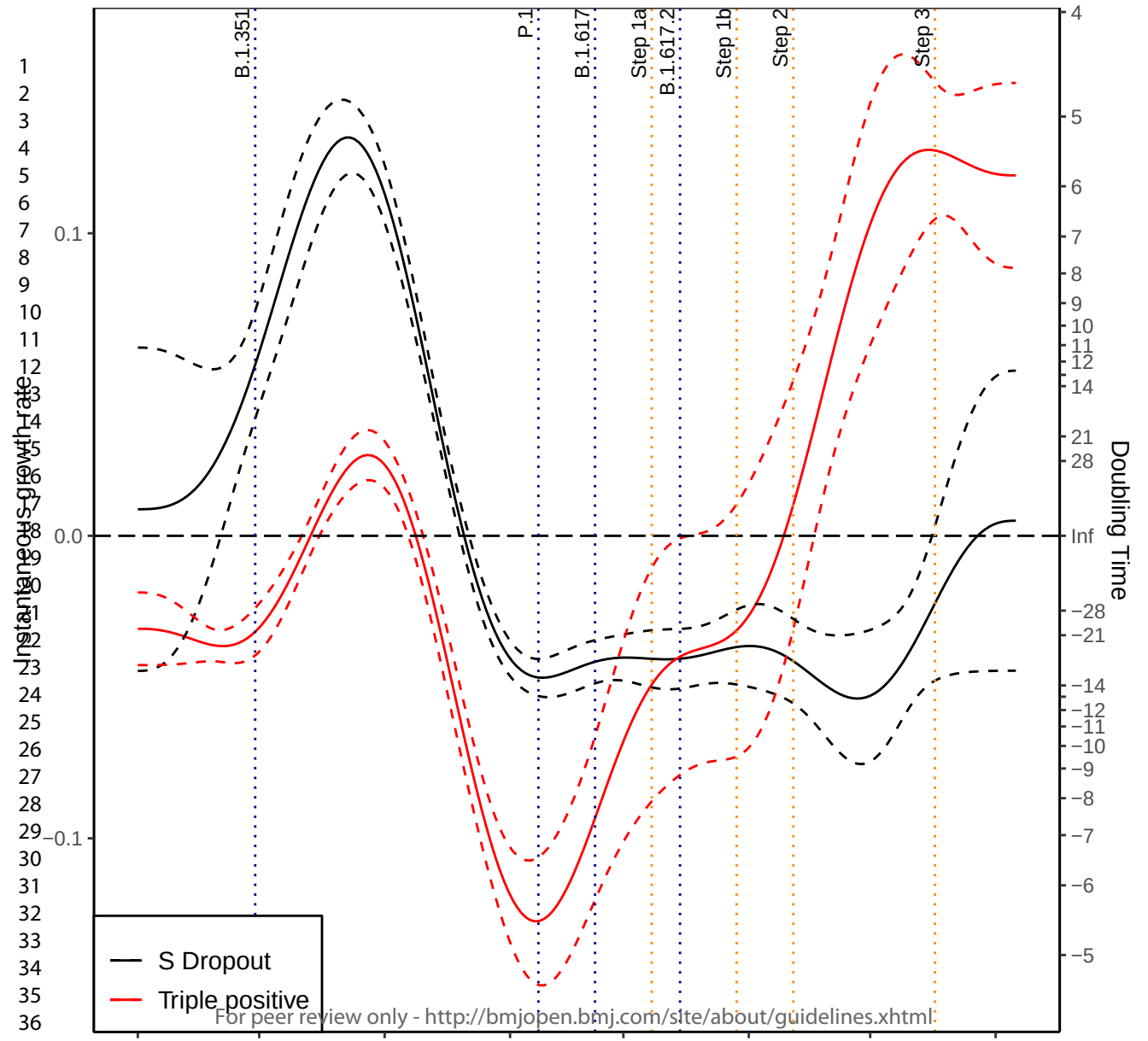
Hospitalisations Region Name: South_East

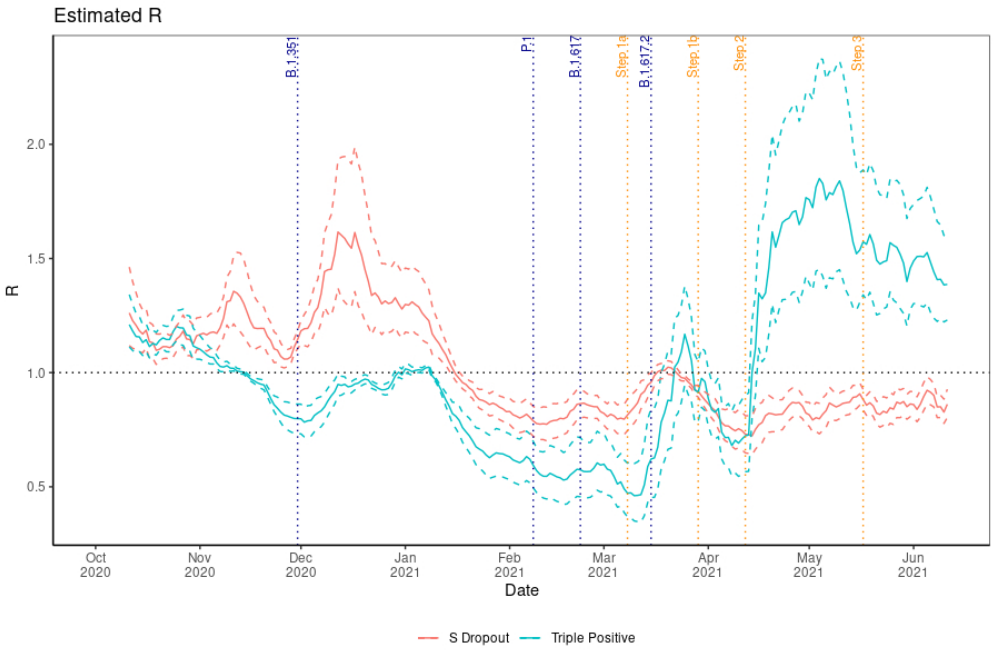


Hospitalisations Region Name: South_East



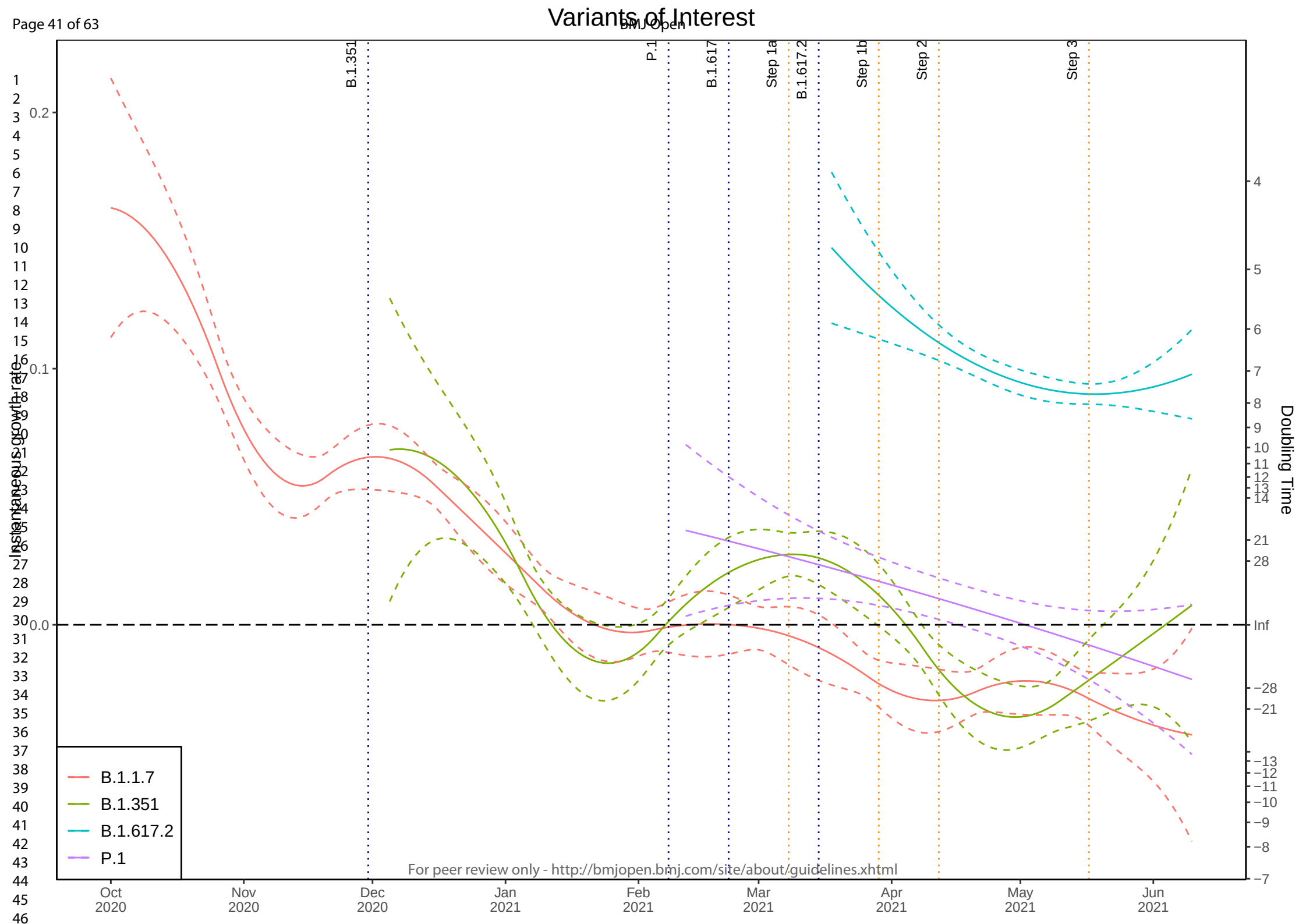
Hospitalisations Region Name: North_West



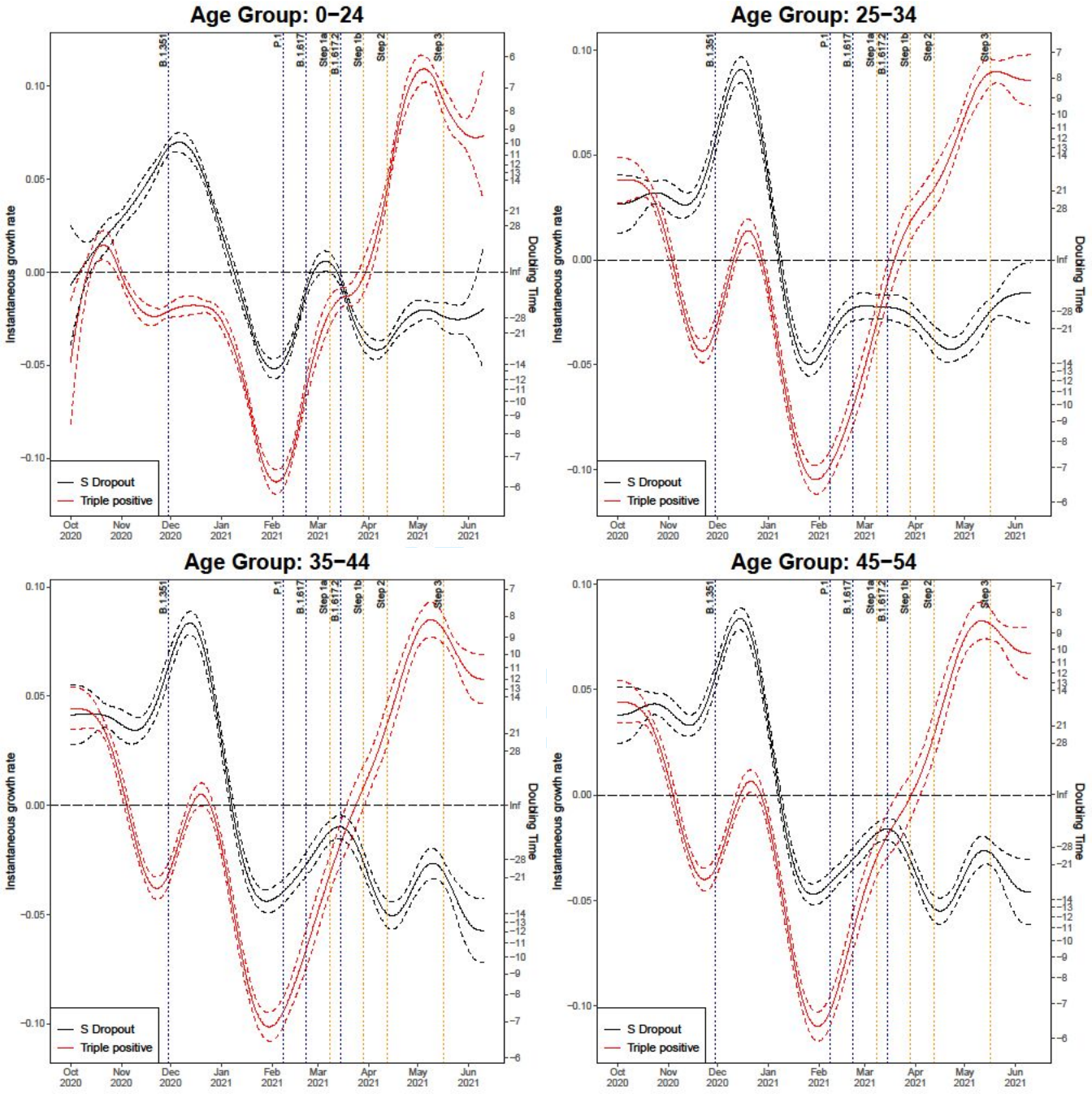


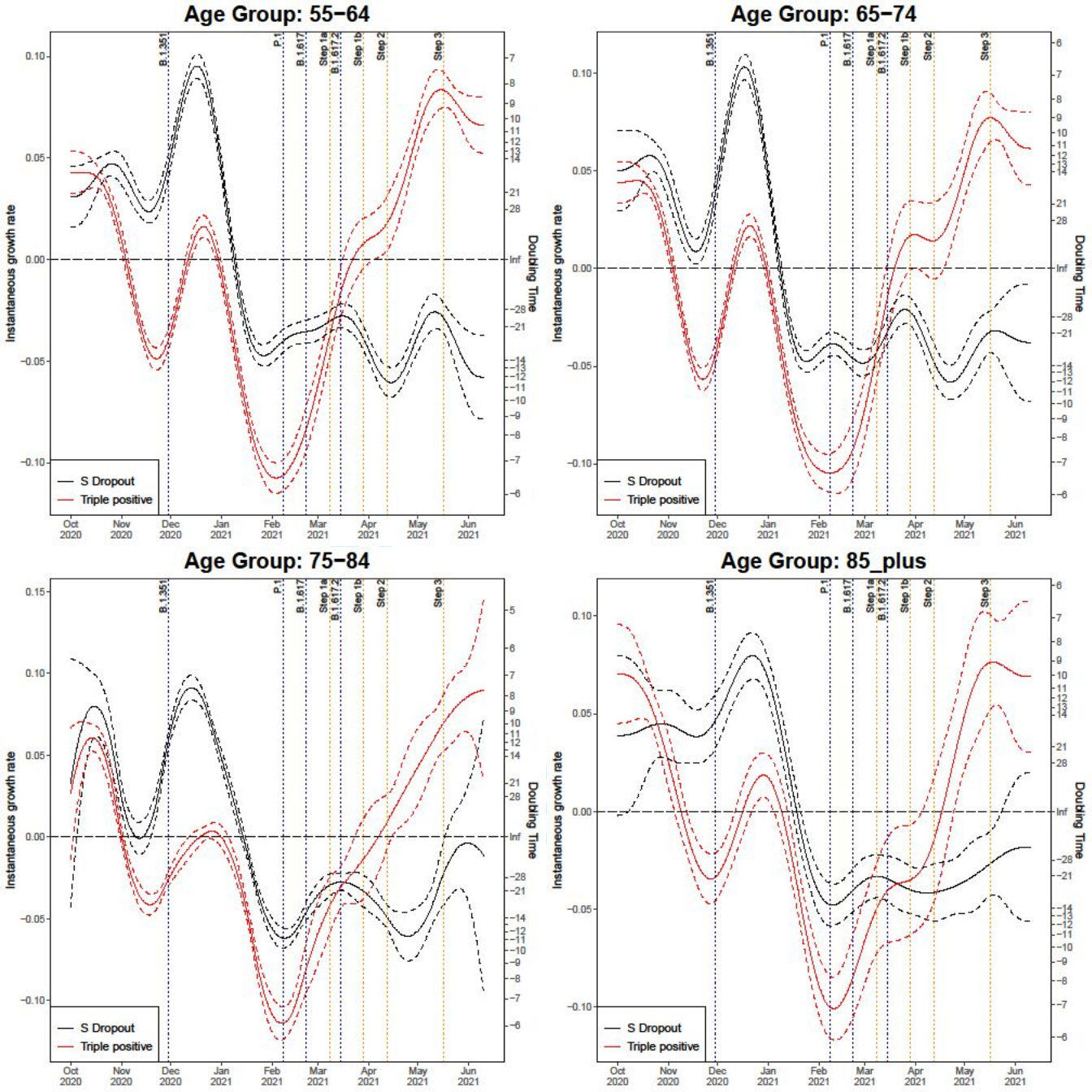
601x401mm (38 x 38 DPI)

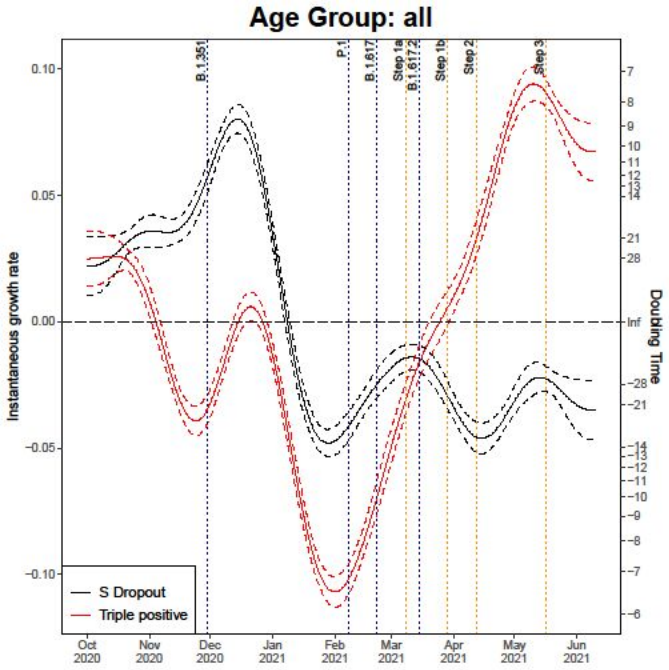
Variants of Interest



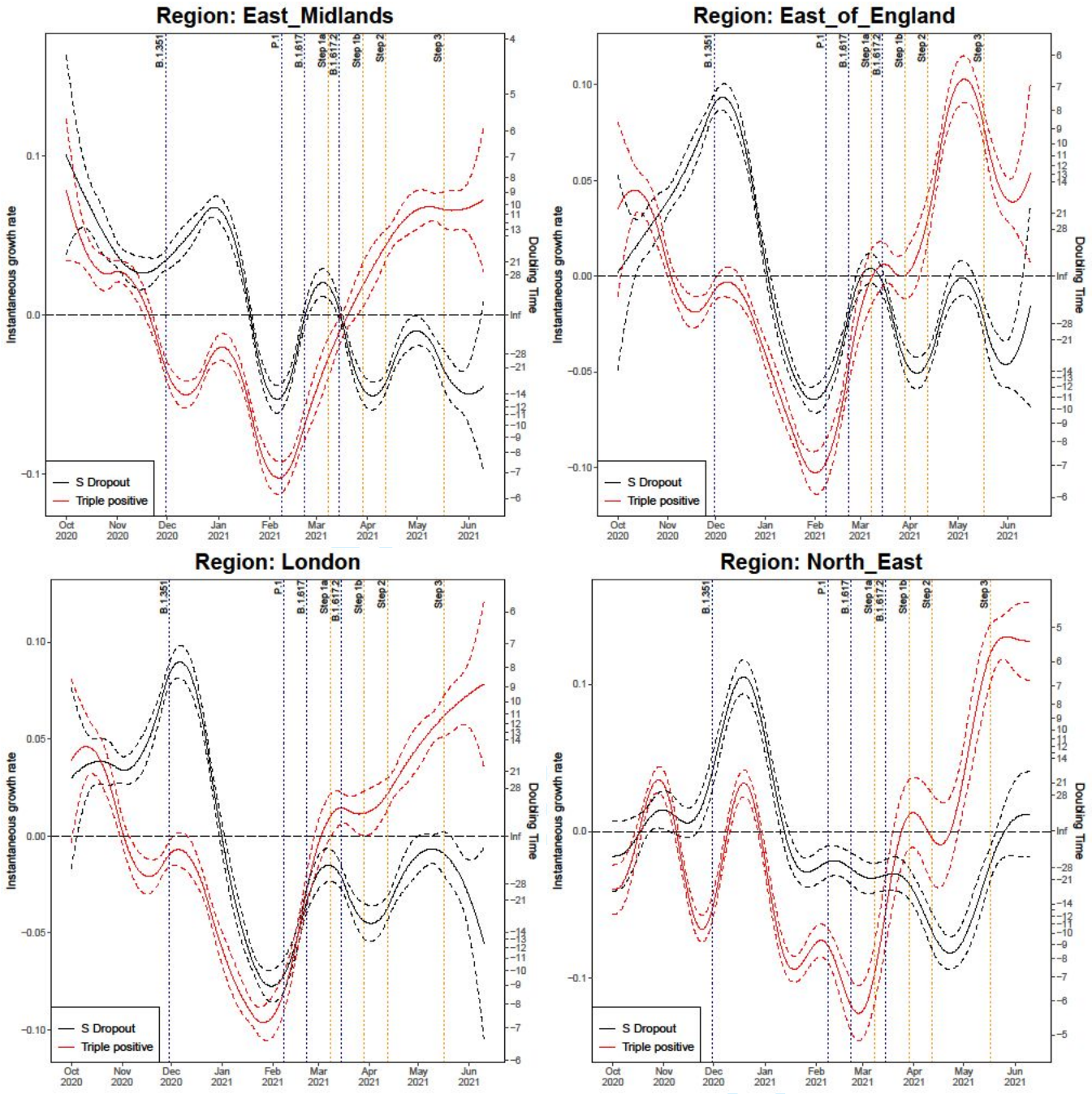
Appendix A: Cases Age Groups

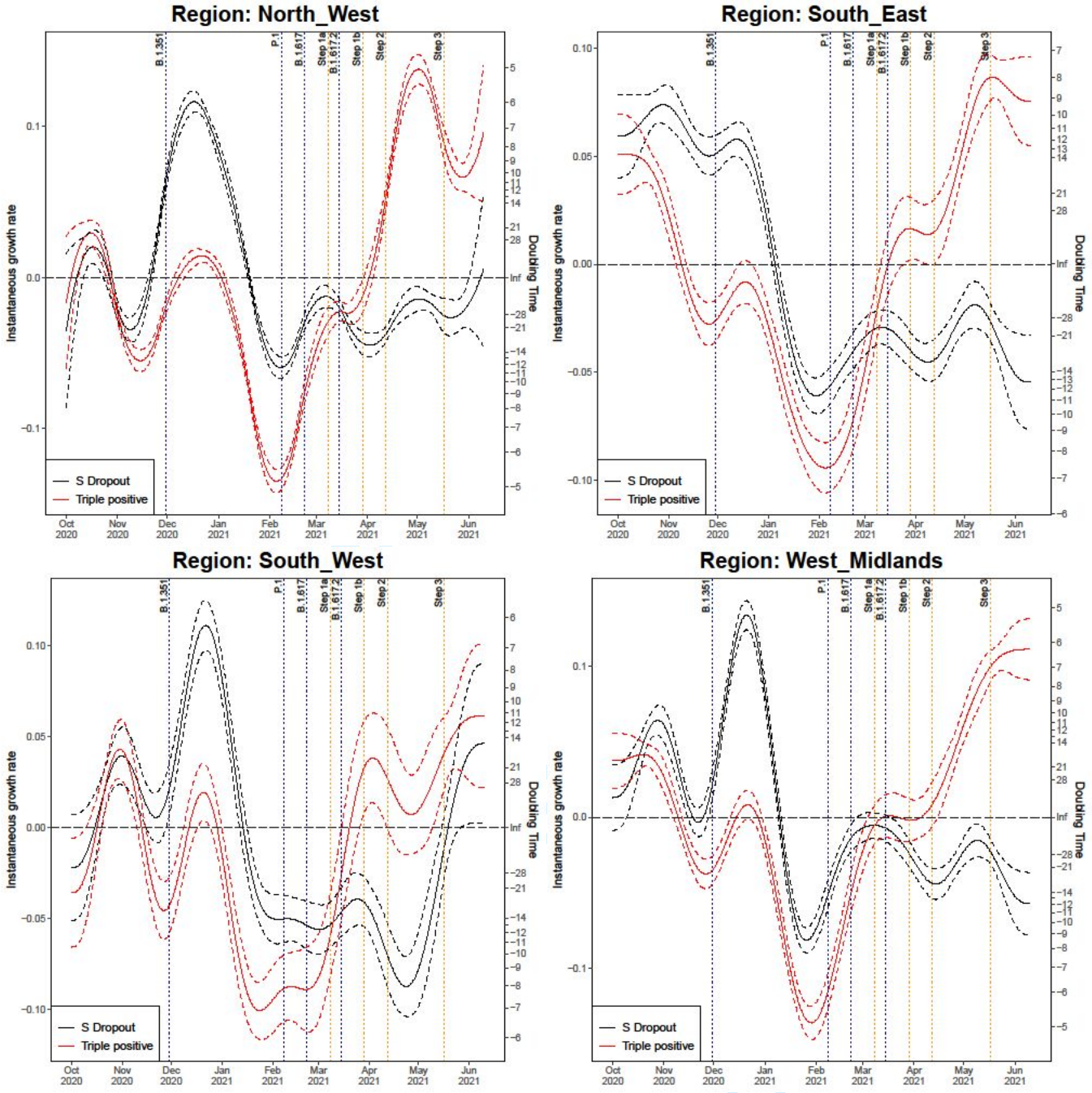


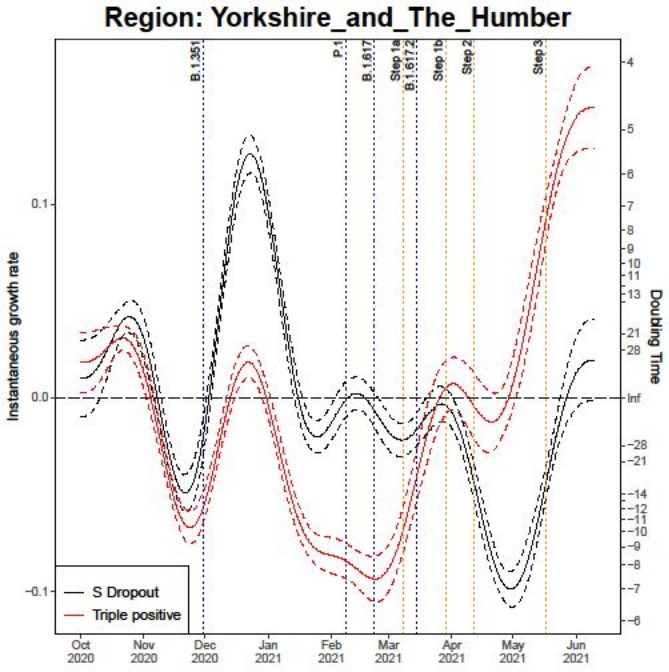




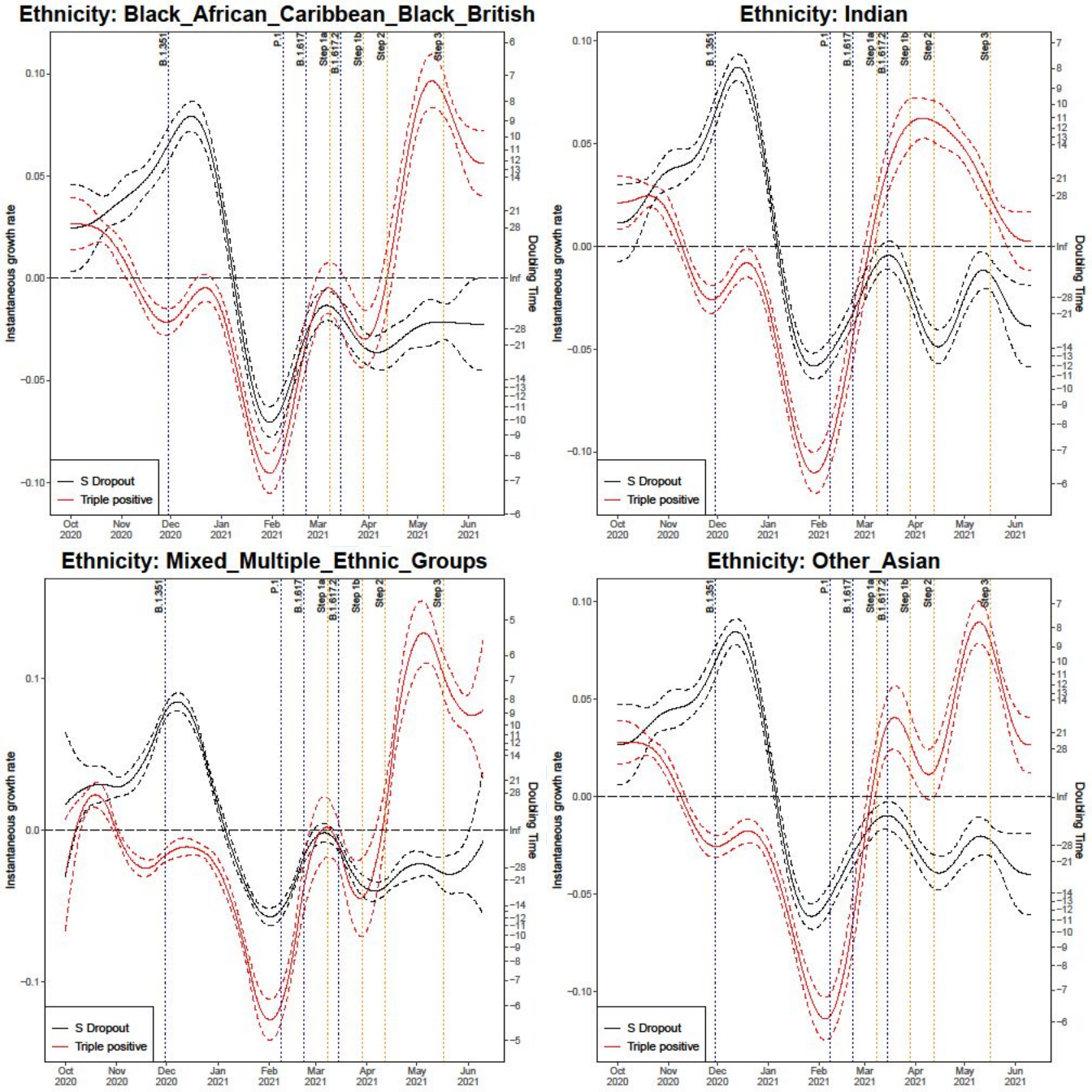
Appendix B: Cases Regions

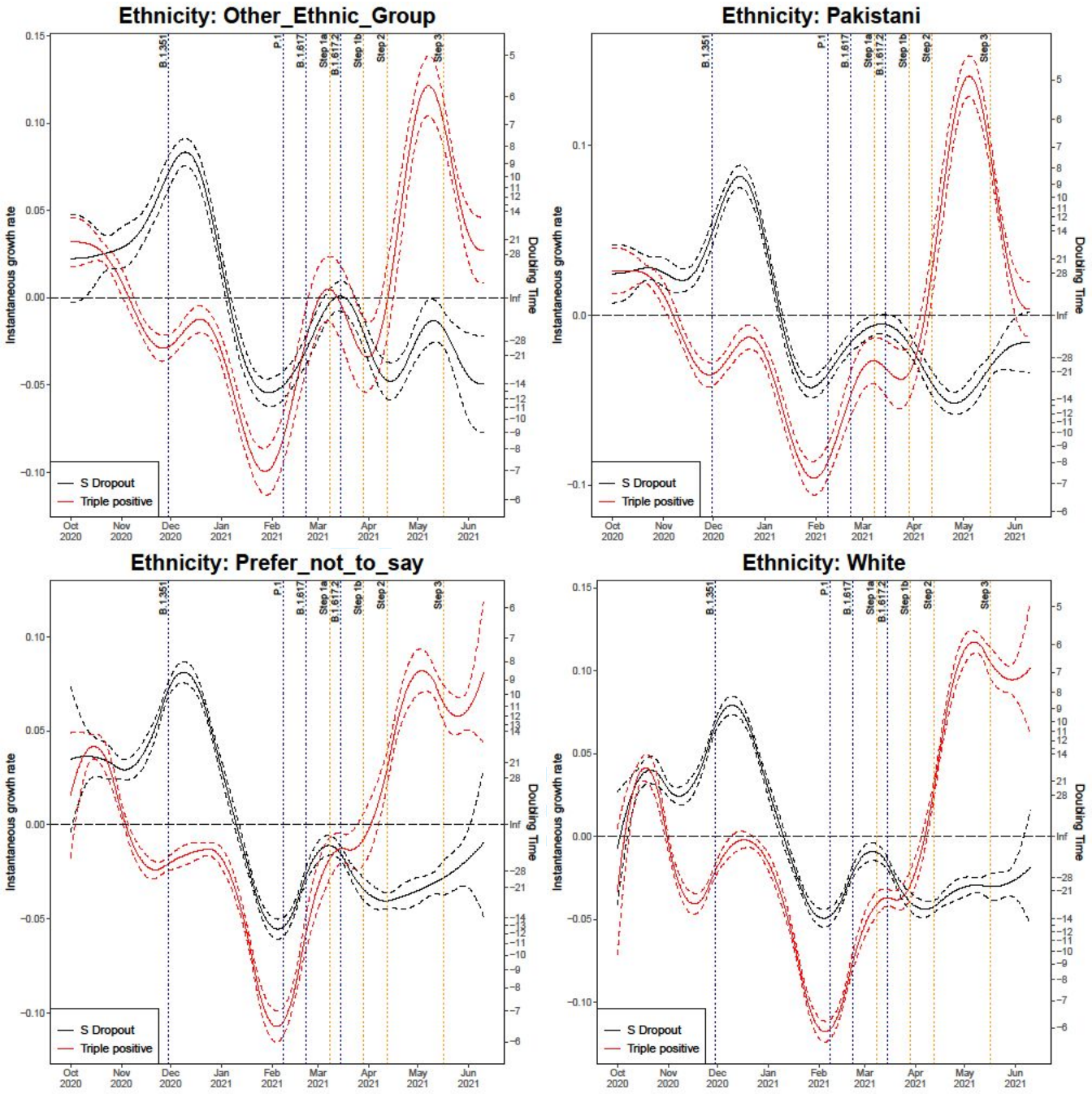




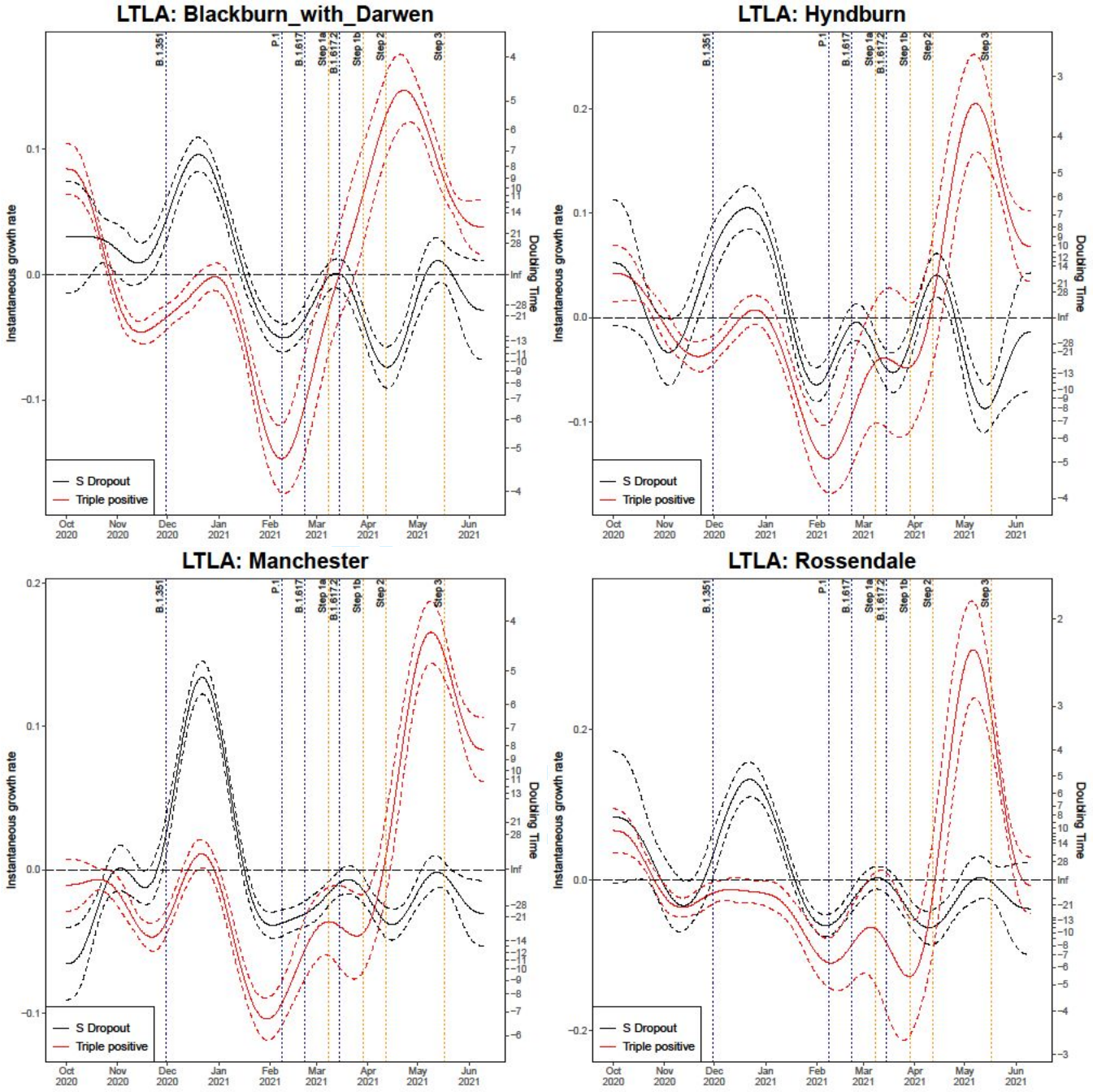


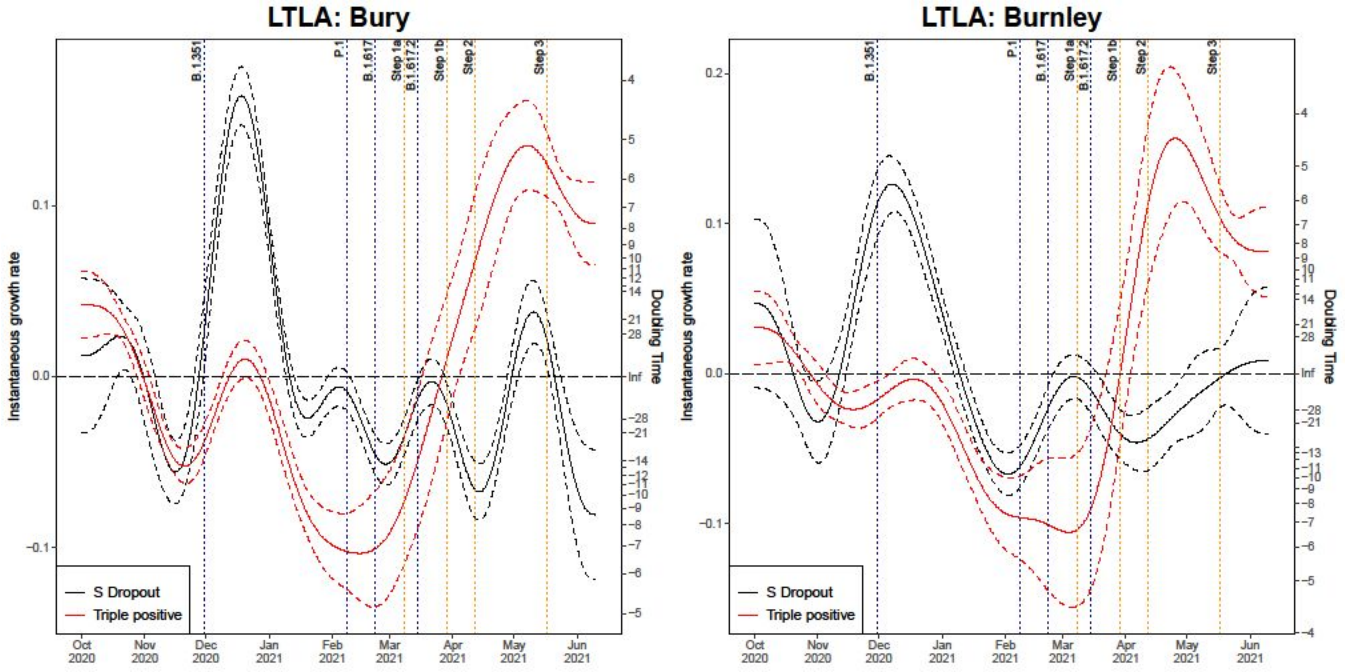
Appendix C: Cases Ethnicity

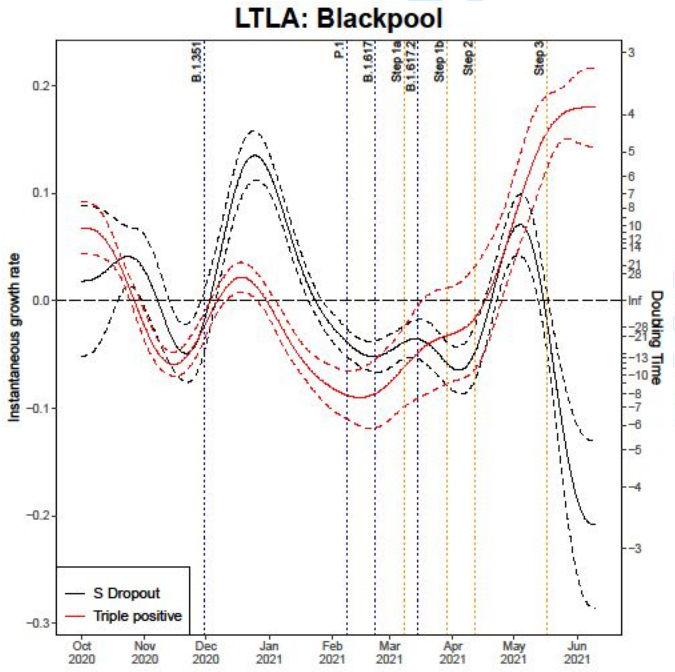


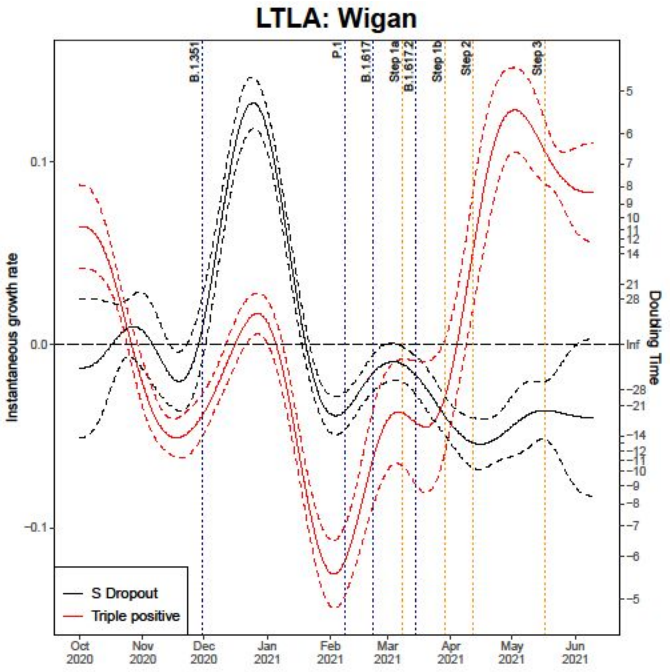


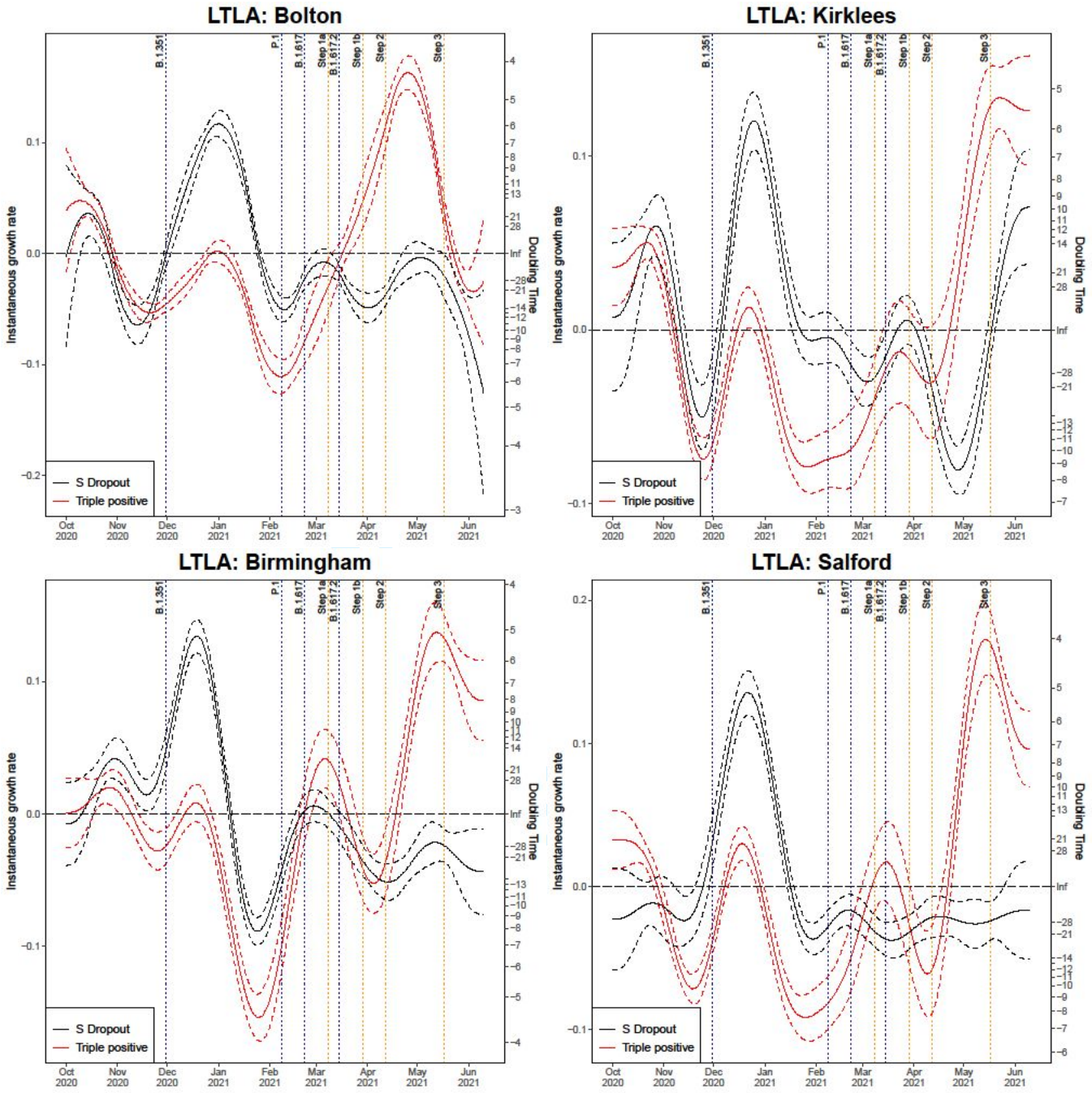
Appendix D: Cases LTLA

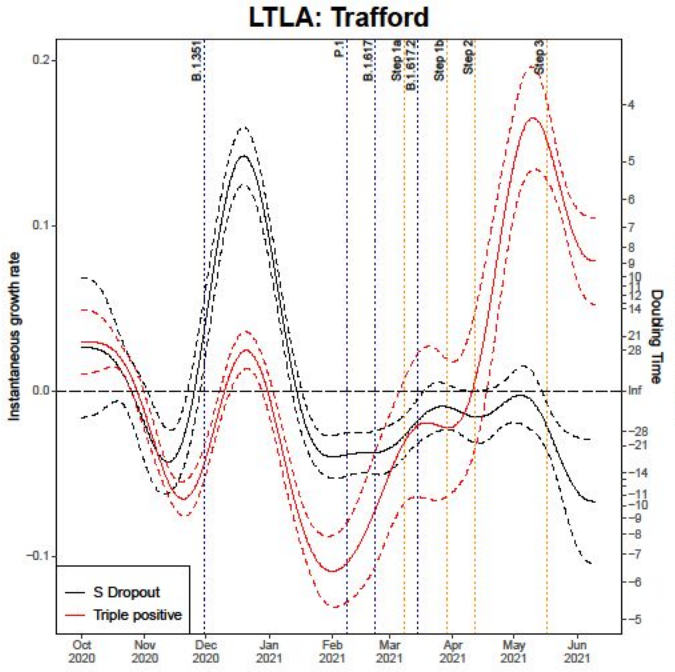


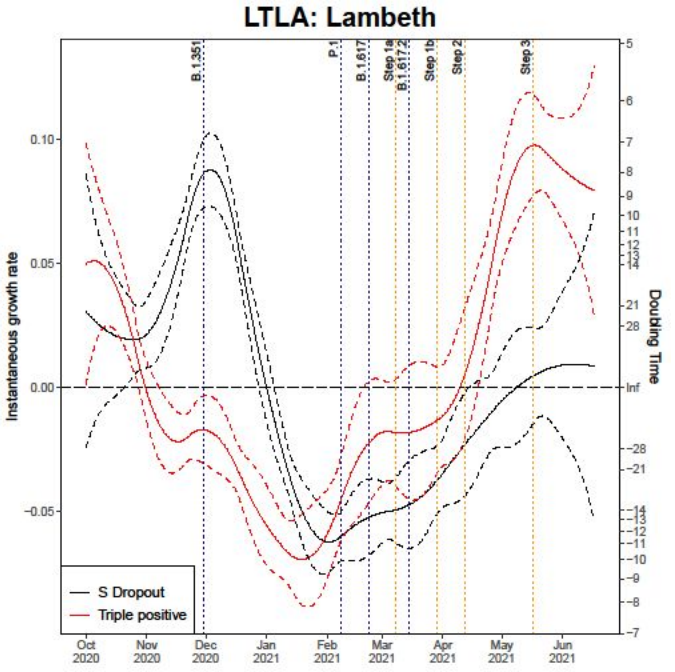




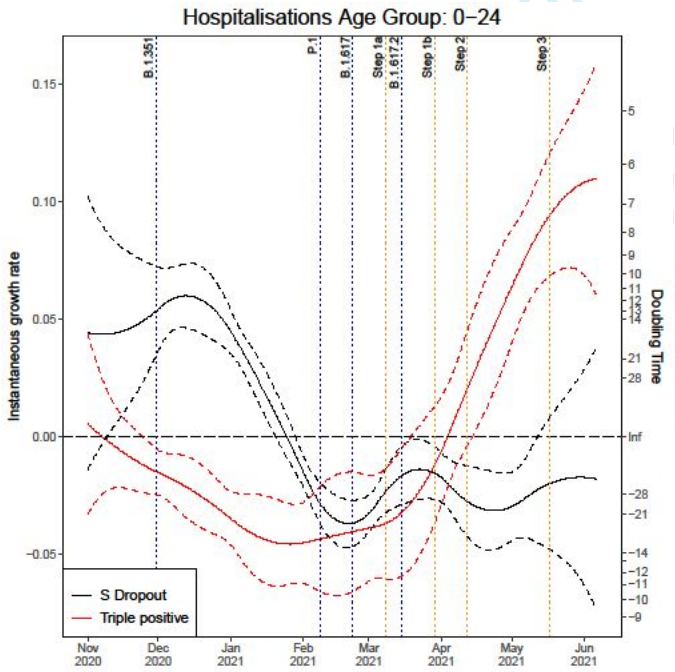


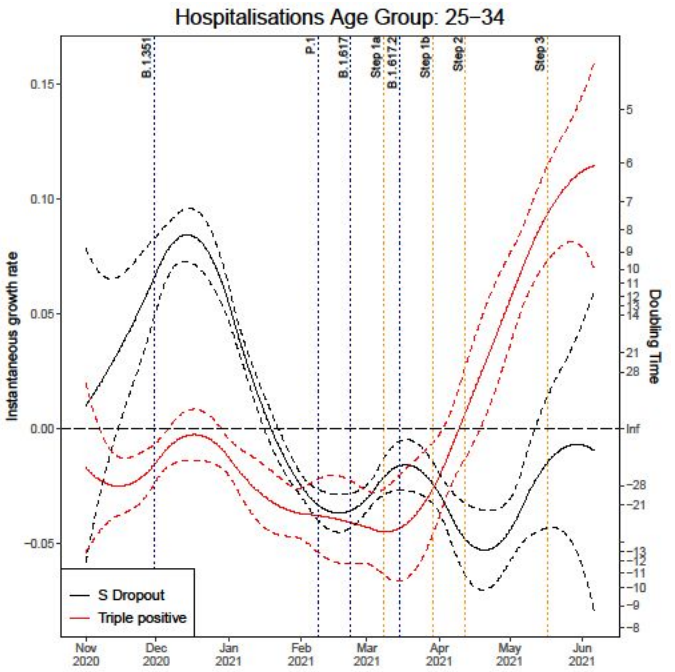


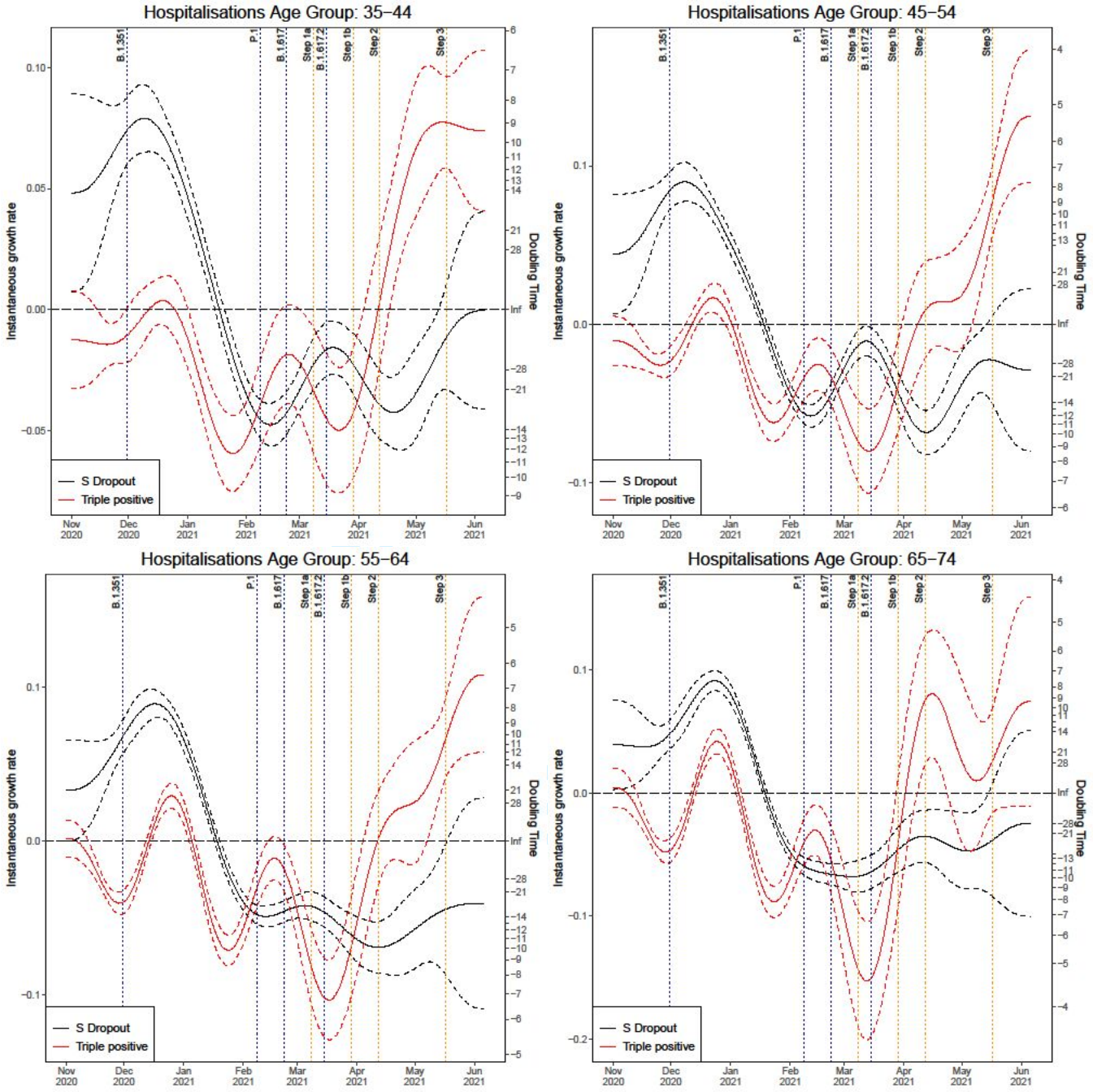


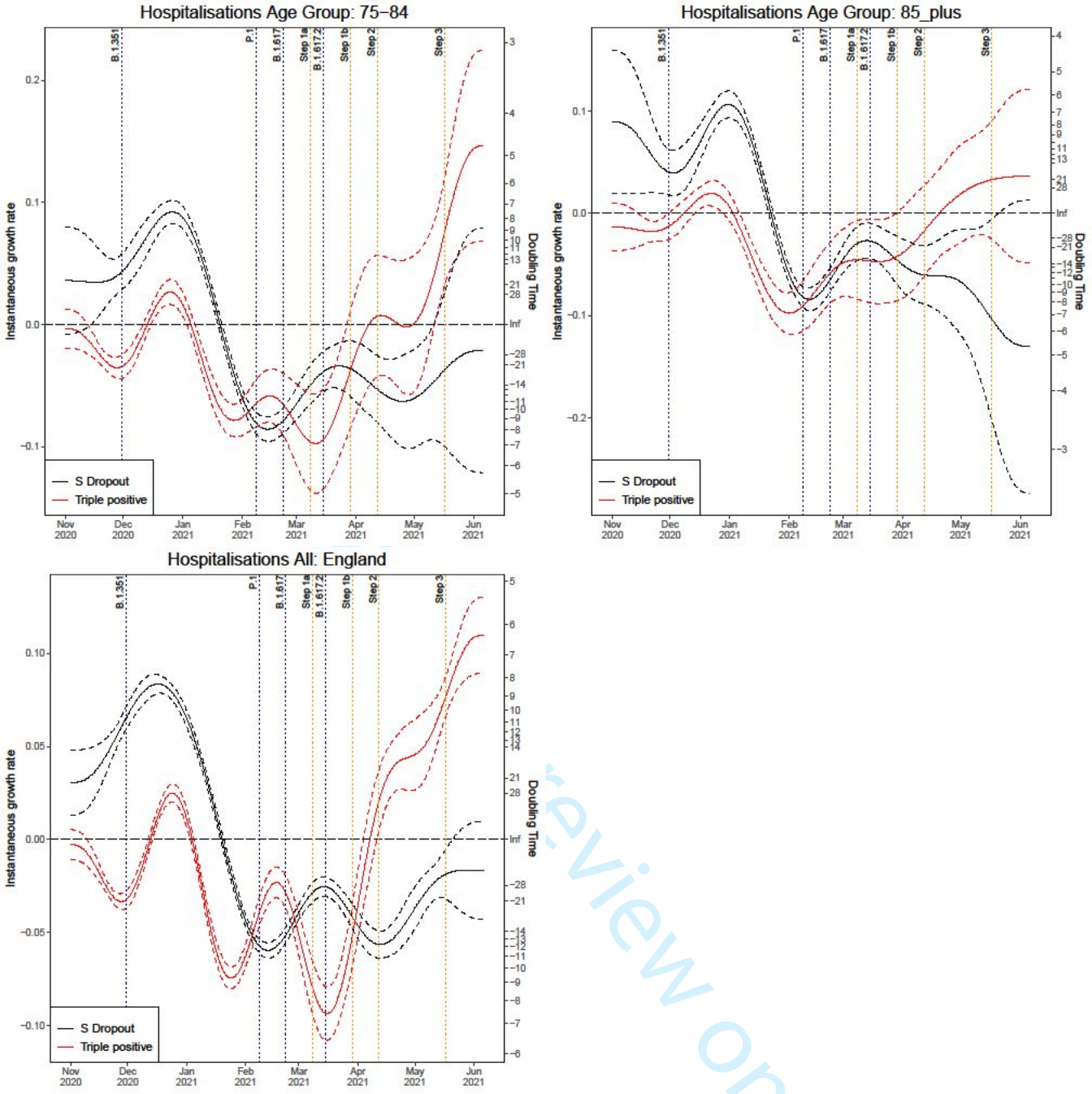


Appendix E: Hospitalisation Age Groups

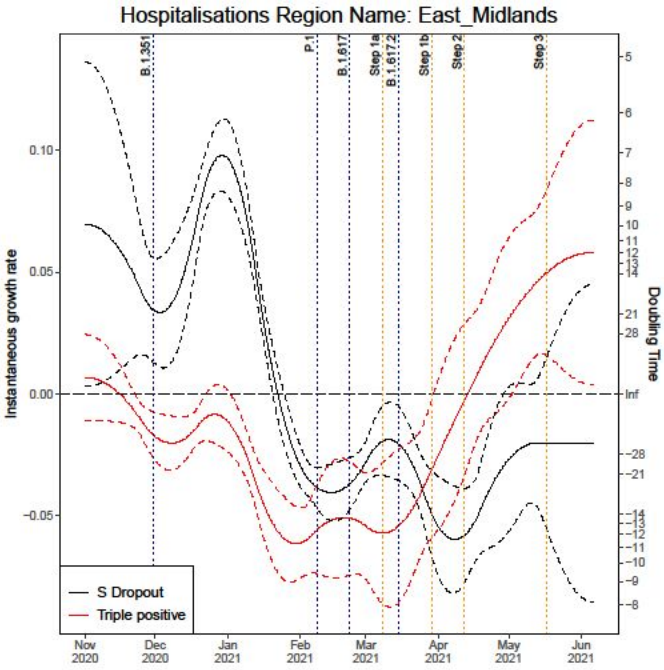


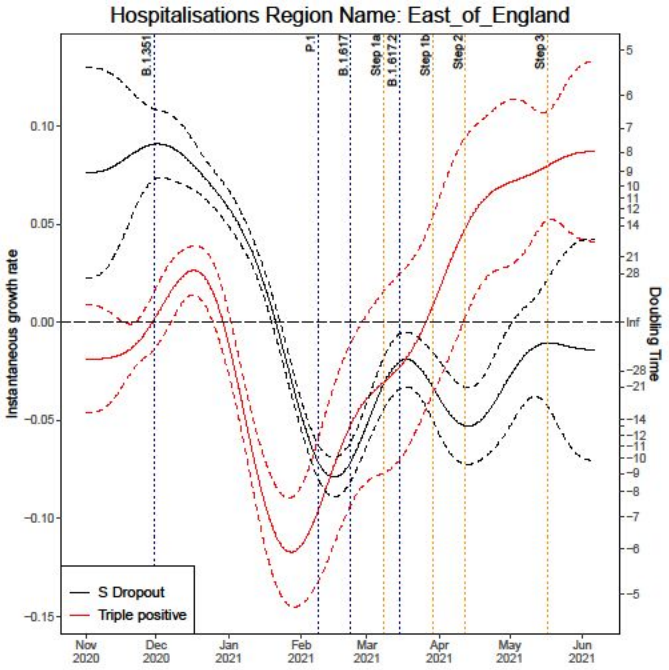


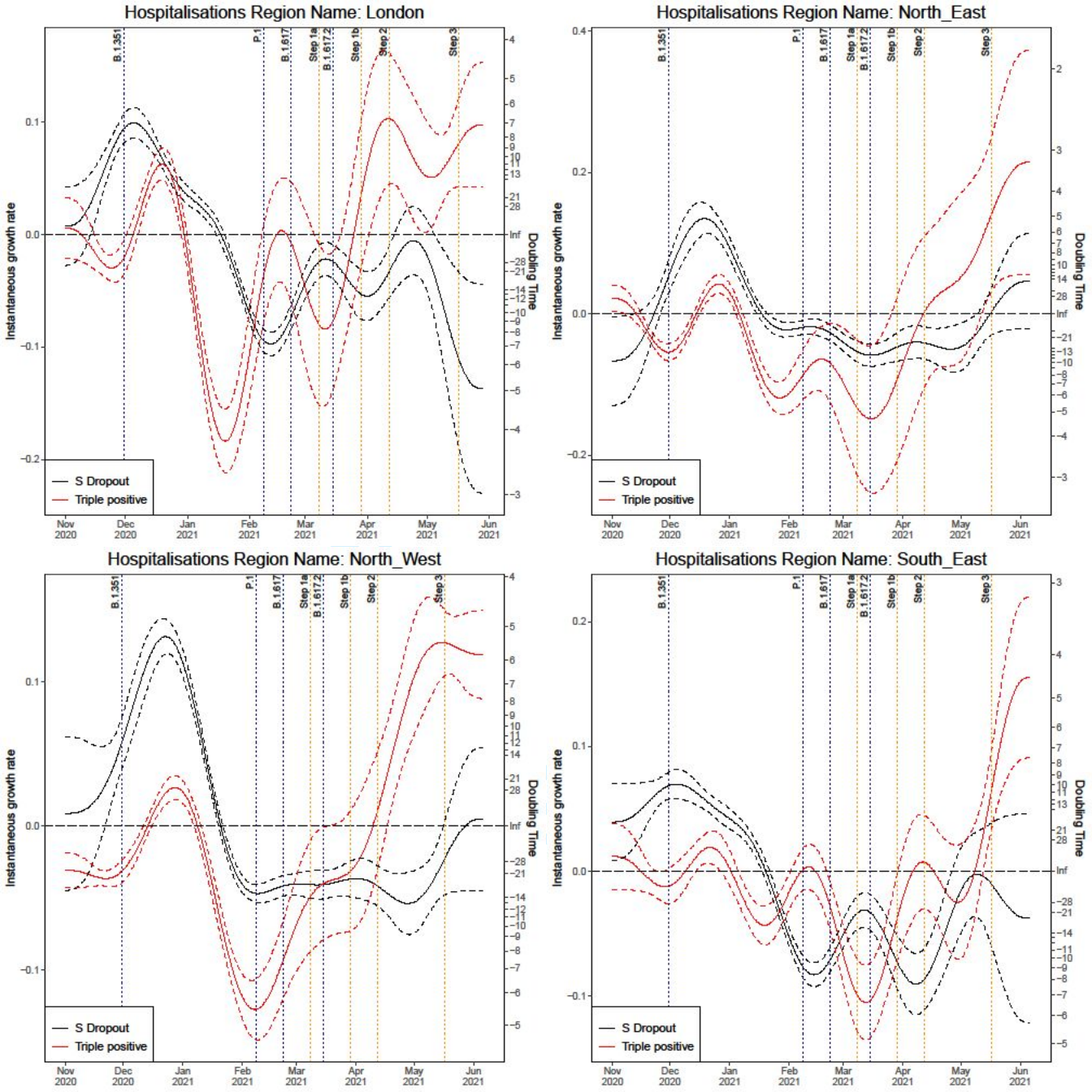


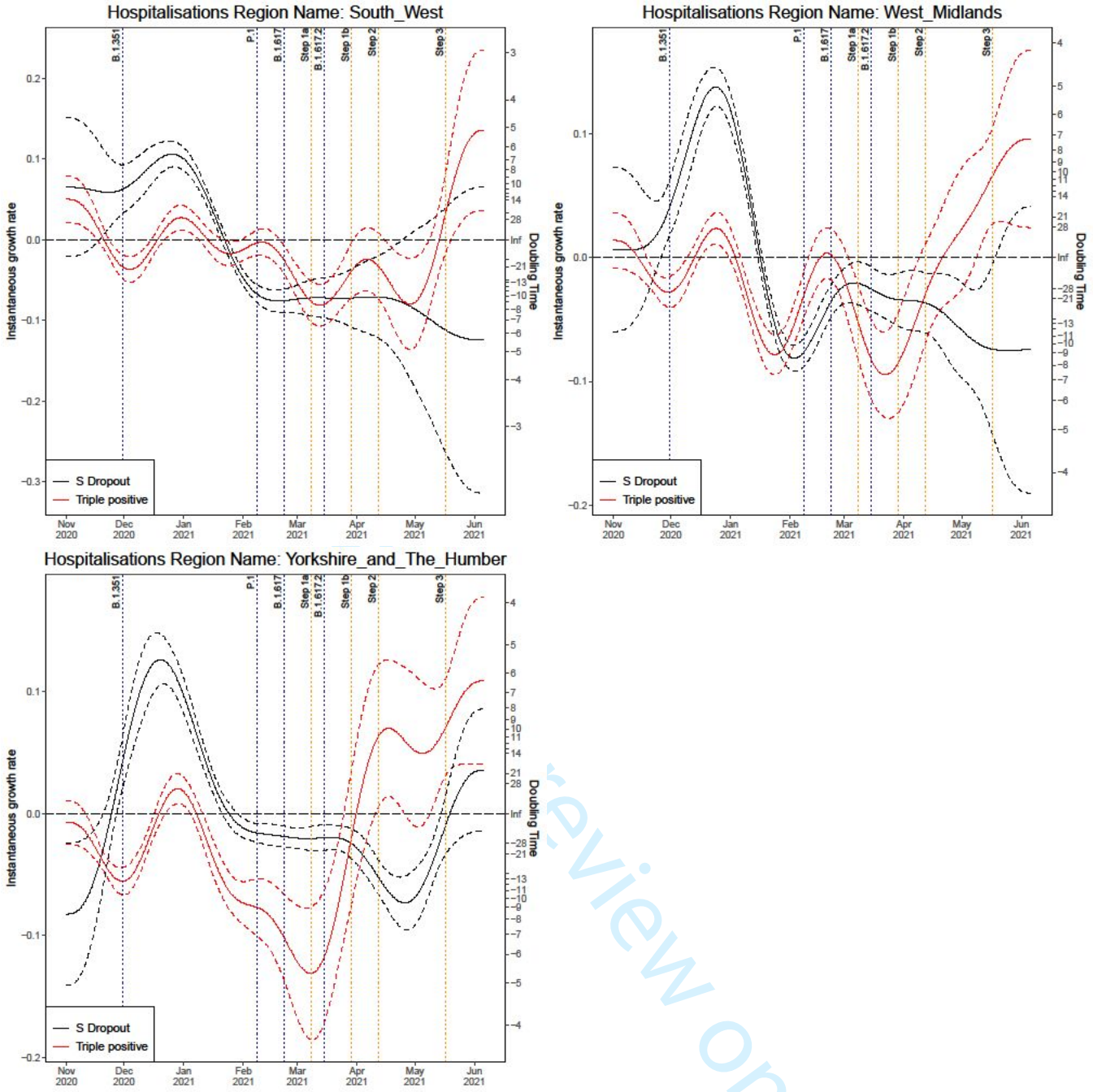


Appendix F: Hospitalisation Region

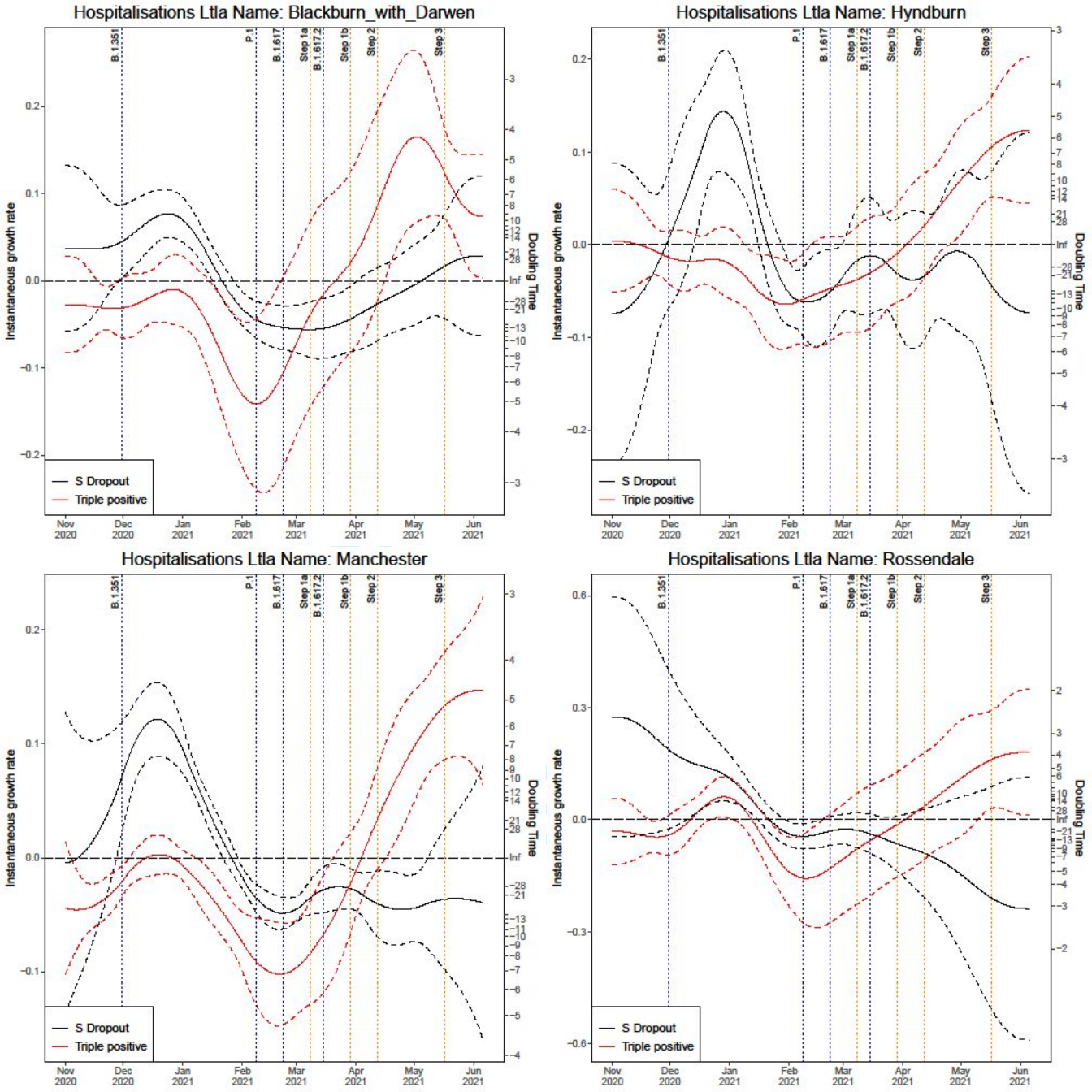


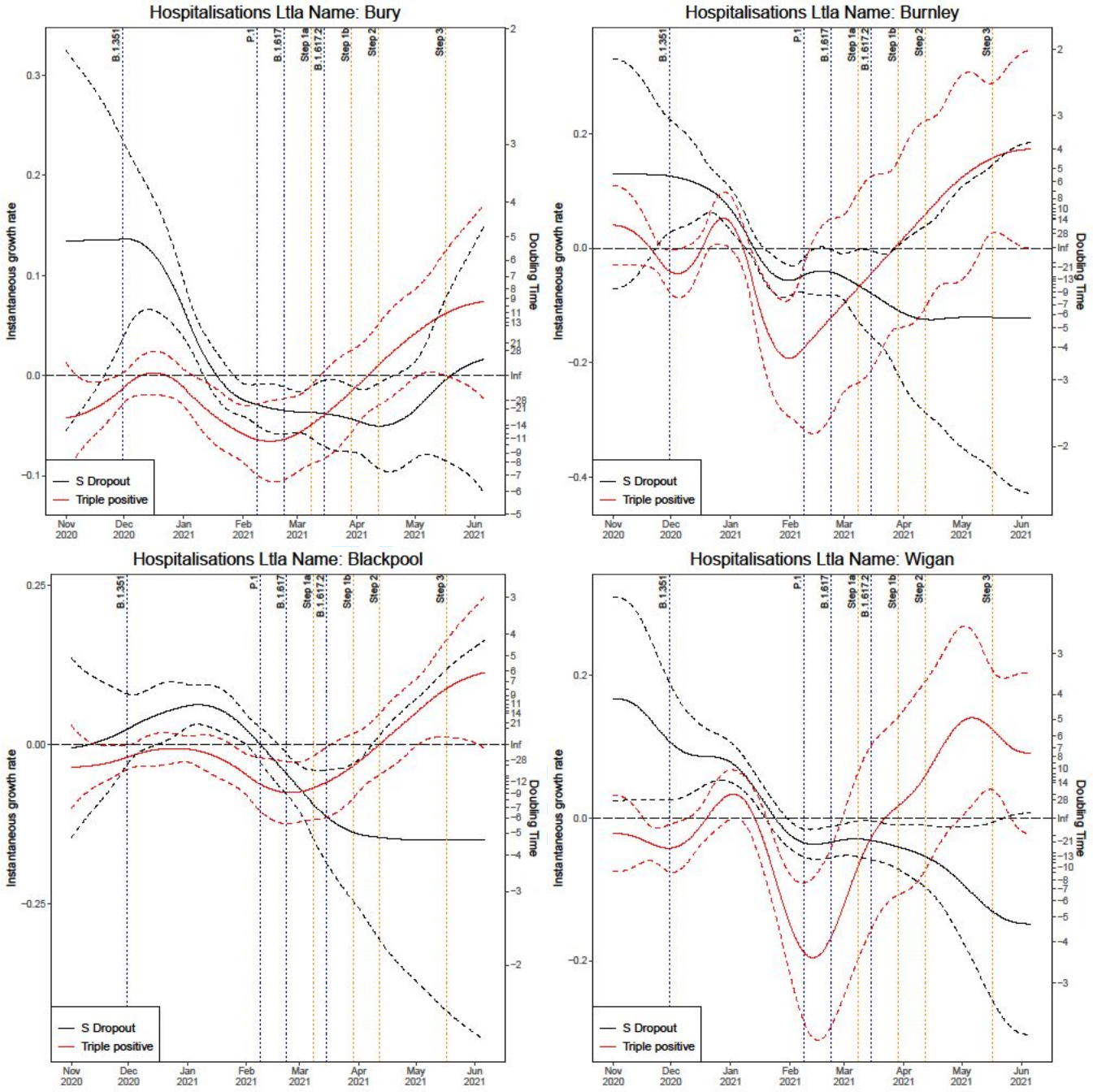


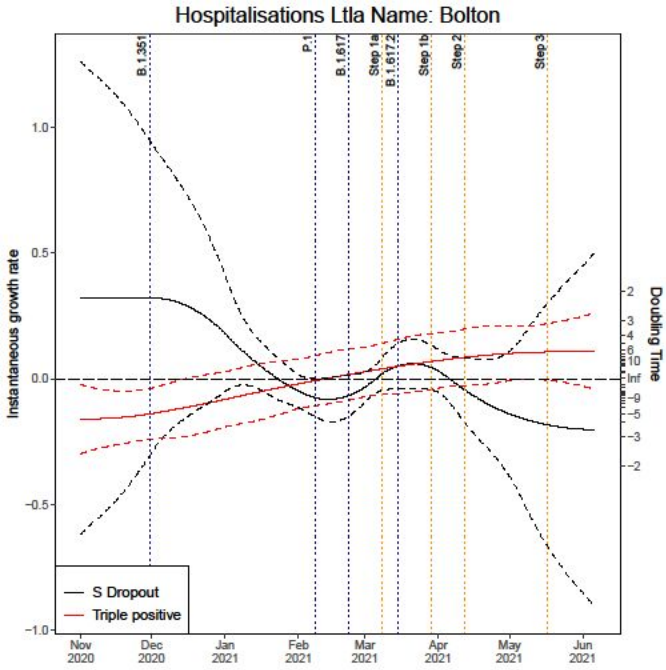


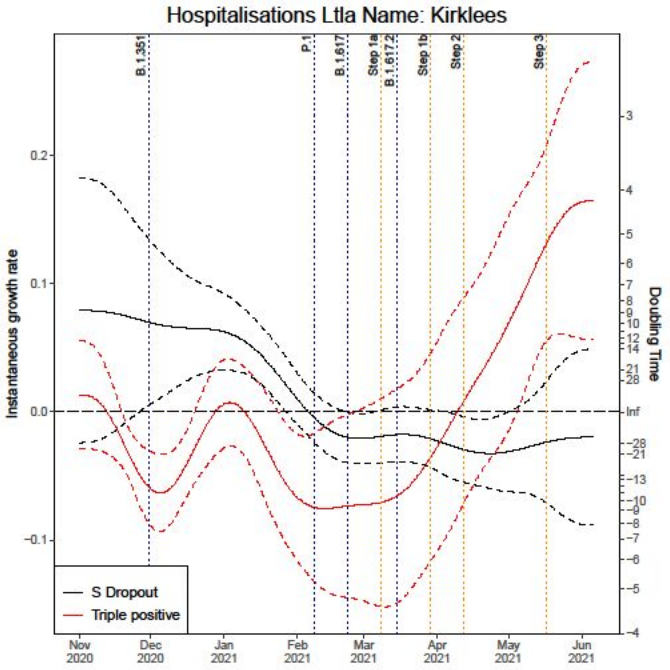


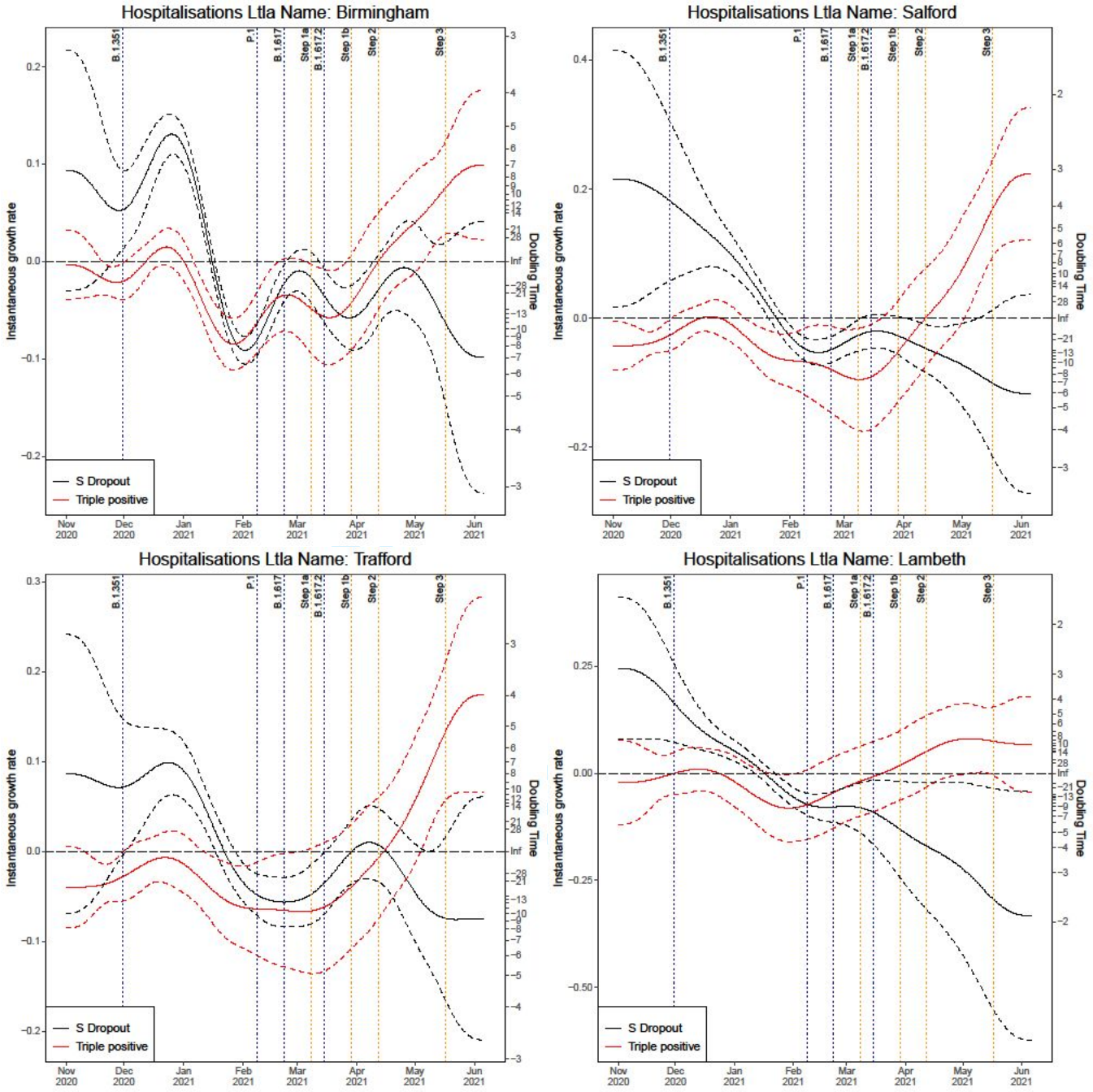
Appendix G: Hospitalisation LTLA











Appendix H

Time from Symptom Onset to Specimen Date

Start Date	End Date	N	Fit	Mean	SD	Alpha	Beta	Looic	Waic	Max Rhat	Min Rhat	Bad Pareto
01/01/2020	31/05/2020	557	weibull	3.66875	2.76983	1.34002	3.99321	2520.042	2519.159	1.00519	1.00089	0
01/06/2020	31/08/2020	4069	weibull	2.95717	2.39506	1.2423	3.17018	16859.34	16857.72	1.00025	0.999731	0
01/09/2020	30/09/2020	10000	weibull	2.6209	2.11027	1.24985	2.81378	39058.67	39055	1.00013	0.999542	0
01/10/2020	31/10/2020	10000	weibull	2.18193	2.05758	1.06113	2.2331	36226.77	36230.66	1.00124	0.99963	0
01/11/2020	30/11/2020	10000	weibull	1.97344	2.18076	0.906533	1.8826	34588.96	34602.01	1.00749	1.00303	0
01/12/2020	31/12/2020	10000	weibull	2.45101	2.17378	1.12993	2.56172	38238.76	38241.01	1.00077	1.00026	0
01/01/2021	31/01/2021	10000	weibull	2.07839	2.06587	1.00626	2.0836	35433.81	35439.34	1.00173	1.00054	0
01/02/2021	28/02/2021	10000	weibull	1.94196	2.00735	0.967796	1.91389	34250.36	34259.23	1.00184	0.999833	0
01/03/2021	31/03/2021	10000	weibull	1.96348	2.00263	0.98072	1.94674	34429.29	34437.08	1.00288	0.999865	0
01/04/2021	30/04/2021	10000	weibull	2.10586	2.06323	1.02091	2.12367	35648.06	35652.98	1.00075	0.999712	0
01/05/2021	31/05/2021	10000	weibull	2.01347	2.07044	0.972811	1.98905	34898.79	34906.32	1.00235	1.00046	0

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	3	State specific objectives, including any prespecified hypotheses	2
Methods			
Study design	4	Present key elements of study design early in the paper	2,3,4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2,3,4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	2,3,4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*2,3,4*
Bias	9	Describe any efforts to address potential sources of bias	2,3,4
Study size	10	Explain how the study size was arrived at	2,3,4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	2,3,4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	2,3,4
		(*b*) Describe any methods used to examine subgroups and interactions	2,3,4
		(*c*) Explain how missing data were addressed	
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	2,3,4

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	4-16
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*4-16*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	4-16
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	16-17
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	
Generalisability	21	Discuss the generalisability (external validity) of the study results	
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Growth, reproduction numbers, and factors affecting the spread of SARS-CoV-2 novel variants of concern in the United Kingdom from October 2020 to July 2021: a modelling analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-056636.R1
Article Type:	Original research
Date Submitted by the Author:	14-Oct-2021
Complete List of Authors:	Ward, Thomas; Public Health England, Department of Health and Social Care Glaser, Alex; United Kingdom Department of Health and Social Care Johnsen, Alexander; United Kingdom Department of Health and Social Care Xu, Feng; The University of Manchester Hall, Ian; University of Manchester, Department of Mathematics Pellis, Lorenzo; The University of Manchester
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Infectious diseases
Keywords:	Epidemiology < INFECTIOUS DISEASES, COVID-19, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Growth, reproduction numbers, and factors affecting the spread of SARS-CoV-2 novel variants of concern in the United Kingdom from October 2020 to July 2021: a modelling analysis

Thomas Ward[1*], Alex Glaser[2], Alexander Johnsen[2], Feng Xu[3], Ian Hall[3], Lorenzo Pellis[3]

1. Public Health England, London UK
2. Department of Health and Social Care, London UK
3. University of Manchester, Manchester UK

*Corresponding Author: Tom.Ward@dhsc.gov.uk

Abstract

Objectives: Importations of novel variants of concern (VOC), particularly B.1.617.2, have become the impetus behind recent outbreaks of SARS-CoV-2. Concerns around the impact on vaccine effectiveness, transmissibility, and severity are now driving the public health response to these variants. This paper analyses the patterns of growth in hospitalisations and confirmed cases for novel variants of concern by age groups, geography, and ethnicity in the context of changing behaviour, non-pharmaceutical interventions (NPIs), and the UK vaccination programme. We seek to highlight where strategies have been effective and periods that have facilitated the establishment of new variants.

Design: We have algorithmically linked the most complete testing and hospitalisation data in England to create a dataset of confirmed infections and hospitalisations by SARS-CoV-2 genomic variant. We have utilised these linked datasets to analyse temporal, geographic, and demographic distinctions.

Setting and participants: The setting is England from October 2020 to July 2021. Participants included all COVID-19 tests that included RT-PCR CT gene target data or underwent sequencing and hospitalisations that could be linked to these tests.

Methods: To calculate the instantaneous growth rate for variants of concern we have developed a Generalised Additive Model (GAM) fit to multiple splines and varying day of the week effects. We have further modelled the instantaneous reproduction number R_t for the B.1.1.7 and B.1.617.2 variants and included a doubly interval censored model to temporally adjust the confirmed variant cases.

Results: We observed a clear replacement of the predominant B.1.1.7 by the B.1.617.2 variant without observing sustained exponential growth in other novel variants. Modelled exponential growth of triple positive cases was initially detected in the youngest age groups, although we now observe across all ages a very small doubling times of 10.7 (95% CI: 9.1, 13.2) days and 8 (95% CI: 6.9, 9.1) days for cases and hospitalisations, respectively. We observe that growth in triple positive cases was first detected in the Indian ethnicity group in late February, with a peak of 0.06 (95% CI: 0.07, 0.05) in the instantaneous growth rate, but is now maintained by the white ethnicity groups, observing a doubling time of 6.8 (95% CI: 4.9, 11) days. R_t analysis indicates a reproduction number advantage of 0.45 for B.1.617.2 relative to B.1.1.7, with the R_t value peaking at 1.85 for B.1.617.2.

Conclusions: Our results illustrate a clear transmission advantage for B.1.617.2 and the growth in hospitalisations illustrates that this variant is able to maintain exponential growth within age groups that are largely doubly vaccinated. There are concerning signs of intermittent growth in the B.1.351 variant, reaching a 28 day doubling time peak in March 2021, although this variant is presently not showing any evidence of a transmission advantage over B.1.617.2. Step 1b, of the UK national lockdown easing, was sufficient to precipitate exponential growth in B.1.617.2 cases for most regions and younger adult age groups. The final stages of NPI easing appeared to have a negligible impact on the growth of B.1.617.2 with every region experiencing sustained exponential growth from Step 2. Nonetheless, early targeted local NPIs appeared to markedly reduced growth of B.1.617.2. Later localised interventions, at a time of higher prevalence and greater geographic dispersion of this variant, appeared to have a negligible impact on growth.

Strengths and limitations of this study

- There is currently no study published that looks at the growth in hospitalisations by variants of concern (VOC) in England or illustrates the impact of demographics, geography and behaviour for the introduction and growth of novel variants.

- The study calculates the growth and R_t in the clinical and testing data therefore, conclusions regarding extrinsic factors are inferential and descriptive.
- Geographic bias for laboratories that supplied RT-PCR CT data for the genes OR, N and S was evident in England, which led to larger confidence intervals observed in some regions.
- We did not algorithmically link vaccination status to each individual as this subset the data further and reduced the feasibility of meaningful growth rate analysis and therefore, age groups were used a proxy indicator.
- To calculate R_t by variant we had to assume there to be reasonably consistent ascertainment bias in the testing data.

Introduction

The SARS-CoV-2 virus has had an unprecedented impact and global reach since the first officially confirmed case in December 2019 [1]. Periods of high global prevalence of the virus has allowed the emergence of novel mutations through antigenic drift, with evidence this is largely a reaction to the host immune response [2]. Furthermore, we may now begin to see selective mutation in response to natural and vaccine induced immunity. The rate of mutation for coronaviruses has been poorly understood; however, evidence from seasonal human coronaviruses HCoV-229E and HCoV-OC43 illustrate that the evolution of SARS-CoV-2 may have parallels with the influenza A virus (IAV), including more concerning adaptive changes to the receptor binding domain (RBD) [3]. In addition, viruses akin to SARS-CoV-2, that are RNA based, tend to show high rates of mutation, which are likely to be related to insufficient proofreading abilities [4]. Imports of novel variants of COVID-19 are now of great concern as they become the impetus behind localised outbreaks in the United Kingdom [5]. A UK government modelling report from June 2021 highlighted the significance of importations [6] and it was estimated that SARS-CoV-2 lineages derived from individuals that had recently travelled had a higher relative reproduction number.

The vaccination campaign began in England on 8th December 2020 with care home residents, the most clinically vulnerable, and hospital staff. This was followed by an age stratified structure that commenced with the over 80s on the 17th January 2021 and reached the 21-30 age group by the 16th June 2021 [7]. The vaccination campaign began with Pfizer/BioNTech and AstraZeneca with first doses prioritised. The age groups over 40 were primarily administered with AstraZeneca; Pfizer/BioNTech and Moderna were administered largely to the younger age groups in response to concerns over haemostatic side effects [8]. The chief concerns around importations of novel variants have been driven by immunological escape. A recent trial in South Africa [9] found that the AstraZeneca vaccine had a two-dose efficacy of 10% against B.1.351 at preventing mild to moderate disease, albeit this study utilised very limited data. Further research found the B.1.617 variant, that was first detected in October 2020 in India, carries two mutations on the RBD and preliminary results indicated this may have an impact upon vaccine effectiveness [10, 11]. B.1.617.2, a sub-lineage of B.1.617, has caused global concern due to the rate of growth that has been observed since it was first sequenced in India. A recent study [12] that analysed the sera of patients infected with B.1.617.2 found that it has 9 spike mutations on the N terminal domain (NTD) and the RBD. This study observed that B.1.617.2 is resistant to neutralisation with the efficacy of the Pfizer vaccine around 3 to 6 times less than observed with B.1.1.7. Nonetheless, there is some evidence that doubly vaccinated individuals may still possess robust neutralisation titres against B.1.617.2 and there is still relatively high vaccine effectiveness against symptomatic disease [13]. However, these results do not take into account that symptomatic status is poorly recorded for PCR tests in England and that sequenced B.1.617.2 variant cases were limited at the time.

The evidence of substantial viral epitopic mutation has necessitated a risk categorisation for novel mutant strains in the United Kingdom. Variants that display epidemiological and immunological characteristics of concern are defined as a Variant Under Investigation (VUI) [14] and after committee evaluation may be escalated to a Variant of Concern (VOC). As of the 12th May 2021 there are eight VUIs and five variants defined as VOCs: B.1.1.7 (VOC-20DEC-01) *Alpha*, B.1.351 (VOC-20DEC-02) *Beta*, P.1 (VOC-21JAN-02) *Gamma*, B.1.1.7 with E484K (VOC-21FEB-02), and B.1.617.2 (VOC-21APR-02) *Delta* [15]. The most concerning VOCs presently are B.1.351 and B.1.617.2 due to evidence of diminished vaccine effectiveness, particularly in the former. There is also growing evidence that B.1.617.2 has acquired mutations that have increased the viral fitness improving the transmissibility of this lineage.

In this paper we have utilised RT-PCR CT data for the genes OR, N, and S and sequenced tests for COVID-19 that have been algorithmically linked to hospitalisation datasets. We assess the temporal variability in the growth of VOCs relative to the previously predominant B.1.1.7 variant across the geography of the United Kingdom. We further assess how the instantaneous growth rate has changed across ages, ethnicity, and in response to the easing of nonpharmaceutical interventions (NPIs). Finally, we assess the relative difference in the reproduction number between B.1.617.2 and the established B.1.1.7 variant.

Methods

Epidemiological and Clinical Data

Accident and Emergency (A&E) dataset is from the SUS suite of datasets. This data was linked with the PCR targets from the Second Generation Surveillance System (SGSS). The linkage allows the patient's pathway to be followed and provides additional information beyond what is obtainable from the standalone datasets. The linkage algorithm has evolved from research and development undertaken on the 2 datasets. The algorithm is primarily based on fields that:

- identify the patient, using a pseudo identifying number

- report the outcome of the A&E attendance, and

- report the method of admission.

Further, the basic principles behind the linkage method are where:

i) the unique patient pseudo identifier is the same in A&E and SGSS data, and
ii) the SGSS Specimen date is between 6 days before or 14 days after the A&E admissions date.

For multi-episode spells the admission date, used for the linkage, comes from the first episode in a spell. Linkage was conducted in a secure research environment and with full anonymisation of the data. The linked hospitalisation data was subset nationally by age, region, and Lower Tier Local Authority (LTLA).

RT-PCR testing for COVID-19 involves a detection of three genes OR, N, and S. The S-gene mutation in B.1.1.7 results in a dropout of S-gene detection, providing an easier prevalence indicator for this variant where information is available on the RT-PCR gene target. Due operational and logistical limitations sequenced viral genomic data was limited and therefore we employed S dropout to identify B.1.1.7 and triple positive (OR, N, and S gene positive) was used as a proxy for the identification of VOCs that include the B.1.351, P1, and B.1.617.2 variants considered in this analysis. Although, from the end of March 2021 B.1.617.2 accounted for almost the entirety of triple positive variant cases. We analysed S dropout and triple positive cases from the Public Health England NPEX dataset, which was subset by travel status, ethnicity, age, region, and LTLA. The RT-PCR data was linked the SUS dataset to acquire hospitalisations for triple positive and S dropout variants. Further genomic sequenced data was acquired through SGSS and suspected variants from the reflex assays. The P.1, B.1.617.2, B.1.351, and B.1.1.7 variants were included in this analysis and other variants were excluded due to low numbers. Owing to limited data, analysis of genomic sequenced and reflex assay data was only conducted at a national spatial resolution.

Instantaneous Growth Rate and Doubling Times

The method for the estimation of the time varying growth rates and doubling times is adapted from a Generalised Additive Model (GAM) with a canonical link [16, 17]. We allow for a varying day of the week fixed effect: no day, weekend, or weekday effect. We further fit to cubic regression splines [18], P-splines [19], thin-plate splines (a low rank isotropic smoother) [20], Duchon splines (allowing for lower orders of the derivative in the penalty relative to the thin plate splines) [21] and Gaussian process smoothers. The model assumes the number of cases $\mathrm{y}(\mathrm{t})$ is proportional to $\exp(s(t))$ for some smoother $\mathrm{s}(\mathrm{t})$ [22]. The over-dispersed noise inherent in both disease dynamics and surveillance data motivates the use of a negative binomial error structure. The instantaneous growth rate is obtained as the time derivative of the smoother, $r_s = \dot{s}(\mathrm{t})$, and the instantaneous doubling time is calculated as $t_D = \log(2)/\dot{s}$. Asymptotic confidence intervals (CIs) on r_s are only indicative of uncertainty on t_D, especially when the variance grows as r_s approaches zero. The number of knots used by the spline is fixed as one twentieth the length of the time-series (for time-series shorter than 200 days the default number of knots is used) to avoid over smoothing the data or loosing signal in the noise. The model for each group, fit to each spline and day of the week effect, is assessed by the leave-one-out (LOO) and the Akaike information criterion (AIC) metrics to select the best model fit.

Included in each plot is the date of the first confirmed case for B.1.1.7, B.1.617.2, B.1.351, and P.1 as they were considered of most concern at this time due to overall volume and phenotypic characteristics. In addition, each step of the national lockdown easing [23] in England has been included:

- Step 1a – Schools and universities are to re-open, care homes allow visitors and recreation within households and support bubbles are allowed.
- Step 1b – The 'stay at home' rule will end and outdoor sports to resume. Furthermore, the rule of 6 begins and two household can meet outdoors.
- Step 2 – Non-essential retail, gyms and outdoor hospitality will reopen.

- Step 3 – A lifting of most legal restrictions on mixing outdoors, events of up to 30 persons can be held, and indoor hospitality can recommence.

Instantaneous Reproduction Number R_t

This model utilised the diagnostic RT-PCR gene target results for positive tests with the S dropout and triple positive cases as a proxy for infections of the B.1.1.7 and B.1.617.2 variants. We calculate the instantaneous reproduction number [24] that corresponds to the average quantity of secondary cases that develop from the primary cases infected at a time period we call t, if conditions remained constant. Defined as:

$$R_t = \frac{I_t}{\sum_{s=1}^{t} I_{t-s} w_s}$$

Here I_t is defined as the quantity of incidence on day t and w_s is the discretised generation time distribution.

This approach was utilised as it is a reflection of the instantaneous transmissibility at a given point in time of the SARS-CoV-2 variant. The limitation of this approach is that it assumes there is a reasonable consistency in the ascertainment bias in the testing data and those tests included for targeted gene sequencing. Individuals that had a flag for recent travel were removed from the dataset prior to analysis.

The serial interval pertains to the duration of time from the onset symptoms of a primary case to the onset of symptoms for a secondary case that was infected by the primary case. To account for uncertainty in the serial interval we utilised a Monte Carlo simulation model of data sourced from a meta-analysis [25] that included studies which have published on the mean and standard deviations of the serial Interval for SARS-CoV-2.

The most complete available testing data for England is recorded at the specimen date of the test. To calculate the instantaneous reproduction number, we would optimally utilise the symptom onset date of positive cases at time t. Therefore, to adjust for this temporal discrepancy we have adapted a Hamiltonian Bayesian Monte Carlo Markov Chain (MCMC) doubly interval censored model from Ward & Johnsen, (2021) [26] to calculate temporal changes over time, for the lag from symptom onset to specimen date fitting to a Weibull distribution.

Patient and Public Involvement Statement

Patients were not involved in the development of the research question and study design.

Results

Across all age groups in England we can observe that the decay rate for triple positive cases peaked at the start of February in *Fig 1*. This was followed by a rapid reduction in the rate of exponential decay for triple positive cases and subsequent exponential growth from the 25th March, which is 10 days after the first confirmed case of B.1.617.2. Conversely, we observe exponential decay in the S dropout cases from the 7th January. We can observe that Step 1a occurred shortly before the exponential growth of triple positive cases and a steeper gradient in the line can be seen after Step 1b and Step 2. This model had a sample size of 1,108,537 triple positive and 1,051,205 S dropout cases and the model fit to the data can be seen in *Fig 2* and *Fig 3*.

Fig 1 Instantaneous growth rate and doubling times for confirmed positive tests in England by triple positive and S dropout variants

Fig 2 The maximum likelihood model fit to the S dropout confirmed positive testing data

Fig 3 The maximum likelihood model fit to the triple positive confirmed positive testing data

We can observe a sharp reduction in the rate of the exponential decay for triple positive hospital admissions from the 14th March in *Fig 4*, which was 1 day before the first confirmed case of B.1.617.2. The data fit to the model can be seen in *Fig 5* and *Fig 6* for S dropout and triple positive cases respectively. After linkage with the targeted gene sequencing data we had a sample

size of 35,435 triple positive and 64,514 S dropout hospitalisations. We observe exponential growth in triple positive hospitalisations from the 8th April, which is 13 days later than exponential growth was detected in the testing data. S dropout admissions have been in exponential decay from the 16th January after a peak in mid December, which preceded the second wave of SARS-CoV-2 hospitalisations in England. Wider confidence intervals of S dropout hospitalisations in mid June are a result of the low numbers now observed.

Fig 4 Instantaneous growth rate and doubling times of hospitalisations in England by variants

Fig 5 The maximum likelihood model fit to the S dropout confirmed hospitalisation data

Fig 6 The maximum likelihood model fit to the triple positive confirmed hospitalisation data

RT-PCR Gene Targets - SARS-CoV-2 Positive Tests

Age Group

Analysis of the testing data across age groups, seen in *Appendix A,* illustrates that the earliest reduction in the decay rate for triple positive variant cases was observed in the youngest age groups from the end of January. The earliest exponential growth was observed in the 25-34 age group and this was followed very shortly by exponential growth in triple positive cases in all ages. The 0-24 age group experienced the largest growth rate in triple positive cases reaching a doubling time of 6.30 (95% CI: 6.74, 5.90) days. Step 1a had a stark impact on the triple positive variant growth rate for the 25-44 age group and a negligible impact on the oldest age groups. It was not until Step 2 that the over 75-year-old age groups began to experience exponential growth in triple positive variant cases. We observe slightly wider confidence intervals in the over 75-year-old age group, which is due to smaller case numbers, at this time, producing greater uncertainty. In these age groups we also observe smaller numbers leading to larger confidence intervals in S dropout growth estimates that overlaps into positive growth.

Region

In the regional testing data we can observe that the North East 0.12 (95% CI: 0.12, 0.14), the North West 0.09 (95% CI: 0.05, 0.14), and Yorkshire and the Humber 0.15 (95% CI: 0.13, 0.17) are currently experiencing the largest exponential growth in triple positive cases, which can be seen in *Appendix B*. We do, however, observe modest growth in the S dropout cases in these regions from the start of May with the largest growth observed in the South West where we observe a doubling time of 15.8 (95% CI: 8.00, 811.1) days. Nonetheless, these estimates produce extremely large confidence intervals and the South West is known to have limited reporting for diagnostic RT-PCR gene target CT values, which will lead to greater uncertainty in growth rate calculations. London saw the earliest exponential growth in the triple positive variant cases and appeared to be the most responsive to Step 1a. Conversely, the North West did not experience exponential growth until Step 1b. Step 2 had the greatest impact on the North East, West Midlands, Yorkshire and the Humber, which hitherto had only experienced limited sporadic growth of these cases in the testing data.

Ethnicity

In *Appendix C* we can observe that exponential growth in triple positive cases began in the Indian ethnicity group from late February reaching a doubling time of 11.3 (95% CI: 9.7, 13.5) days, which is now in exponential decay. Moreover, we observed very steep growth in the Pakistani ethnicity group, from the 10th April that coincided with the religious festival of Ramadan and reached a doubling time of 4.8 (95% CI: 4.4, 5.2) days on the 4th May. It is evident that the growth of the triple positive cases since the 7th April has been strongly sustained by the white ethnicity groups with a doubling time of approximately 7 days.

Lower Tier Local Authority

From the subset of LTLAs analysed, full results can be seen in *Appendix D*, we currently observe the shortest doubling time for triple positive cases to be in Blackpool, 3.9 (95% CI: 3.2, 4.8) days, and Kirklees, 5.6 (95% CI: 4.5, 7.4) days. Birmingham experienced the earliest intermittent growth of the triple positive cases shortly prior to Step 1a. However, Bolton and Blackburn with Darwin saw the earliest sustained exponential growth in triple positive cases reaching 0.16 (95% CI: 0.15, 0.18) and 0.14 (95% CI: 0.16, 0.11) respectively before interventions, that included surge testing, increased vaccination, and public health awareness campaigns [27], appeared to slow growth substantially. Nonetheless, the interventions in Burnley and Kirklees [28] had limited success, which is apparent in *Appendix D*. It is striking in the LTLAs in the North West like Manchester the rate of

exponential decay began to reduce from the end of January, which was over a month before the relaxation in NPIs began with Step 1a. This is not observed for S dropout cases and this illustrates that triple positive variant transmission was able to increase despite a strict national lockdown.

RT-PCR Gene Targets - SARS-CoV-2 Positive Hospitalisations

Age Group

In *Appendix E,* we observe that the 65-74 age group saw the earliest initial wave of growth in triple positive variant attributed hospitalisations at the end of March, after Step 1b. Nonetheless, most age groups only experienced exponential growth post Step 2, resulting from the wave of infections caused by earlier NPI restriction easing steps. Akin to the results observed for triple positive tests, we now see corresponding strong growth in hospitalisations for the youngest age groups 0-24: 0.11 (95% CI: 0.07, 0.15) , 25-34: 0.12 (95% CI: 0.08, 0.15) and, 35-44: 0.07 (95% CI: 0.04, 0.11). This is indicative of where the epidemic was growing in the population and therefore, the ages that had seen the largest concentration of infections. The analysis for the over 65-year-old age groups indicates congruent growth and very short doubling times although, there is greater uncertainty in these estimates due to smaller numbers presently observed.

Region

The region that observed the earliest exponential growth in triple positive cases that could be attributed to B.1.617.2 was London, seen in *Appendix F,* followed very shortly by the East of England and Yorkshire and the Humber. This is not consistent with the growth rate analysis of the positive tests for the East of England and Yorkshire and the Humber, which may be indicative of poorer laboratory reporting coverage for diagnostic RT-PCR gene target CT values in these regions that will bias the results to areas of higher coverage. However, this may also be a consequence of triple positive variant infections being less concentrated in the younger ages in the East of England and Yorkshire and the Humber. The wave of triple positive variant attributed hospitalisations observed that occurred across the regions around the introduction of Step 2 are a palpable consequence of Step 1a and 1b with the exception of the South East and South West, which did not experience sustained exponential growth until the effects of the Step 2 restriction easing had impacted transmission. The hospitalisation analysis illustrates the most substantial growth in hospitalisations now observed, with the tightest confidence intervals, can be seen in the North West 0.12 (95% CI: 0.09, 0.15) and also the South East 0.14 (95% CI: 0.08, 0.20). The highest central estimate is in the North East with a doubling time of 3.4 days, but with large confidence intervals that may be related to poor CT value reporting coverage in this area.

Lower Tier Local Authority

Following the early sustained exponential growth in triple positive cases in Bolton and Blackburn with Darwen we also concurrently observe the earliest growth in hospitalisations that can be attributed to triple positive cases in these areas. The strongest growth we presently observe in the LTLAs of concern, included in this analysis, are in Salford and Trafford with doubling times of 3.3 (95% CI: 2.4, 5.8) and 4 (95% CI: 2.5, 10.7) days respectively, with full LTLA analysis in *Appendix G*. However, the tighter confidence intervals in the North West are due to higher proportion of laboratories reporting RT-PCR CT gene target data. We also observe in an LTLA in Yorkshire and the Humber, Kirklees, a comparably short doubling time of 4.2 (95% CI: 2.5, 12.4) days but with larger confidence intervals that are a by-product less diagnostic RT-PCR gene target laboratory reporting in this region.

The Instantaneous Reproduction Number R_t

To parametrise the R_t model we have calculated the minimum, maximum, standard deviation and mean of the values from the Reed, et al., (2021) [25] meta-analysis, with results in *Appendix H,* to create an MC simulation of the distribution. For this model we further calculated the time lag from symptom onset date to specimen date in England that can be seen in *Appendix I*. The results show a marked reduction from the first wave of SARS-CoV-2 in England, from January to May 2020, which may be related to an improved public health message and more effective contact tracing. The sample size used in this model was 1,040,387 and 1,020,664 for S dropout and triple positive variant cases respectively.

Analysis of the instantaneous reproduction number in *Fig* 7 illustrates that from mid-March R_t began to grow for triple positive cases reaching 1.85 by mid-April. The short reduction in R_t for triple positive cases at the end of March is a probable consequence of targeted local interventions [28]. Nonetheless, we observe that B.1.1.7 has been below 1 since January apart from a brief period of growth in March with the reopening of schools. The R_t estimates imply the greater transmissibility for the triple positive dominant variant B.1.617.2 and from the time of the first confirmed case we observe an average reproduction number advantage of 0.45.

Fig 7 The instantaneous reproduction number for s dropout and triple positive cases

Genomic Sequenced and Reflex Assays - SARS-CoV-2 Positive Cases

The sample size for each variant included in this model:

i) B.1.1.7: 225,034
ii) B.1.351: 933
iii) B.1.617.2: 91,960
iv) P.1: 223

Analysis of the sequenced genomic data and reflex assays in *Fig 8* illustrates that since the first detected case of B.1.617.2 in England there has been sustained exponential growth in this variant with a doubling time of 7 days now observed. It is apparent that P.1 has not managed to gain much traction and been in steady decline from the time of first detection and importation. B.1.1.7, as can be observed in the S dropout results, has been in exponential decay since January after a period of high prevalence contributing to the second wave of the SARS-CoV-2 epidemic in the United Kingdom. B.1.351 conversely has seen periods of growth and decay but without a substantial period of sustained growth that would allow this variant to become established and more significant. However, we observed exponential growth in B.1.351 from the start of February during a national lockdown in England, which began to decline after localised NPIs and the introduction of B.1.617.2. There are concerning signs that this variant now continues to experience intermittent growth albeit from a very low baseline. It is not clear that the NPI easing had a substantial impact on increasing B.1.1.7 transmission which, is likely to be related to competition with B.1.617.2 from early 2021.

Fig 8 Instantaneous growth rate and doubling times of confirmed positive cases by the highest priority variants of interest

Discussion

There has been a reduction in the exponential decay rate of triple positive cases since February in England and exponential growth since April. It is evident from *Fig 8* that B.1.617.2 has been the dominant triple positive variant and that other imported variants have thus far failed to compete effectively and have been largely in exponential decay. The results indicate the earliest local outbreaks of B.1.617.2 were concentrated in Bolton and Blackburn with Darwen in the North West of England, which is corroborated by triple positive attributed tests and hospitalisations. Analysis of the testing data at a regional spatial scale, however, indicates the earliest exponential growth of cases that could be attributed B.1.617.2 were in London, the East Midlands, the South East, and the South West. With the North West experiencing exponential growth later than these regions and post Step 1b. Analysis of B.1.617.2 attributed hospitalisations corroborate the early growth in London however, analysis suggests Yorkshire and the Humber and the East of England may have experienced some of the earliest outbreaks. This may be a consequence of these regions having a greater concentration of cases in older age groups. However, these regions also experience less complete laboratory reporting for diagnostic RT-PCR CT gene target data, which may disguise possible outbreaks. S dropout cases, the proxy for B.1.1.7, have been largely in exponential decay since the January national lockdown in the UK with isolated areas of growth. Further research should focus on whether we can observe greater growth in B.1.1.7 in the areas that have a larger proportion of unvaccinated individuals.

A limitation of this study is that it does not directly include the vaccination status of the infections in the analysis and therefore, the analysis employs age as a proxy for vaccination status. This is due to limitations in sequencing and RT-PCR CT gene target data coverage, which after linkage with vaccination status preclude meaningful analysis for the growth and reproduction number of the groups included. The very high rates of vaccination in the oldest age groups seen in *Table 1* and *Table 2* illustrate a clear stratification between ages: by the end of June over 91.1% of those over 60 had received two doses of the vaccination and that most individuals under the age of 50 had not received their second dose of the vaccination by July.

Month	18-29	30-39	40-49	50-59	60+
Dec	0.5	0.8	1.2	1.4	5.3
Jan	4.6	6.0	8.1	10.2	40.0
Feb	9.2	12.6	18.7	28.8	79.3
Mar	15.9	21.9	36.5	81.5	91.8
Apr	17.3	24.9	55.5	84.4	92.4
May	21.8	47.8	74.2	85.9	92.9
Jun	50.2	63.1	76.9	86.7	93.2

Table 1 Proportion of each age group who have recived their first vaccination by the end of each month.

Month	18-29	30-39	40-49	50-59	60+
Dec	0.0	0.0	0.0	0.0	0.0
Jan	0.1	0.2	0.3	0.4	2.6
Feb	0.3	0.5	0.7	0.8	2.9
Mar	2.3	3.3	4.5	5.6	15.7
Apr	6.3	8.4	12.0	17.3	62.0
May	11.9	16.9	27.3	54.7	87.4
Jun	16.7	24.9	50.1	81.9	91.1

Table 2 Proportion of each age group who have recived their second vaccination by the end of each month.

The arrival of a VOC can result in higher rates of testing and sequencing for specific LTLA geographic locations. However, this targeted approach was not conducted to a considerable extent for B.1.617.2 due to how quickly transmission became widespread in England. Nonetheless, the sporadic growth we observe in B.1.351, which has failed to maintain growth in the absence of importations, has been influenced by surge testing and enhanced contact tracing of locations where this variant has been found. Targeted surge testing in response to importations of B.1.351 began in 2021 for LTLAs on the 9th February in Lambeth [29] and ran until the 16th March in Sandwell [30]. These interventions showed success in slowing the growth of this variant in April, apparent in *Fig 8*, before exponential growth was again observed in June. The primary limiting factor for the analysis conducted is geographic reporting bias, with the North West observing the highest proportion of laboratories reporting diagnostic RT-PCR CR with gene target data across the period of this study, which can be seen in *Fig 9*. Conversely, the lowest levels of coverage have been observed in the South West, East of England, and East Midlands. Although, this geographic bias is temporally variable as changes in laboratory capabilities evolve over the pandemic.

Fig 9 A map of the proportion of RT-PCR tests with gene target data by English region from October 2020 to July 2021

We observe in *Appendix C* that exponential growth in triple positive cases was initially seen in February within the Indian ethnicity group, which was due to importations of the B.1.617.2 variant. Growth in this group has subsequently declined and we can see from our results in early April that B.1.617.2 has now been largely sustained in the white and black British ethnicity groups. This illustrates from April the variant was no longer dependent upon importations to maintain exponential growth in England. Interestingly, we observe exponential growth in the Pakistani ethnicity group around the holiday of Ramadan and this is indicative of the significance of public and religious events in driving significant outbreaks of SARS-CoV-2. Further exemplified by the Christmas period in the UK when we observed similar growth in the B.1.1.7 variant. We can discern that the first phase of the relaxation of the national lockdown Step 1a [23], the reopening of schools, appeared to have the earliest impact on London for triple positive variant growth, albeit this region was already experiencing the beginnings of exponential growth prior to this step. Exponential growth was observed across all regions before the final easing of national restrictions in Step 3. The earliest targeted local NPIs to contain B.1.617.2 variant began in Bolton on the 30th Match 2021 [31] and on the 14th May 2021 in Blackburn with Darwen [32] with further measures brought in later for these LTLAs. The targeted interventions showed some success in precluding triple positive case growth in these areas, evidenced by *Appendix D*. Later interventions that targeted further LTLAs to limit the transmission of B.1.617.2 [28] experienced limited success which is apparent for instance, in Kirklees and Burnley (seen in *Appendix D*). Therefore, once transmission of the B.1.617.2 variant became more widespread in England the efficacy of these targeted approaches including surge testing, vaccination campaigns and travel restrictions was limited. The East Midlands, that had the slowest prior growth of S dropout cases in December, had a subsequent wave of exponential growth in March that was not present or weakly observed in other regions. The high level of prevalence for SARS-CoV-2 now observed in England is facilitating sporadic growth of S dropout cases that can be seen in the North West, North East, South West and, Yorkshire and the Humber where it is growing from a very small baseline.

The implications for the strong growth in triple positive cases followed by similar patterns in the hospitalisations is very significant for the implications of vaccine effectiveness. We presently observe significant growth in the younger age groups that have a low infection hospitalisation rate for SARS-CoV-2 infection [33] and have only largely received one dose of a vaccine at the time of this study [34]. Nonetheless, growth within the hospitalisations will be largely indicative of the demographic groups where infections are primarily concentrated at that time. Significantly, we observe early trends from late March in the, largely doubly vaccinated, over 65-year-old groups, which is now particularly pronounced in the 75-84 group where we can observe a doubling time of almost 4 days. The regions that are seeing the most concerning growth in triple positive hospitalisations are the North West, North East and South East, which is very much in line with where we observed the earlier growth in positive cases, although it is now evident that the variant is in exponential growth throughout England.

We can observe in *Fig 7* that the R_t number showed some growth in the B.1.1.7 attributed infections in March when overall incidence began to initially surge across England in reaction to Step 1a. However, since this time it has hovered around 0.8 and we observe largely exponential decay across the country, with some sporadic growth evidenced by the regional analysis from *Appendix B*. If R_t continues to be < 1 then transmission of this variant within England is likely to decay to insignificance. There is believed to be an increased risk of within household transmissibility of 60% for B.1.617.2 [27] relative to B.1.1.7. Similarly, we find a transmission advantage for B.1.617.2, with the mean difference for R_t found in this study to be 0.45.

Conclusion

To conclude, the sustained exponential growth in cases of sequenced B.1.617.2 and the exponential decay of other triple positive variants illustrates that this variant now causes most of the RT-PCR triple positive case transmission in England. The reduction in the exponential decay rate for confirmed triple positive variant cases in February 2021 indicate that B.1.617.2 appeared quite a lot earlier than the first confirmed case in March and the relaxation of NPIs coincided with exponential growth in this variant. We can see that growth of B.1.617.2 was initially concentrated in the LTLAs Bolton and Blackburn with Darwen in the North West. However, regional analysis suggested earlier and greater geographic dispersion with Yorkshire and the Humber, the East of England, and London experiencing the earliest exponential growth for B.1.617.2 attributed hospitalisations. The B.1.617.2 variant has now spread across the country with a doubling time of 8 (95% CI: 6.9, 9.1) days for hospitalisations. The study illustrates a substantial transmission advantage for the B.1.617.2 variant relative to B.1.1.7 and we estimate the reproduction number advantage is around 0.45. There have been small indications of growth in B.1.1.7 with R_t above 1 in March, which is in line with increases in B.1.617.2 but it is now clear that there has been a replacement of the predominant B.1.1.7 variant. We have observed some worrying trends in B.1.351 although it has failed to gain traction and a sustained enough period of growth for this variant to become a substantial public health concern.

Contributorship Statement

TW conceived the idea of the article. TW wrote the article. TW, LP, IH, FX and AG developed the model methodology. AG and TW created the graphical representations. LP, AJ, TW, and IH reviewed the final draft.

Conflict of Interest

The authors have declared that no competing interests exist. The authors were employed by the Department of Health and Social Care but received no specific funding for this study.

Ethics statements

Patient consent for publication

Not applicable.

Ethics approval

This study was conducted in line with national data regulations. It only employed and accessed fully anonymised population level data from Public Health England in a secure research environment.

Data Availability Statement

To access the data used for this study, an application can be made to Public Health England, Department of Health and Social Care. Data requests can be made to the Office for Data Release (https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data) and contacting odr@phe.gov.uk. All requests to access data are reviewed by the ODR and are subject to strict confidentiality provisions in line with the requirements of:

- the common law duty of confidentiality
- data protection legislation (including the General Data Protection Regulation)
- 8 Caldicott principles
- the Information Commissioner's statutory data sharing code of practice
- the national data opt-out programme

References

[1] WHO, “Archived: WHO Timeline - COVID-19,” Geneva, 2020.

[2] R. Wang, Y. Hozumi, Y.-H. Zheng, C. Yin and G.-W. Wei, “Host Immune Response Driving SARS-CoV-2 Evolution,” *Viruses,* vol. 12, no. 10, p. 1095, 2020.

[3] W. Jo, C. Drosten and J. Drexler, “The evolutionary dynamics of endemic human coronaviruses,” *Virus Evol ,* vol. 7, no. 1, 2021.

[4] C. Yin, “Genotyping coronavirus SARS-CoV-2: methods and implications,” *Genomics,* vol. 112, no. 5, pp. 3588-3596, 2020.

[5] Department of Health and Social Care, “Further measures in additional areas to tackle Delta (B1.617.2) variant,” 8 6 2021. [Online]. Available: https://www.gov.uk/government/news/further-measures-in-additional-areas-to-tackle-delta-b16172-variant.

[6] A. Kucharski, N. Davies, R. Eggo and S. Funk, “Dynamics of B.1.617.2 in England NHS regions from importations, traveller-linked and non-traveller-linked transmission,” Report for SPI-M, 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993232/S1272_LSHTM_Modelling_Paper_B.1.617.2.pdf.

[7] Department of Health and Social Care, “UK COVID-19 vaccines delivery plan,” 2021. [Online]. Available: https://www.gov.uk/government/publications/uk-covid-19-vaccines-delivery-plan/uk-covid-19-vaccines-delivery-plan#prioritisation-1.

[8] A. Pottegård, L. C. Lund, Ø. Karlstad, J. Dahl, M. Andersen, J. Hallas, Ø. Lidegaard, G. Tapia, H. L. Gulseth, P. L.-D. Ruiz, S. V. Watle, A. P. Mikkelsen, L. Pedersen and H. T. S, “Arterial events, venous thromboembolism, thrombocytopenia, and bleeding after vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: population based cohort study,” *BMJ,* vol. 373, 2021.

[9] S. Madhi, V. Baillie, C. L. Cutland, M. Voysey, A. Koen, L. Fairlie, F. Paeds, S. Padayachee, K. Dheda, S. Barnabas and Q. E. Bhorat, “Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant,” *The New England Journal of Medicine,* vol. 384, no. DOI: 10.1056/NEJMoa2102214, pp. 1885-1898, 2021.

[10] I. Ferreira, R. Datir, S. Kemp, G. Papa, P. Rakshit, S. Singh, B. Meng, R. Pandey, K. Ponnusamy, V. Radhakrishnan, T. I. S.-C.-2. G. C. (INSACOG) and T. C.-N. B. COVID-19, “ SARS-CoV-2 B.1.617 emergence and sensitivity to vaccine-elicited antibodies,” *BioRxiv ,* no. 2021.05.08.443253, 2021.

[11] M. Hoffmann, H. Hofmann-Winkler, N. Krüger, A. Kempf, I. Nehlmeier, L. Graichen, A. Sidarovich, A.-S. Moldenhauer, M. Winkler, S. Schulz, H.-M. Jäck, M. Stankov, G. Behrens and S. Pöh, “SARS-CoV-2 variant B.1.617 is resistant to Bamlanivimab and evades antibodies induced by infection and vaccination,” *bioRxiv,* no. 2021.05.04.442663., 2021.

[12] D. Planas, D. Veyer, A. Baidaliuk, M. M. Rajah, C. Planchais, M. Prot, F. Gallais, P. Gantner, A. Velay, J. L. Guen, N. Kassis-Chikhani, D. Edriss and L. Belec, “Reduced sensitivity of infectious SARS-CoV-2 variant B.1.617.2 to monoclonal antibodies and sera from convalescent and vaccinated individuals,” *biorxiv,* no. doi.org/10.1101/2021.05.26.445838, 2021.

[13] J. L. Bernal, N. Andrews, C. Gower, E. Gallagher, R. Simmons, S. Thelwall, J. Stowe, E. Tessier, N. Groves, G. Dabrera, R. Myers, C. Campbell, G. Amirthalingam, M. Edmunds, M. Zambon and K. Brown, "Effectiveness of COVID-19 vaccines against the B.1.617.2 variant," *medRxiV,* no. doi.org/10.1101/2021.05.22.21257658, 2021.

[14] Public Health England, "Variants: distribution of cases data," 12 05 2021. [Online]. Available: https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-cases-data.

[15] Public Health England, "SARS-CoV-2 variants of concern and variants under investigation in England," 7 05 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984274/Variants_of_Concern_VOC_Technical_Briefing_10_England.pdf.

[16] S. Wood, Generalized Additive Models, New York: Chapman and Hall/CRC, 2017.

[17] S. Wood, "Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL," *R package version https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/mgcv-package.html,* pp. 1-8, 2018.

[18] S. Wood, "Generalized Additive Models: an introduction with R," *CRC,* 2006.

[19] P. Eilers and B. Marx, "Flexible Smoothing with B-splines and Penalties," *Statistical Science,* vol. 11, no. 2, pp. 89-121, 1996.

[20] S. Wood, "Thin plate regression splines," *J.R.Statist.Soc.B ,* vol. 65, no. 1, pp. 95-114, 2003.

[21] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Solobev spaces," in *Construction theory of functions of several variables*, Berlin, Springer, 1977, pp. 85-100.

[22] L. Pellis, F. Scarabel, H. Stage, C. Overton, L. Chappell, K. Lythgoe, E. Fearon, E. Bennett, J. Curran-Sebastian, R. Das, M. Fyles, H. H. Lewkowicz, X. Pang, B. Vekaria and I. Hall, "Challenges in control of Covid-19: short doubling time and long delay to effect of interventions," *Philosophical Transactions of the Royal Society,* 2021.

[23] Cabinet Office, "COVID-19 Response - Spring 2021 (Summary)," 2021. [Online]. Available: https://www.gov.uk/government/publications/covid-19-response-spring-2021/covid-19-response-spring-2021-summary#step-1---8-and-29-march.

[24] A. Cori, N. Ferguson, C. Fraser and S. Cauchemez, "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics," *Am J Epidemiology,* vol. 178, no. 9, pp. 1505-1512, 2013.

[25] I. Reed, E. Walker and E. Landguth, "SARS-CoV-2 Serial Interval Variation, Montana, USA, March 1–July 31, 2020," *Emerging Infectious Disease,* vol. 27, no. 5, p. 1486–1491, 2021.

[26] T. Ward and A. Johnsen, "Understanding an Evolving Pandemic: An Analysis of the Clinical Time Delay Distributions of COVID-19 in the United Kingdom Plos One," *Plos One,* 2021.

[27] Public Health England, "Confirmed cases of COVID-19 variants identified in UK," 2021. [Online]. Available: https://www.gov.uk/government/news/confirmed-cases-of-covid-19-variants-identified-in-uk.

[28] Department of Health and Social Care, "Further measures in additional areas to tackle Delta (B1.617.2) variant," 2021. [Online]. Available: https://www.gov.uk/government/news/further-measures-in-additional-areas-to-tackle-delta-b16172-variant.

[29] Department of Health and Social Care, “Surge testing to be deployed in areas around Lambeth in connection to new variant,” 2021. [Online]. Available: https://www.gov.uk/government/news/surge-testing-to-be-deployed-in-areas-around-lambeth-in-connection-to-new-variant.

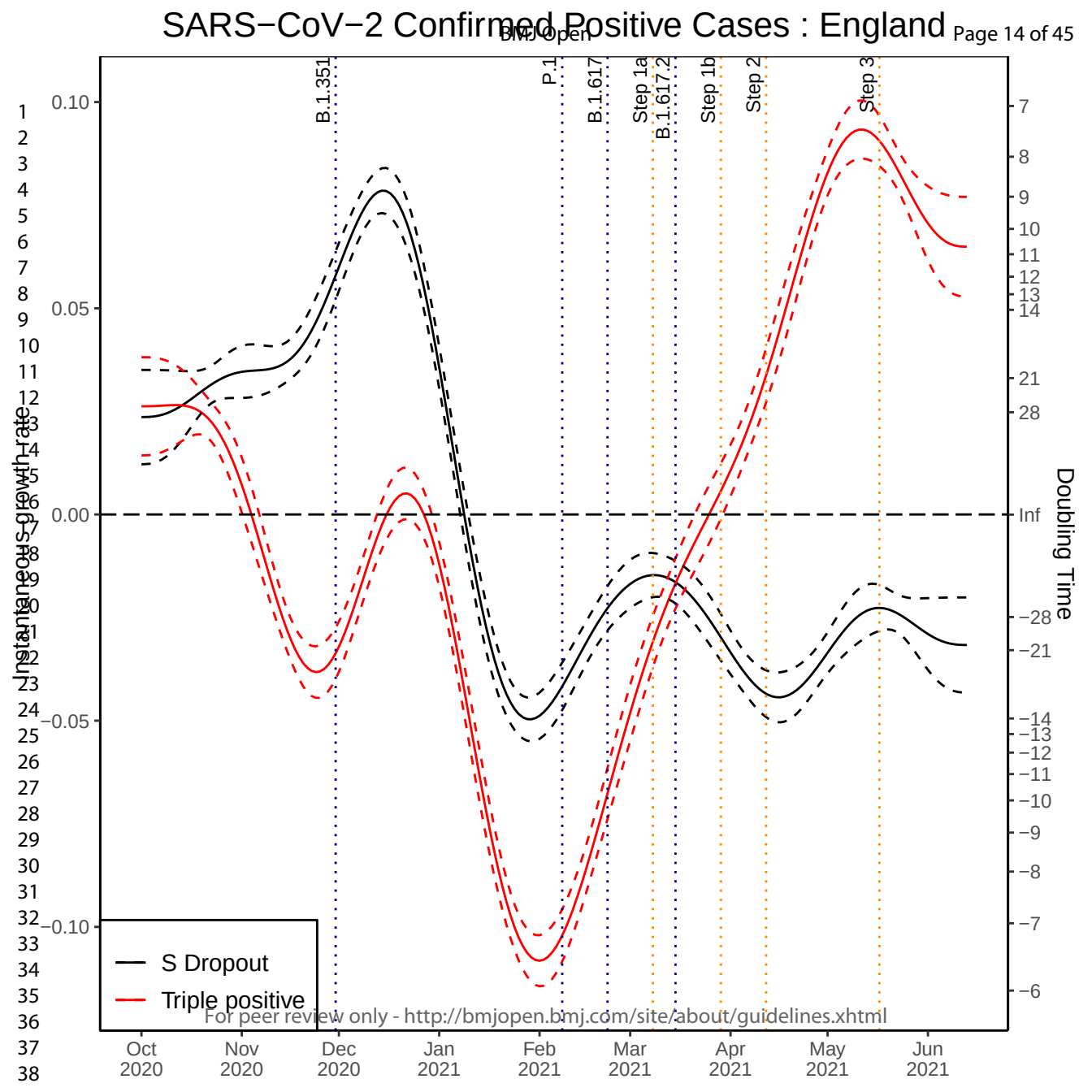
[30] Department of Health and Social Care, “Surge testing to be deployed in targeted areas in Sandwell,” 2021. [Online]. Available: https://www.gov.uk/government/news/surge-testing-to-be-deployed-in-targeted-areas-in-sandwell.

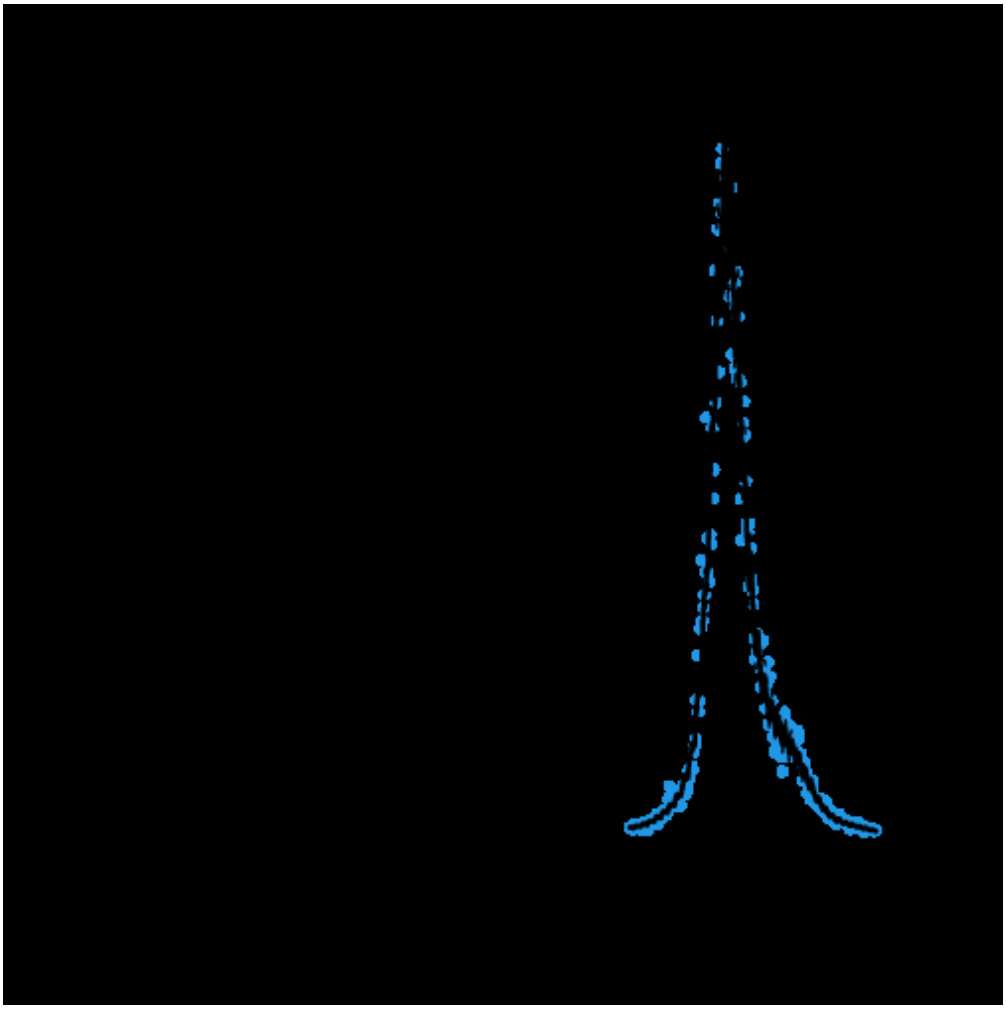
[31] Department of Health and Social Care, “Surge testing to be deployed in targeted areas in Bolton,” 2021. [Online]. Available: https://www.gov.uk/government/news/surge-testing-to-be-deployed-in-targeted-areas-in-bolton.

[32] Department of Health and Social Care, “Surge testing to be deployed in Blackburn with Darwen,” 2021. [Online]. Available: https://www.gov.uk/government/news/surge-testing-to-be-deployed-in-blackburn-with-darwen.

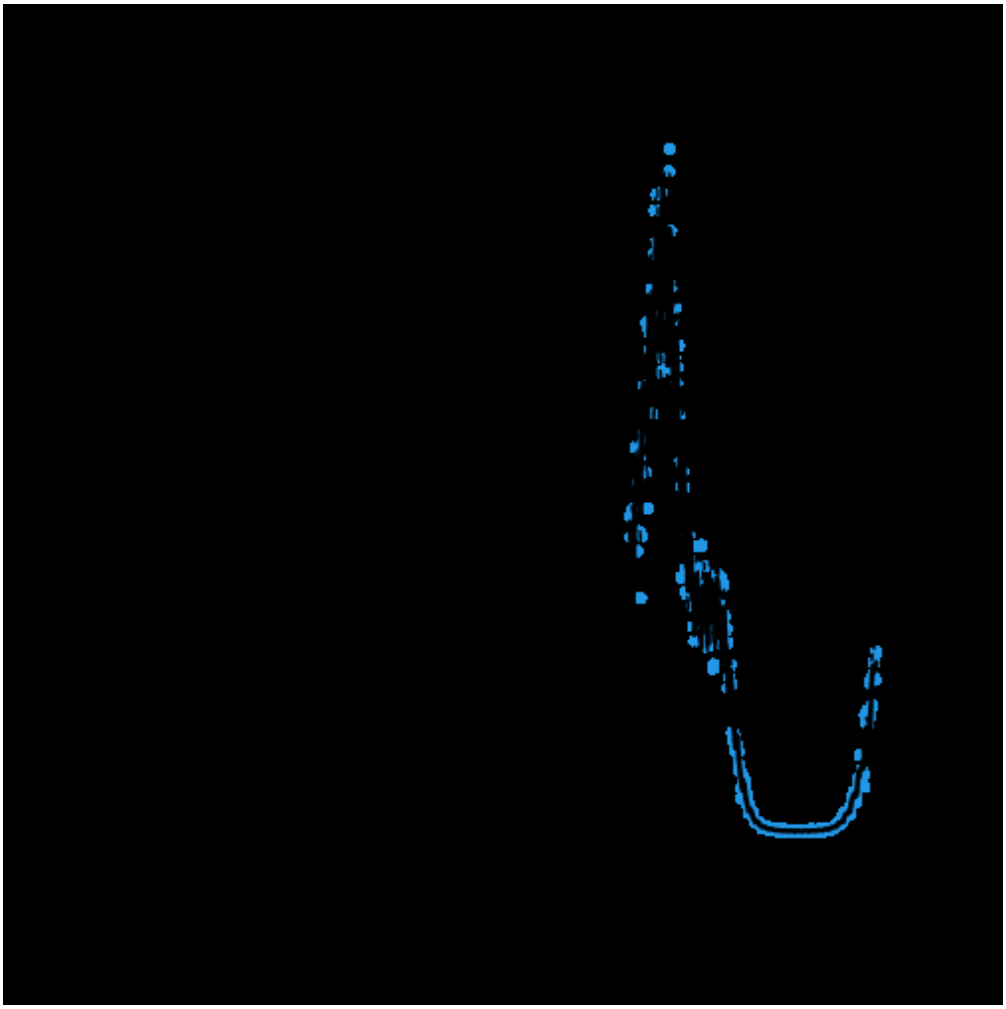
[33] P. Birrell, J. Blake, E. v. Leeuwen, M. B. U. C.-1. W. Group and D. D. Angelis, “COVID-19: nowcast and forecast,” 2021. [Online]. Available: https://www.mrc-bsu.cam.ac.uk/now-casting/nowcasting-and-forecasting-15th-july-2021/.

[34] Public Health England, “COVID-19 vaccination first phase priority groups,” 2021. [Online]. Available: https://www.gov.uk/government/publications/covid-19-vaccination-care-home-and-healthcare-settings-posters/covid-19-vaccination-first-phase-priority-groups.





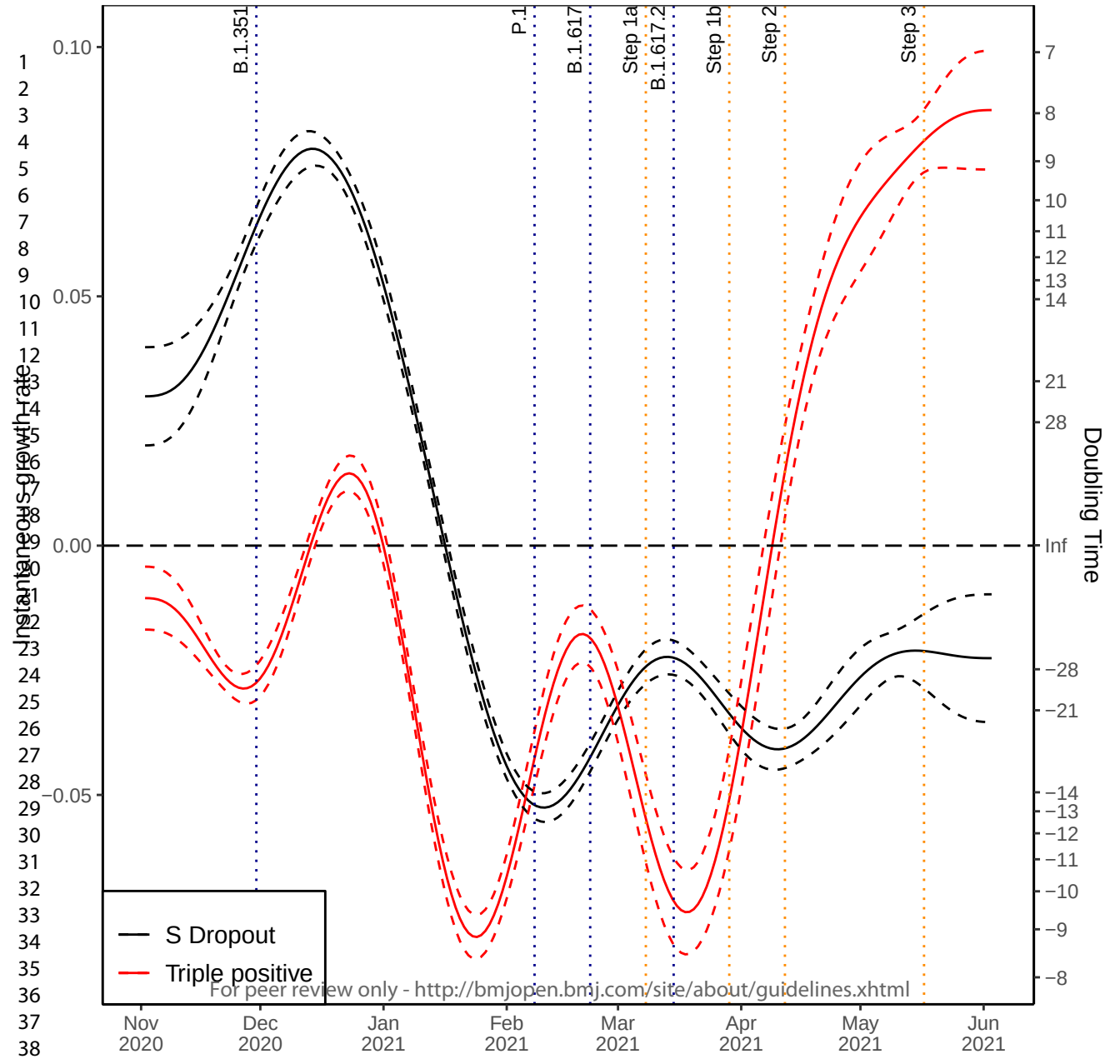
336x336mm (38 x 38 DPI)

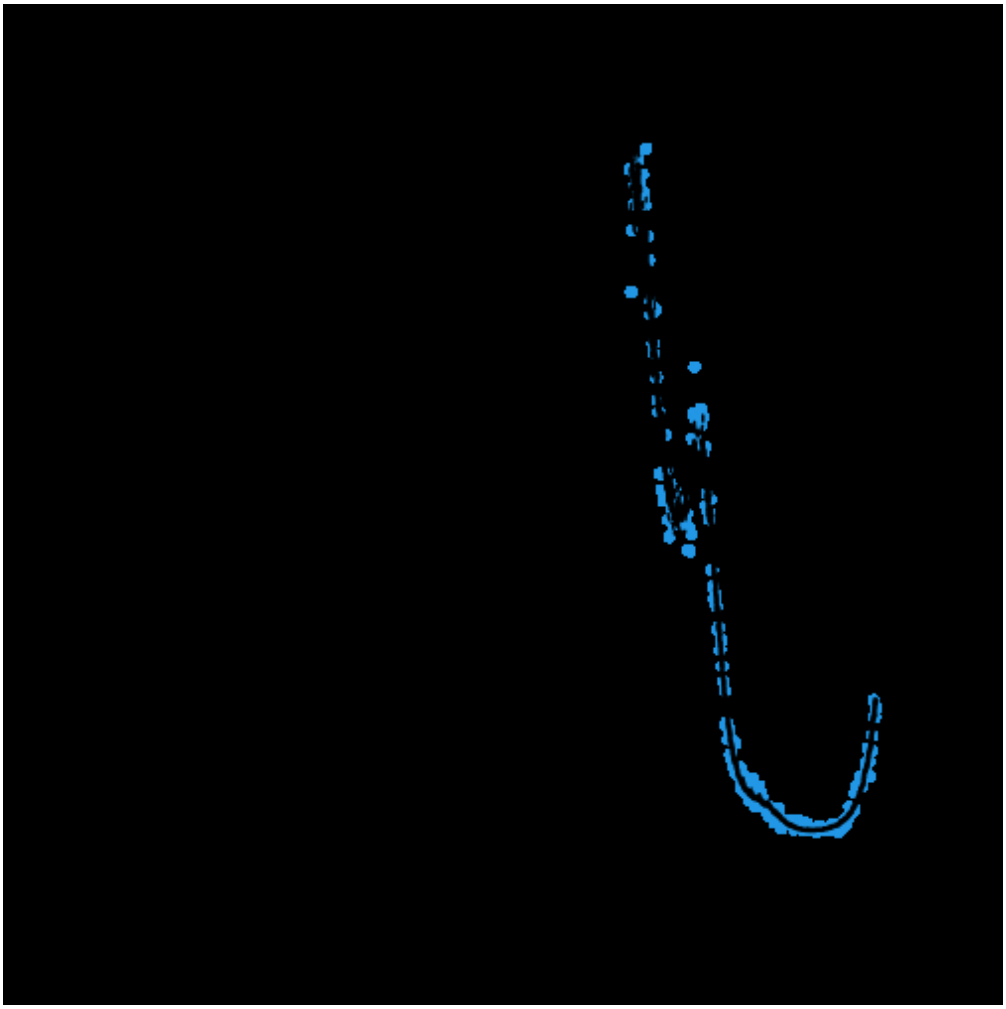


336x336mm (38 x 38 DPI)

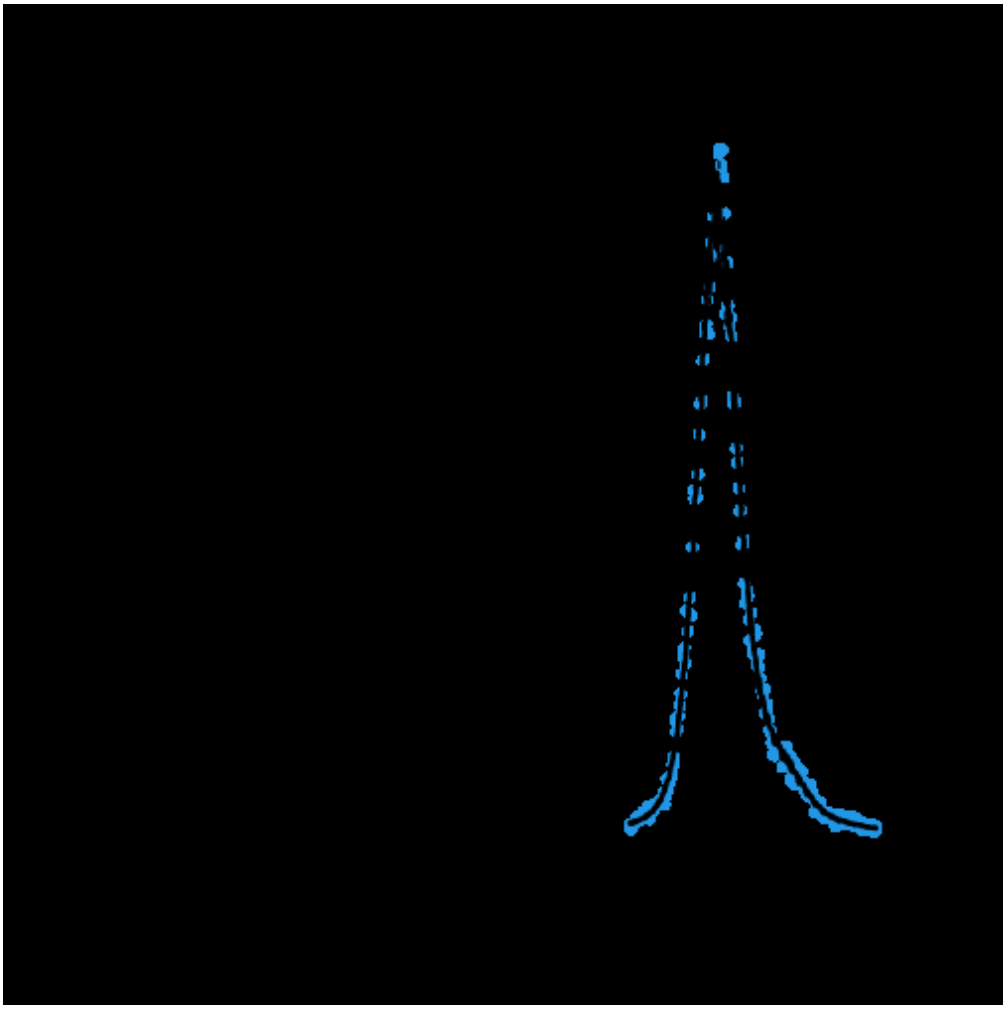
SARS CoV−2 Hospitalisations : England

BMJ Open

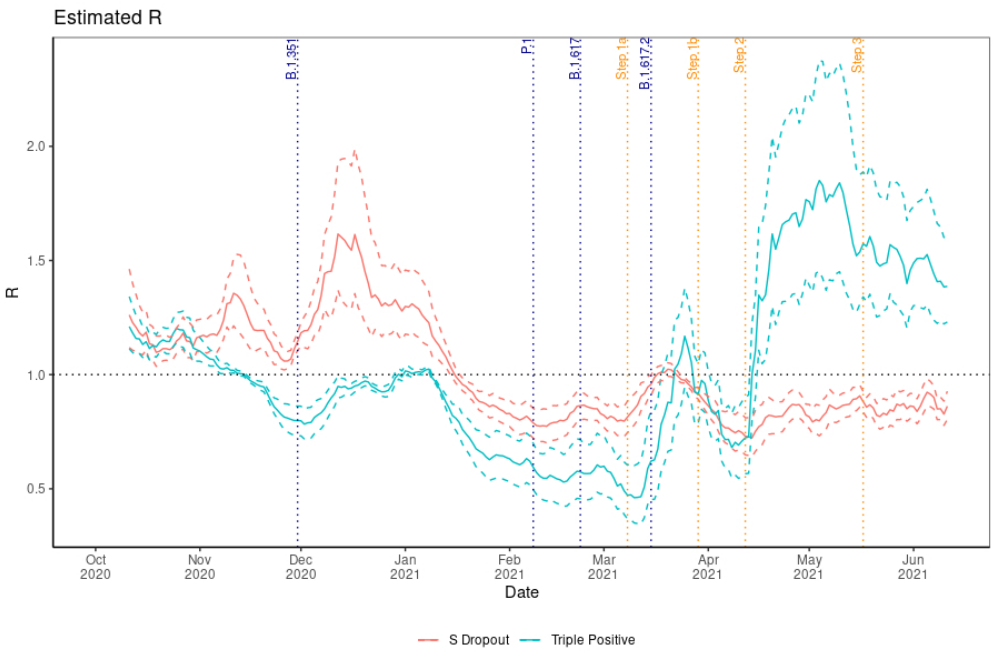




336x336mm (38 x 38 DPI)

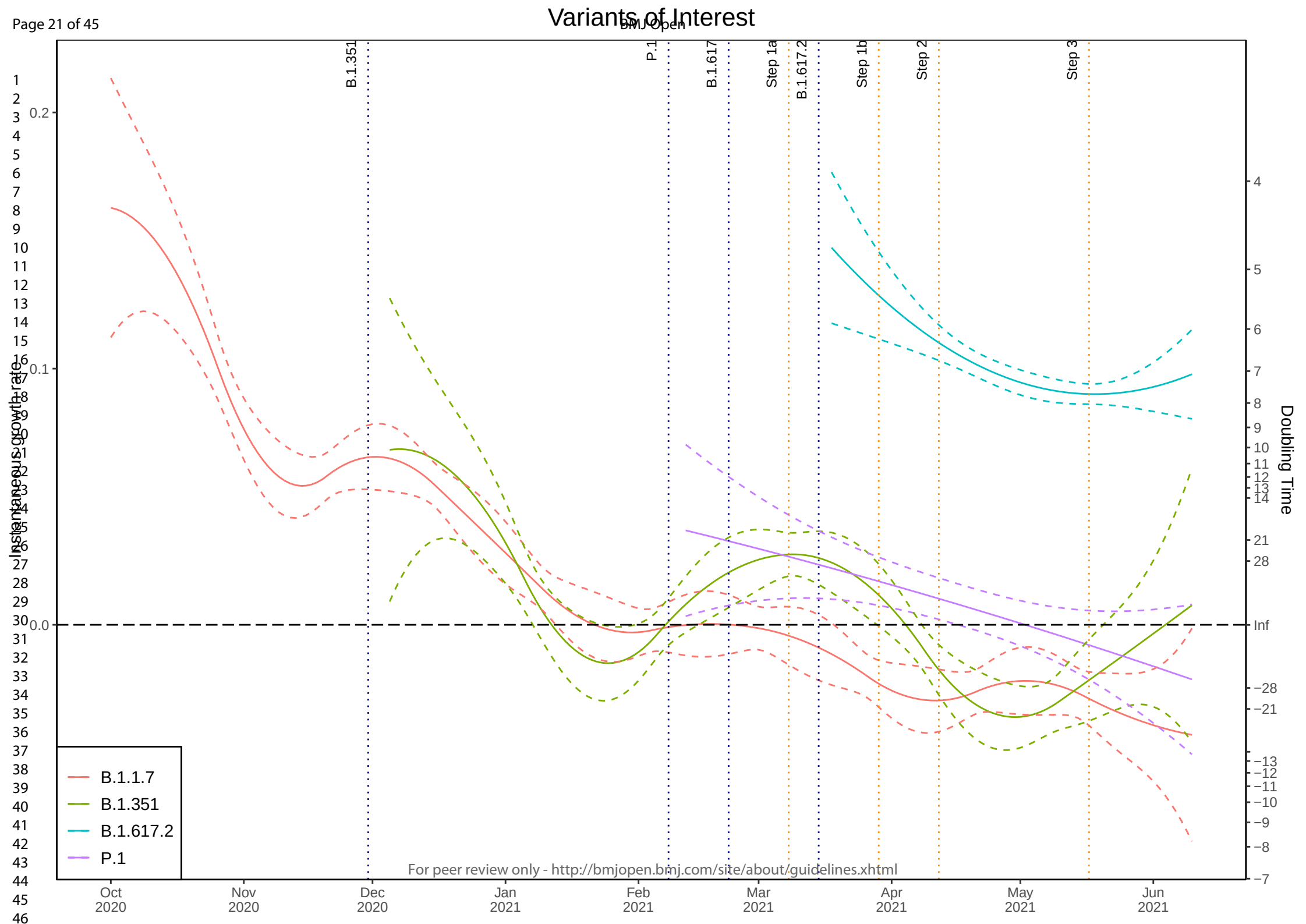


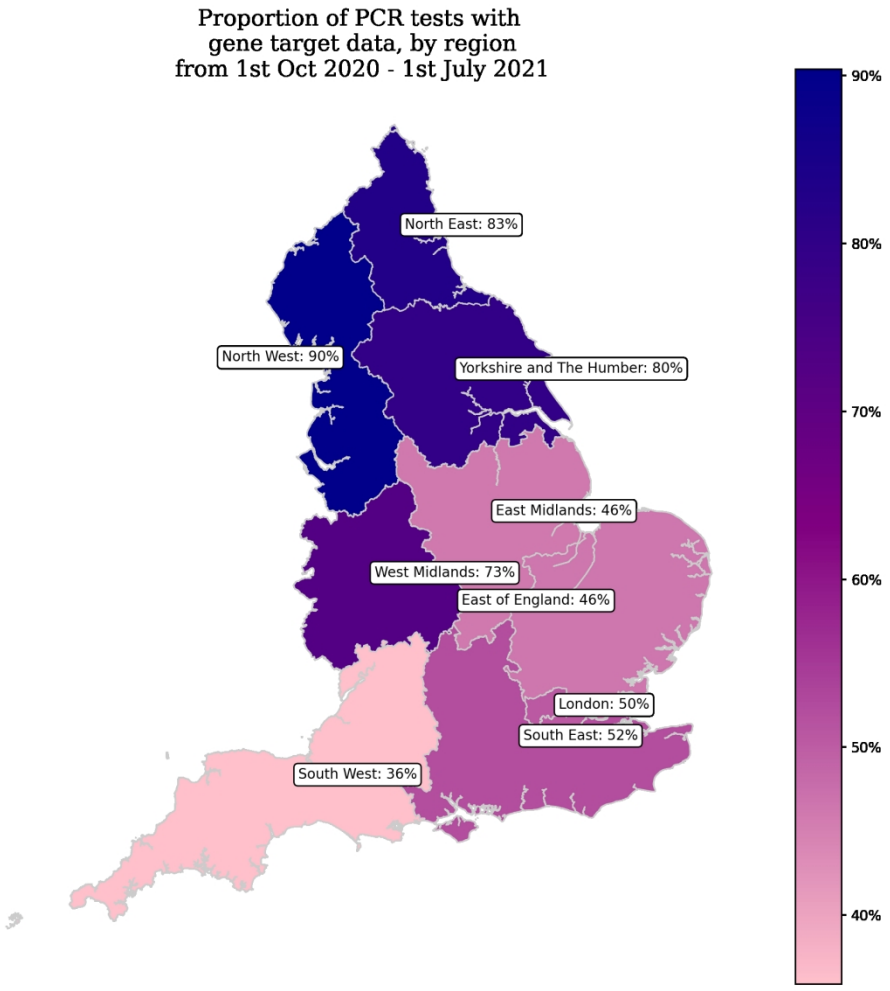
336x336mm (38 x 38 DPI)



601x401mm (38 x 38 DPI)

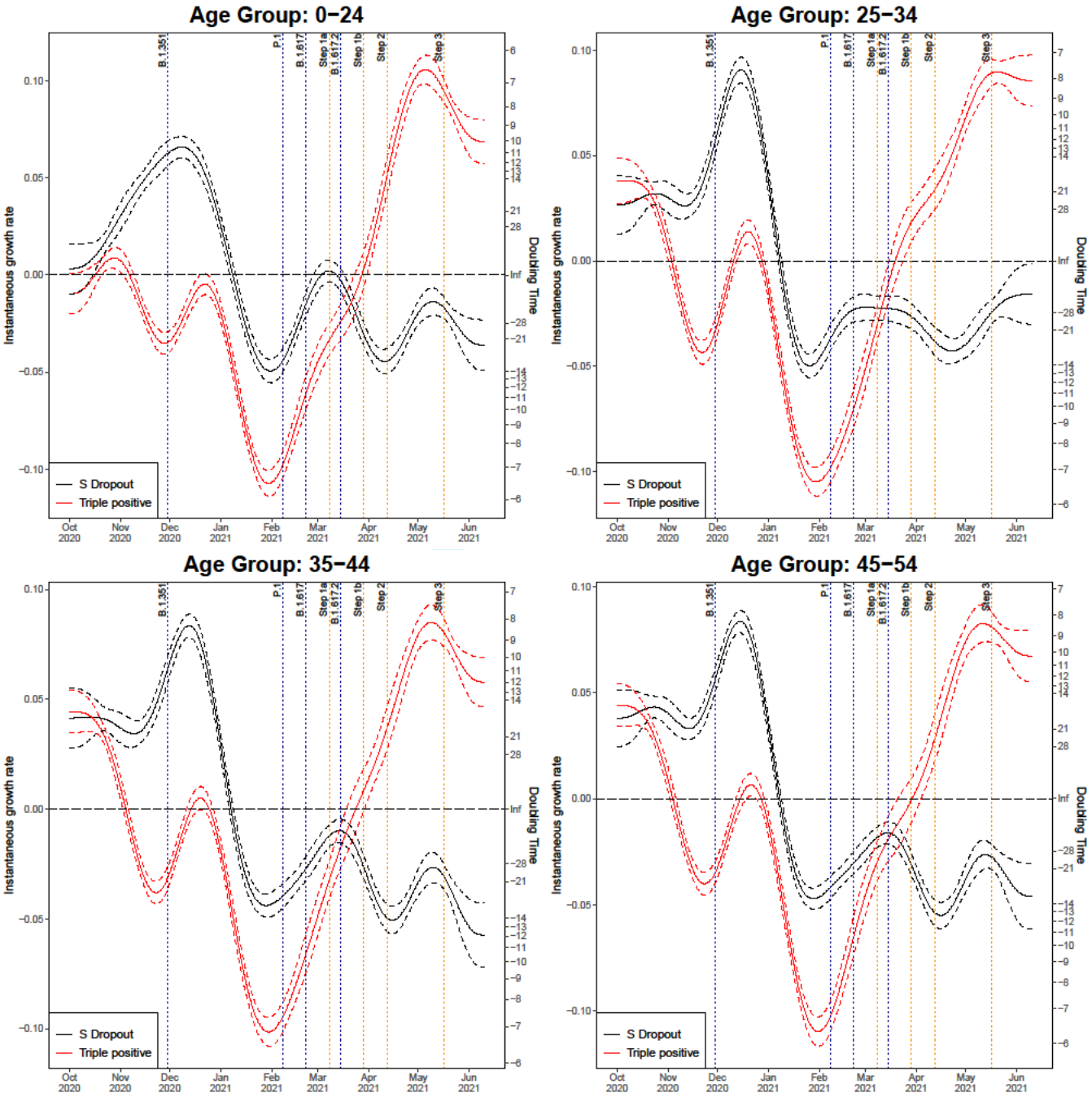
Variants of Interest

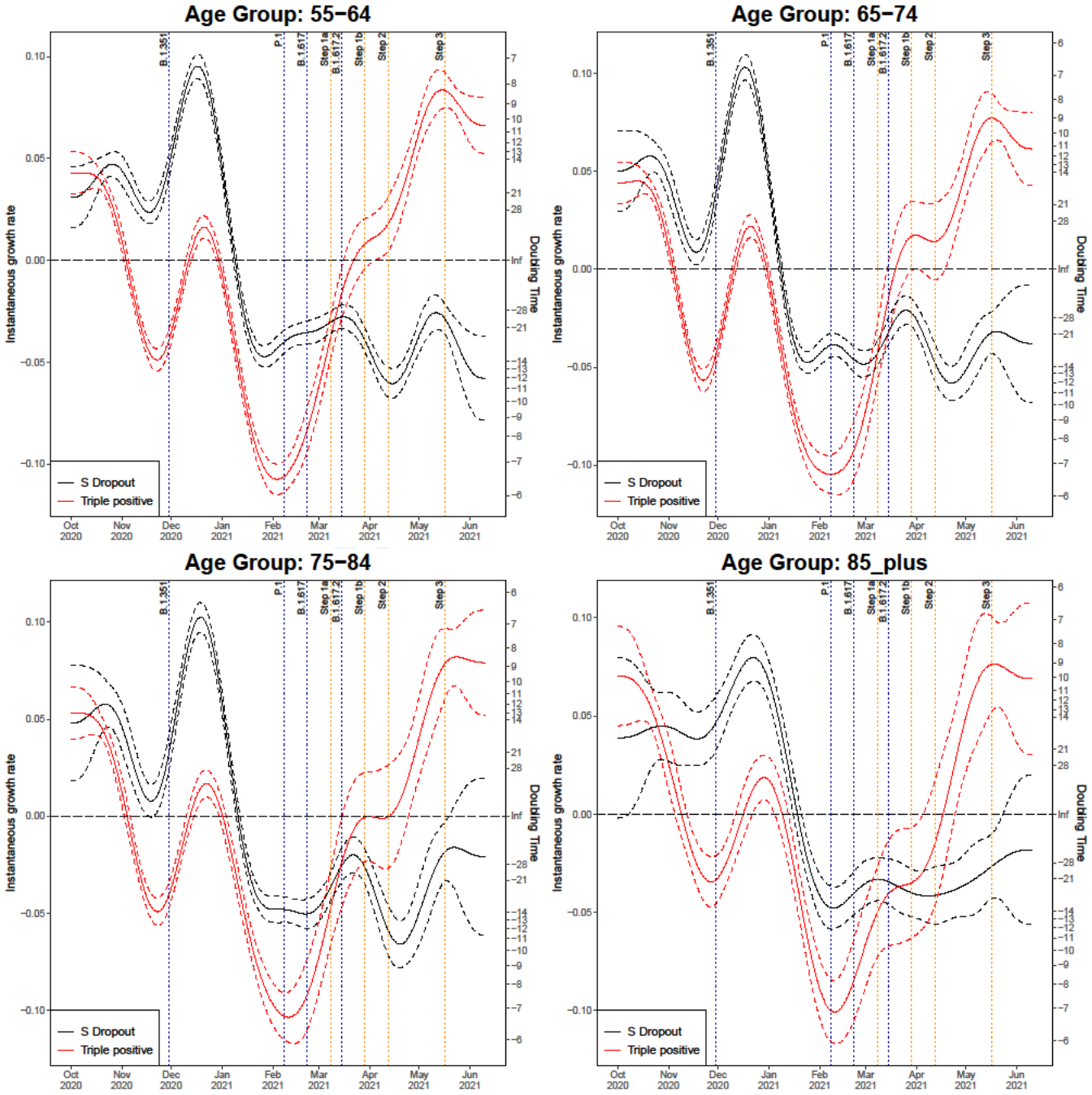


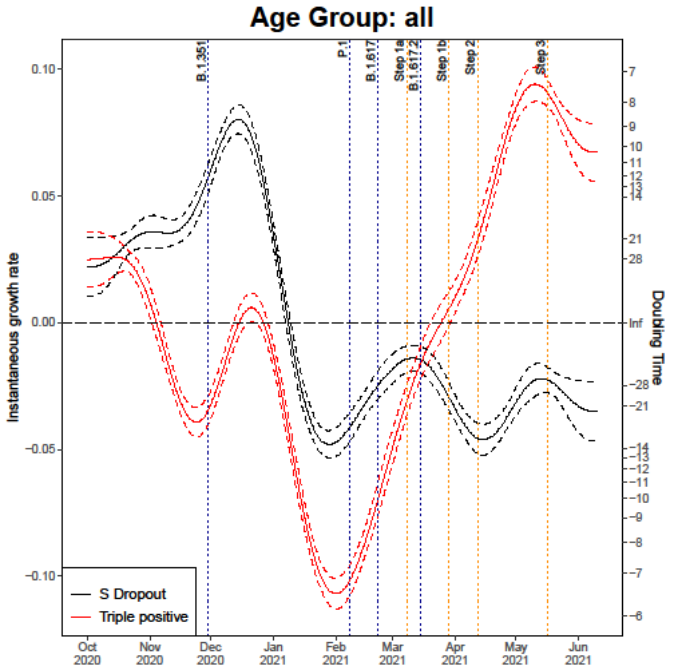


648x648mm (78 x 78 DPI)

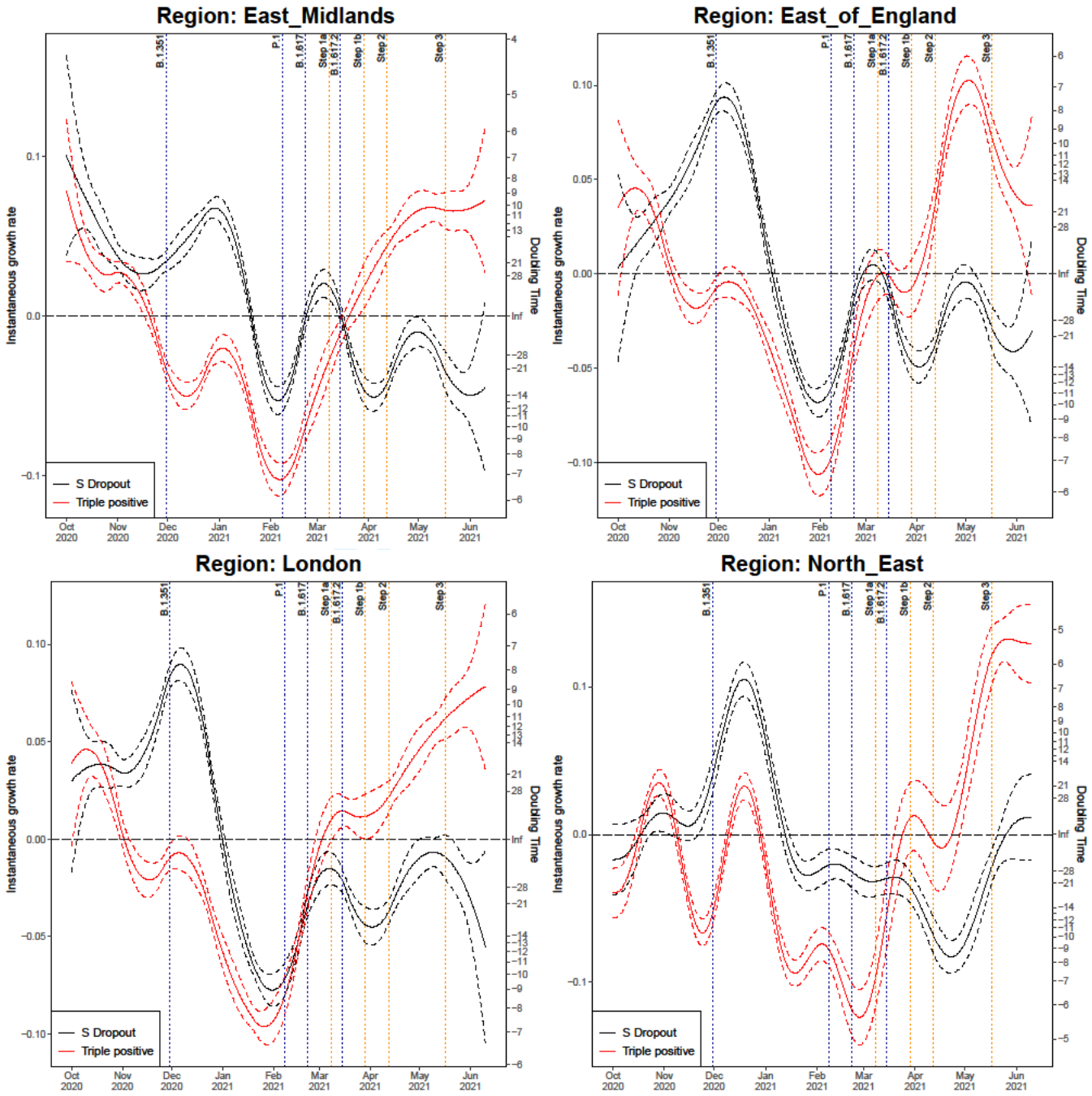
Appendix A: Cases Age Groups

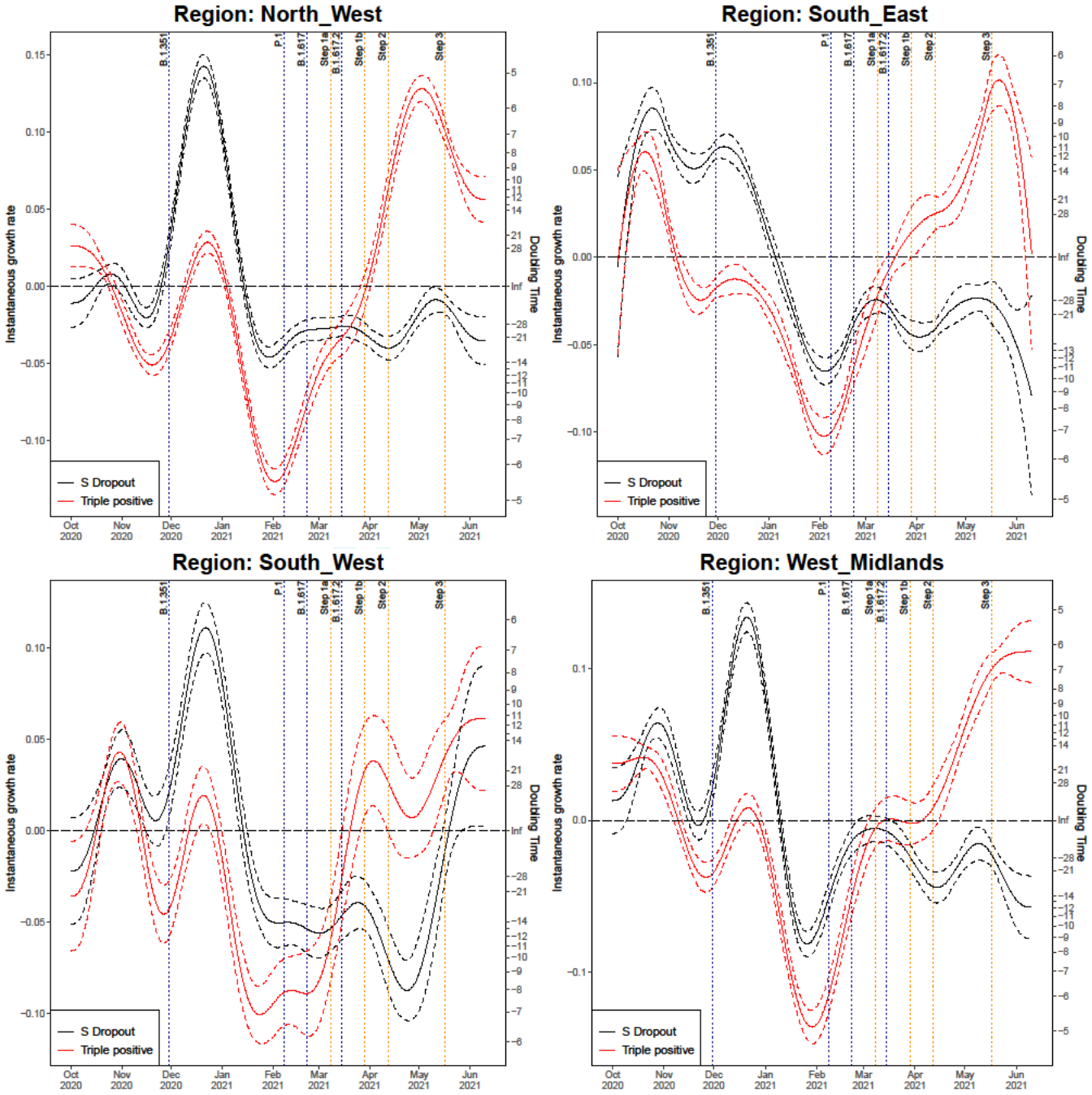


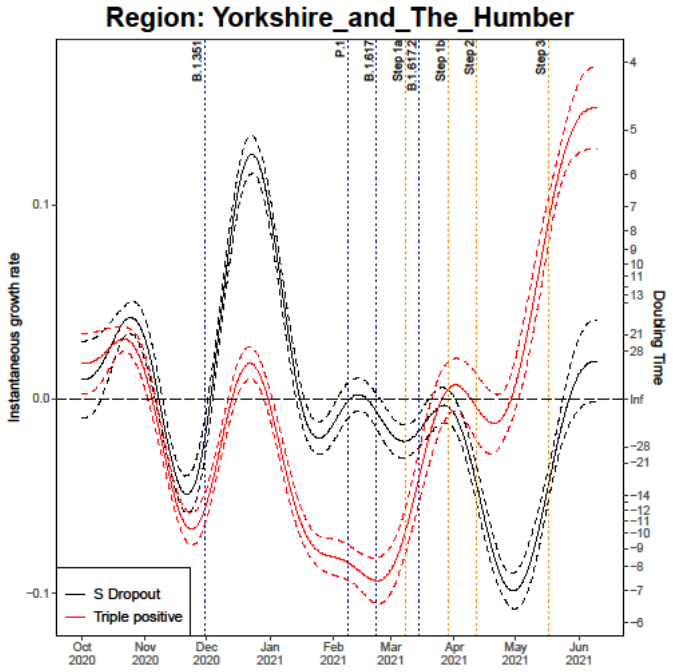




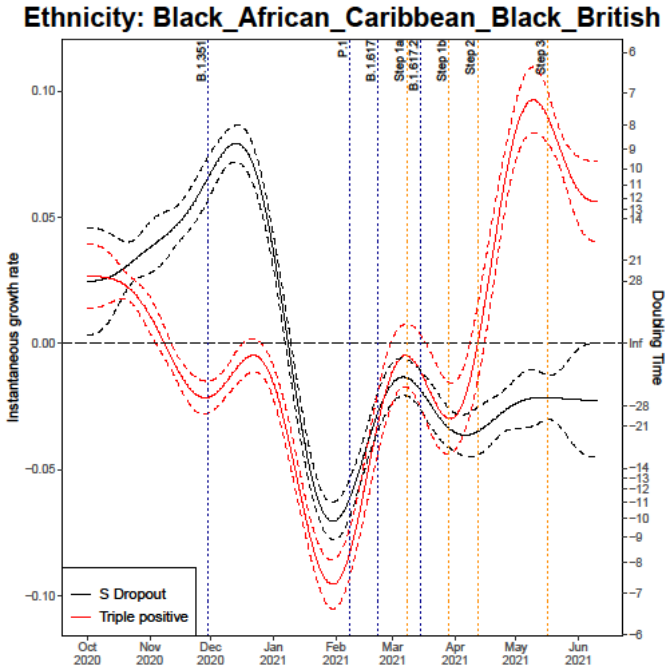
Appendix B: Cases Regions

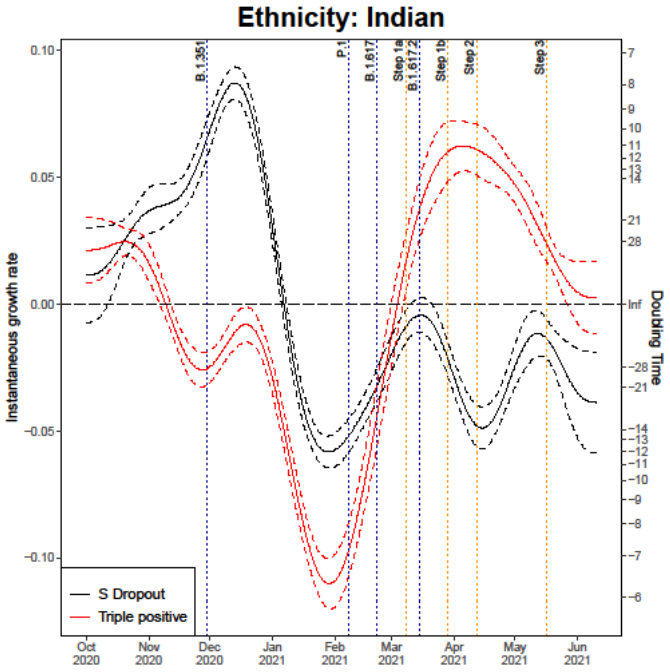


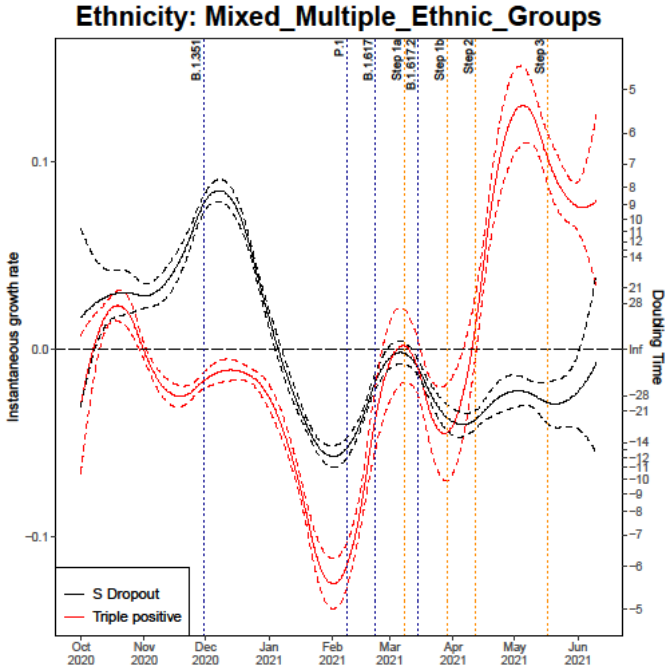


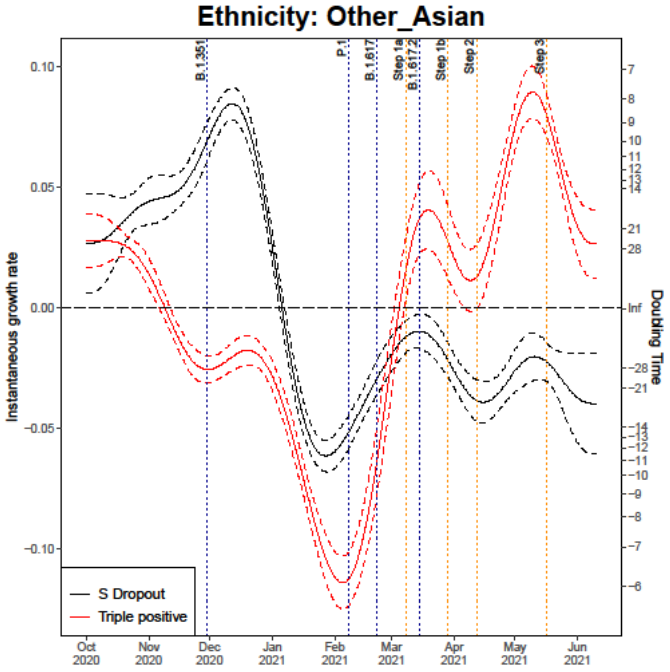


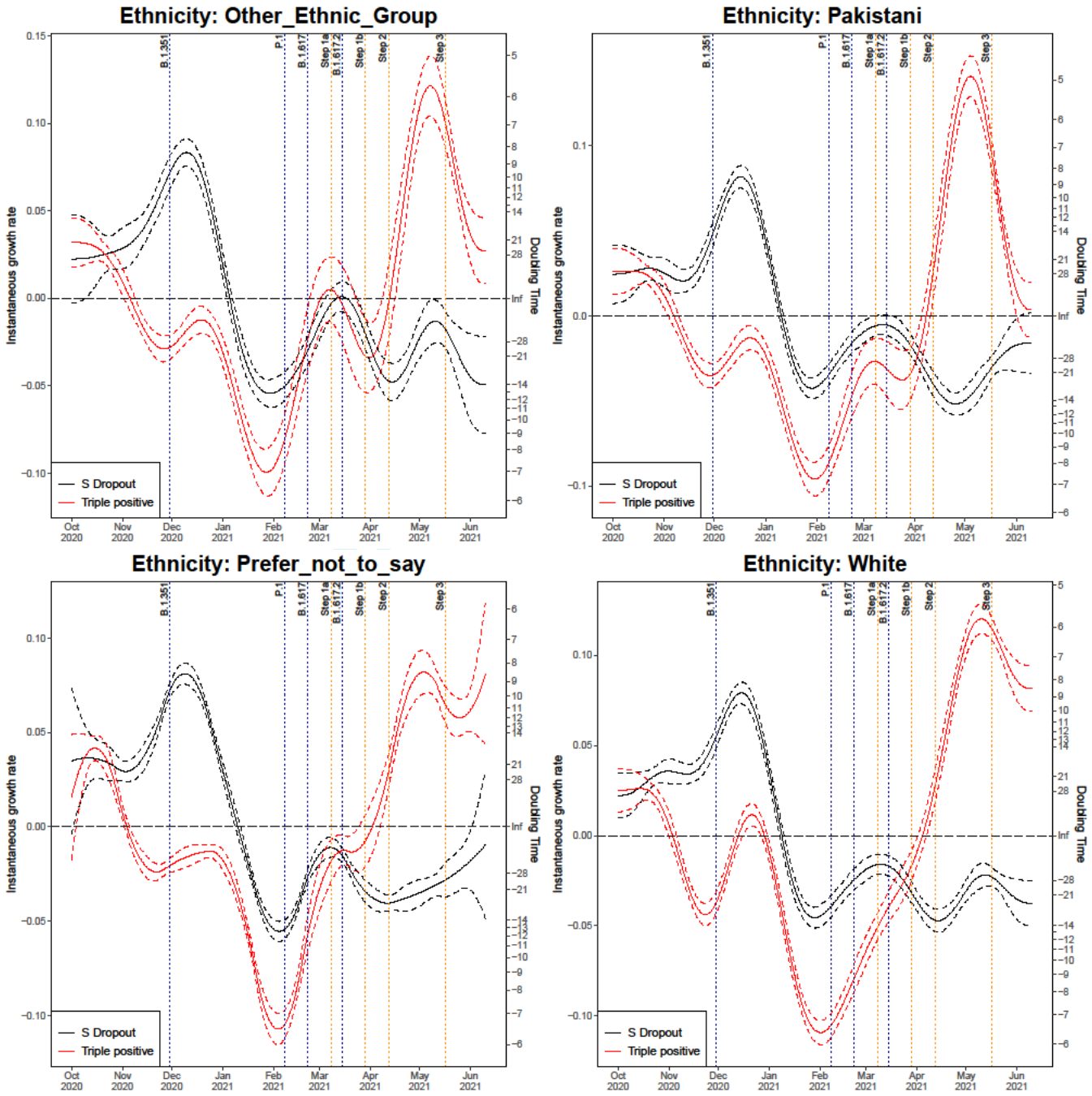
Appendix C: Cases Ethnicity



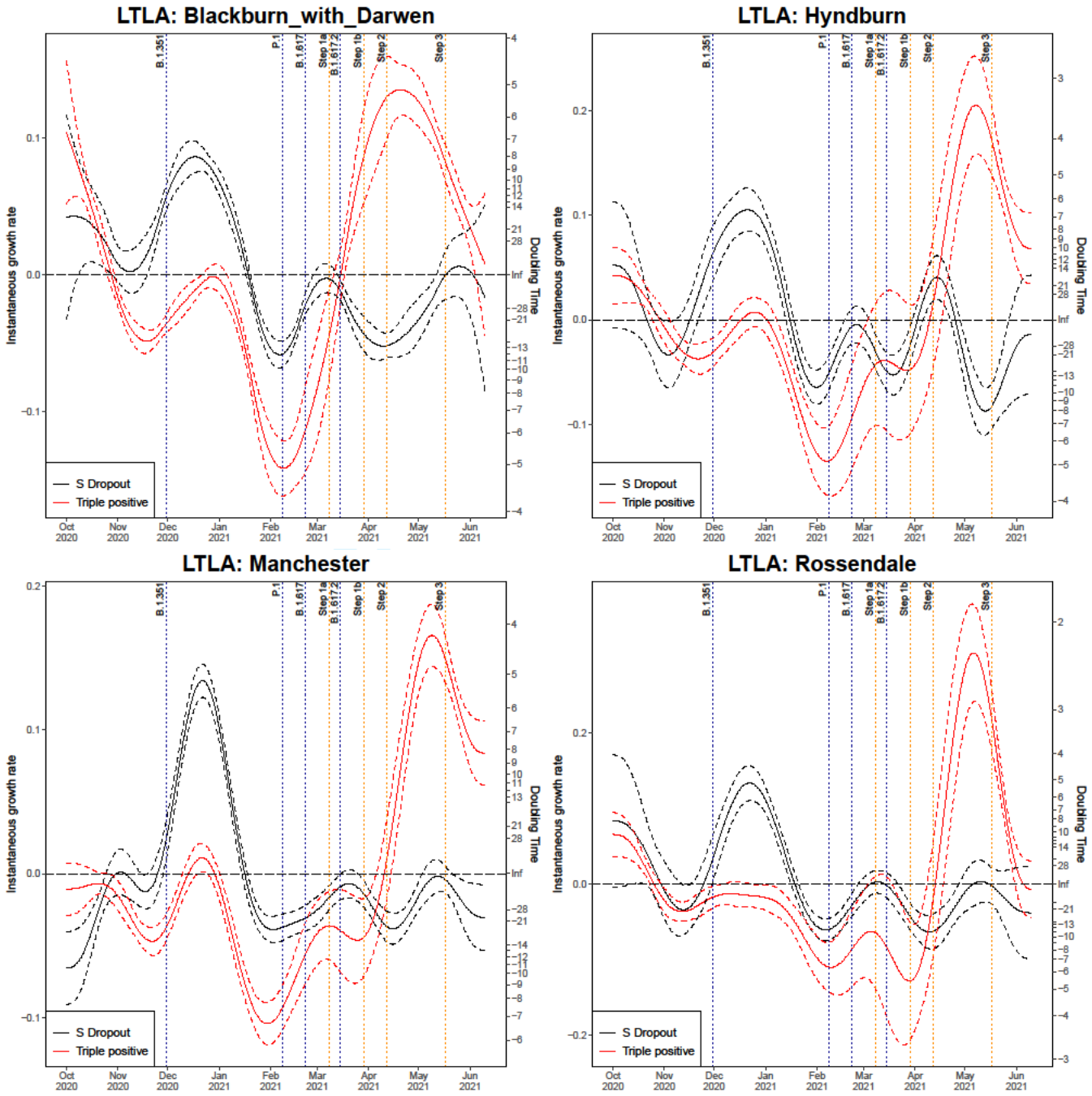


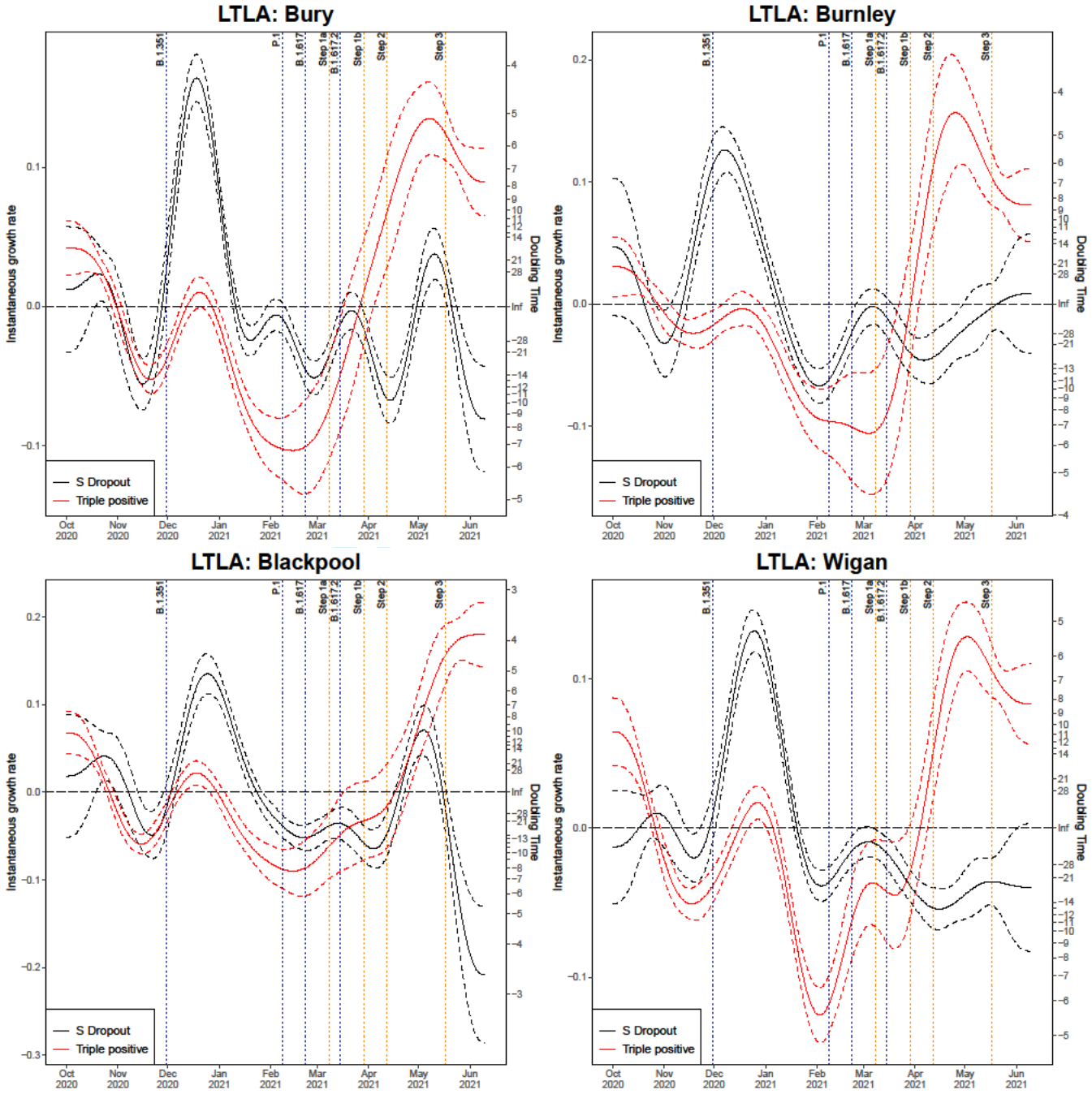


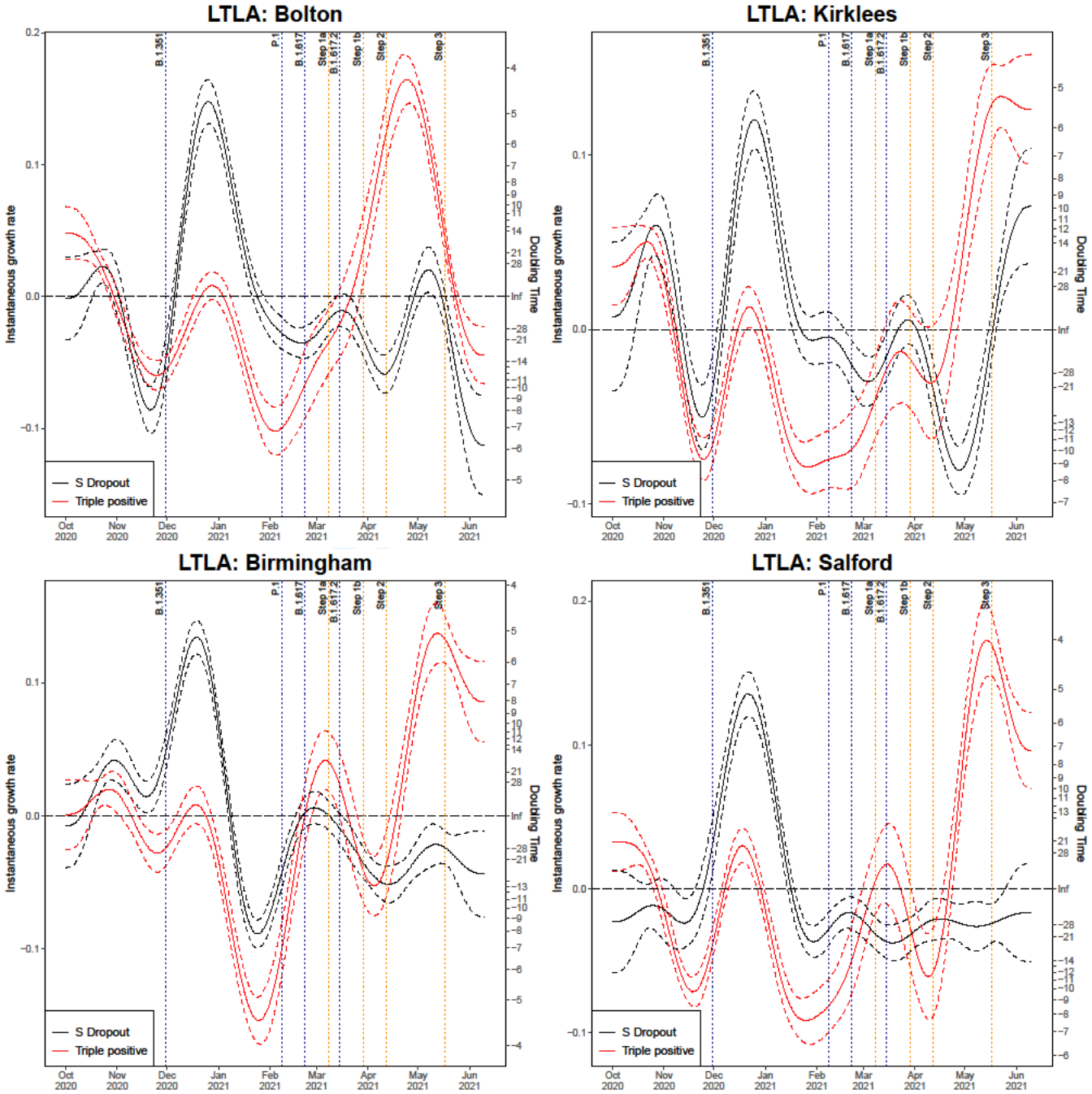


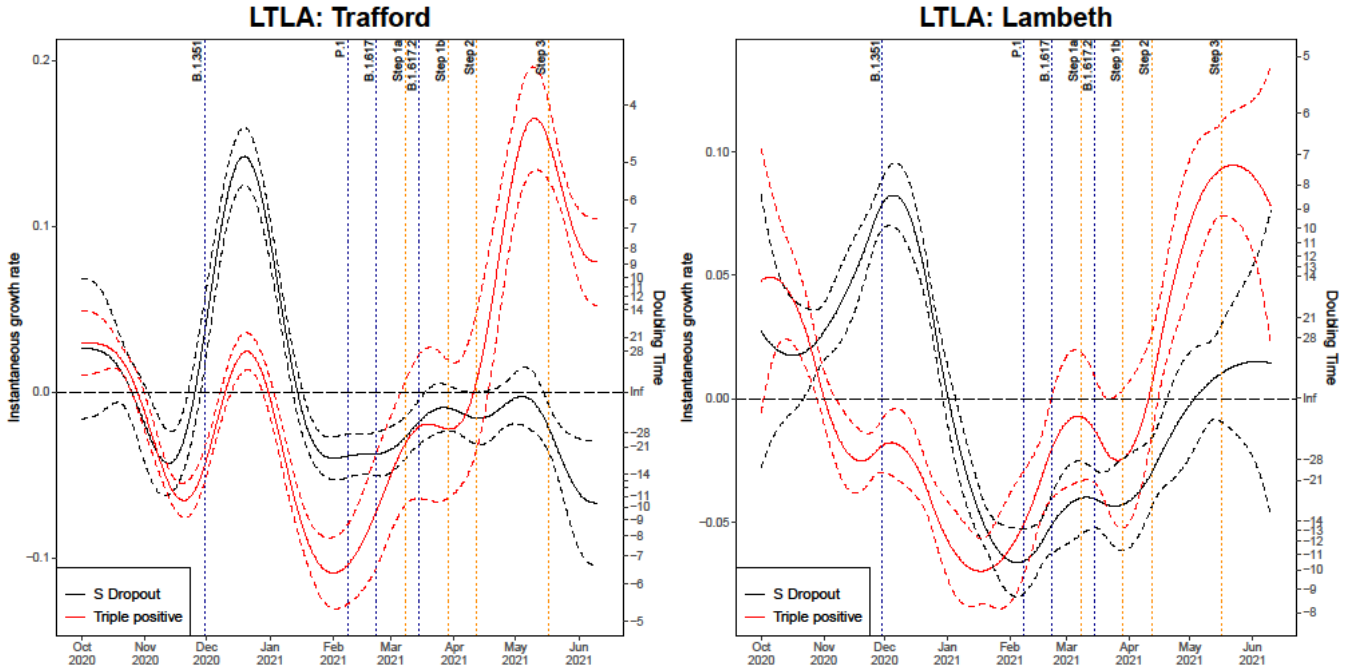


Appendix D: Cases LTLA

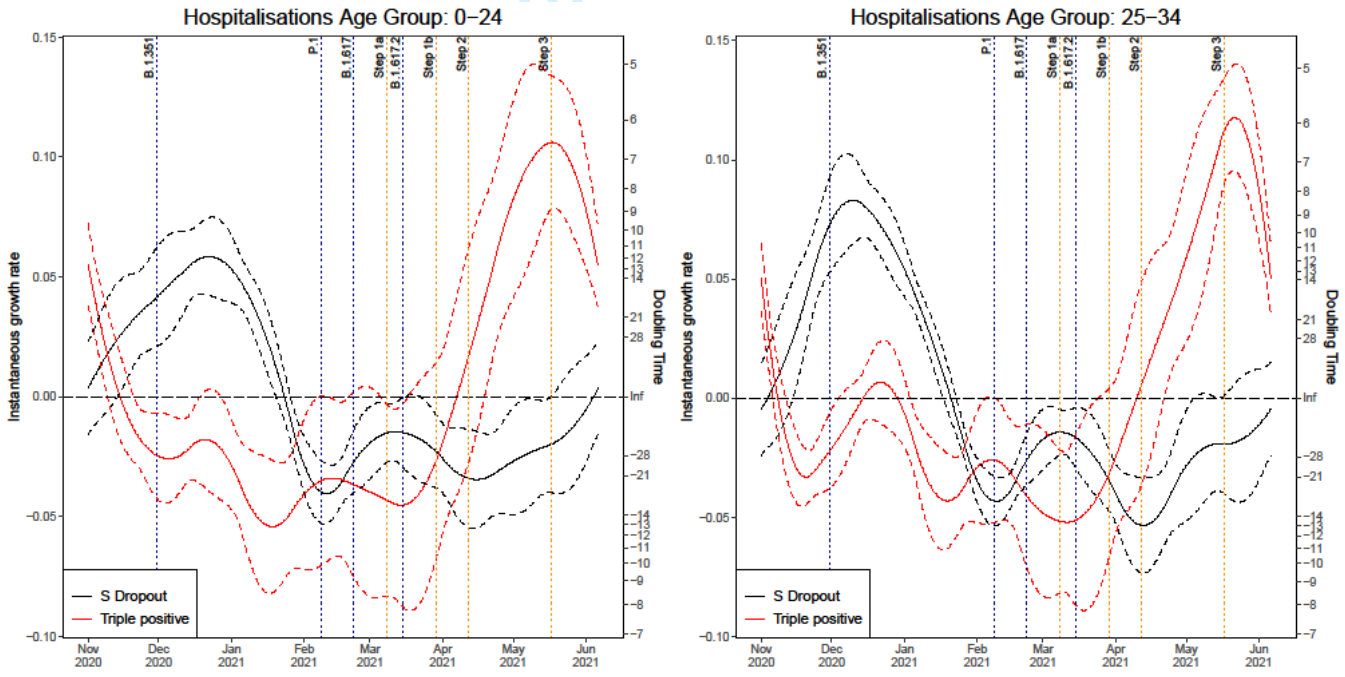


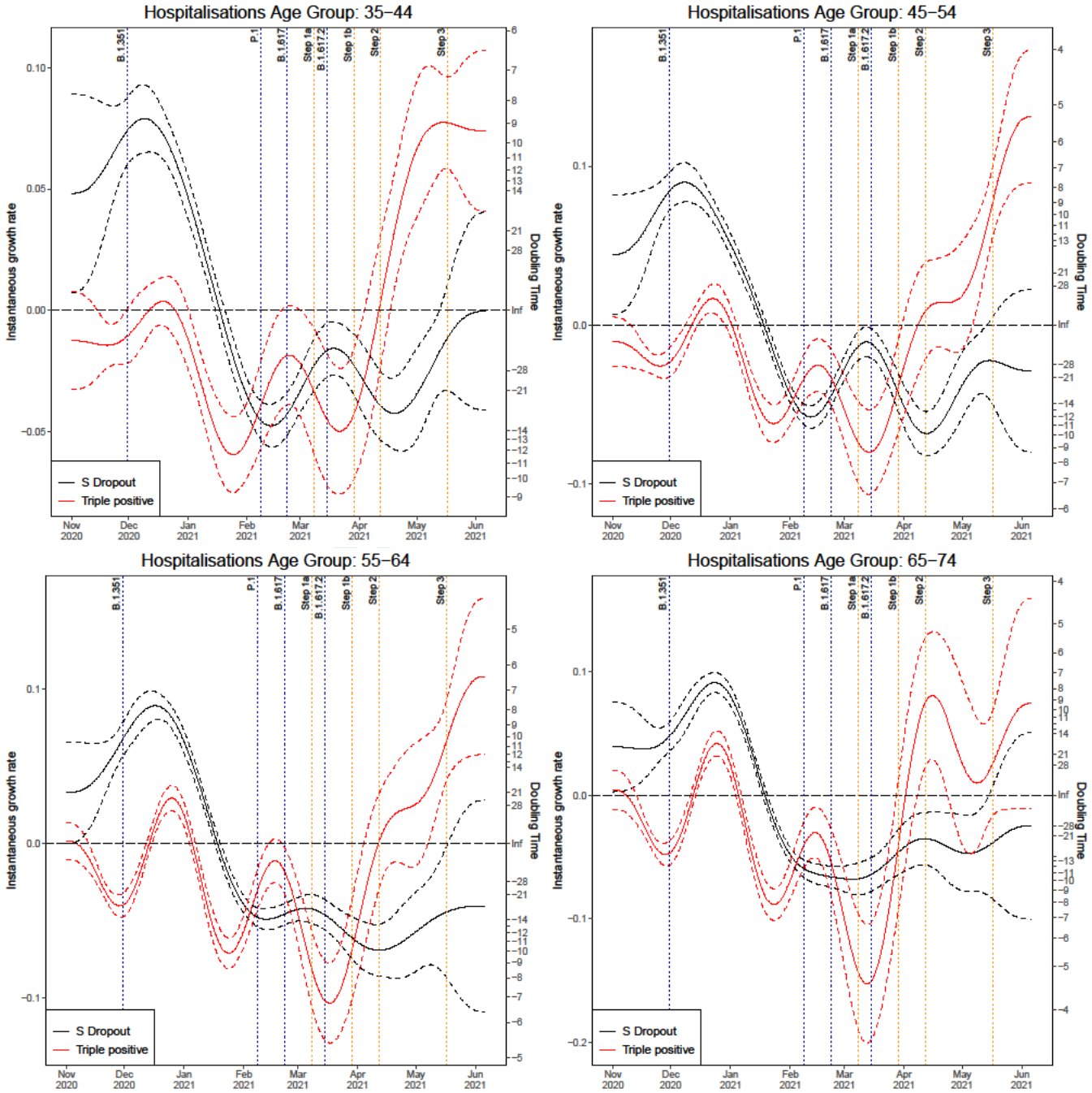


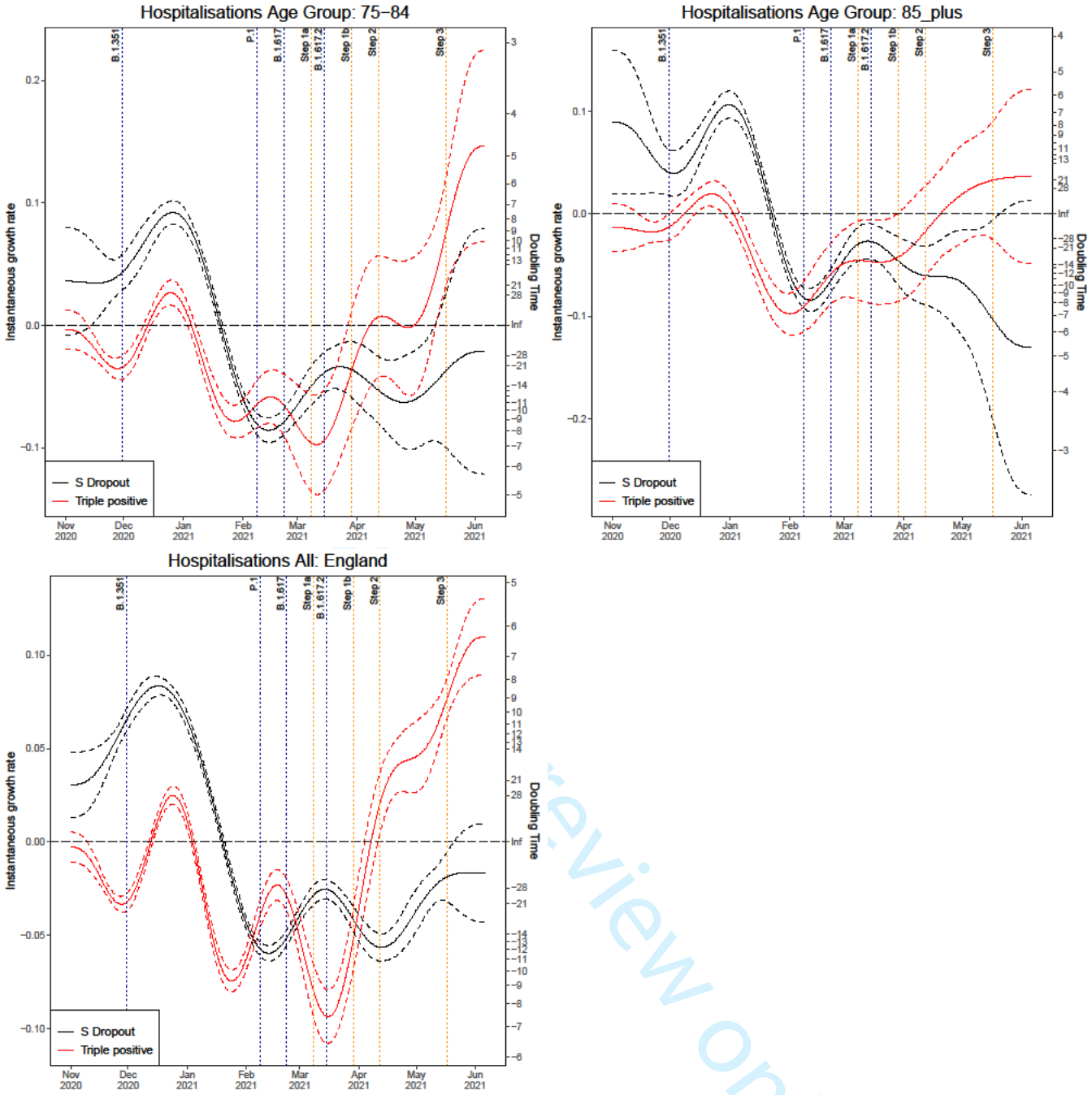




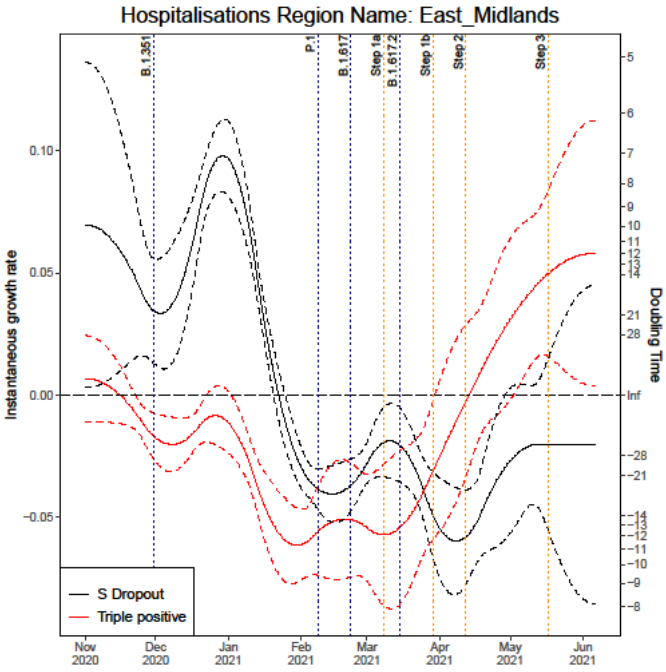
Appendix E: Hospitalisation Age Groups

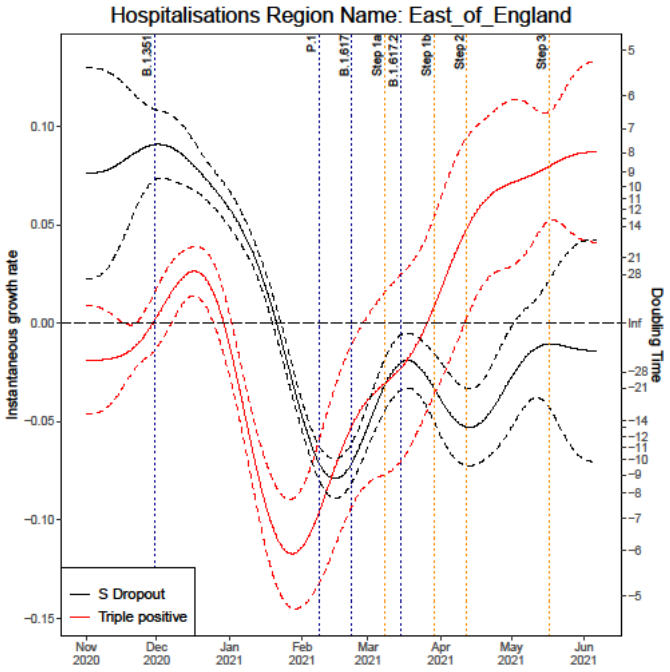


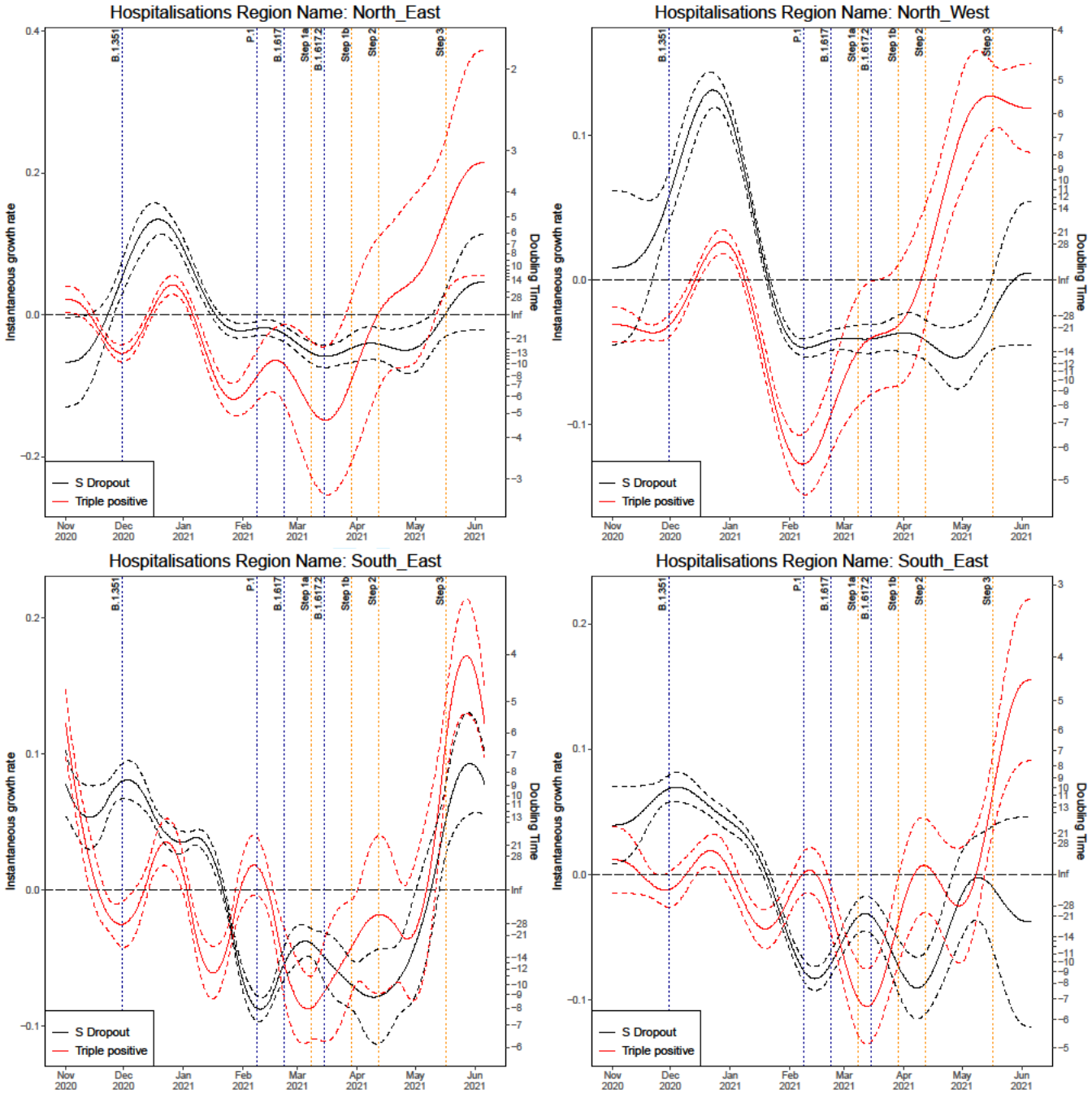


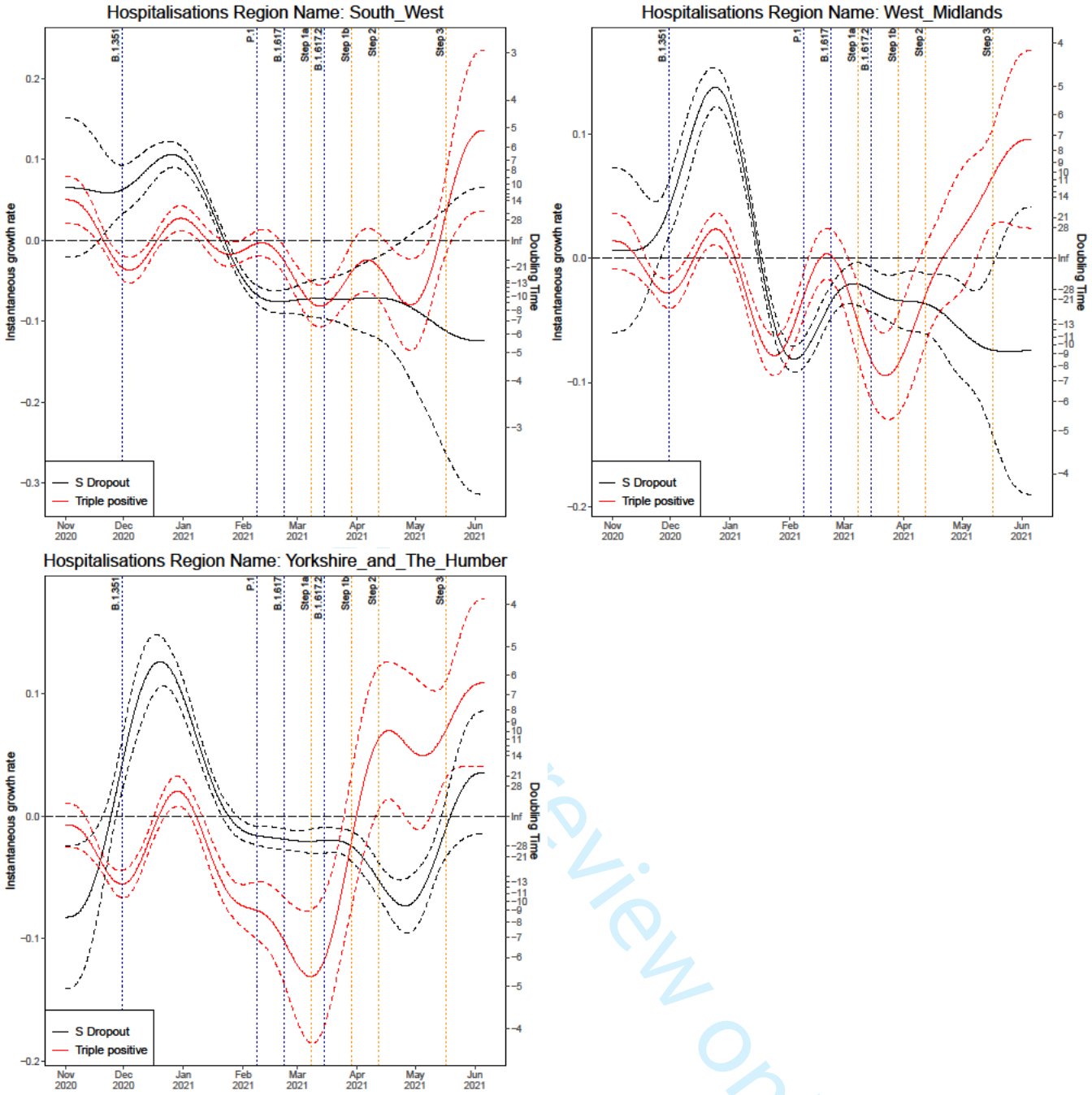


Appendix F: Hospitalisation Region

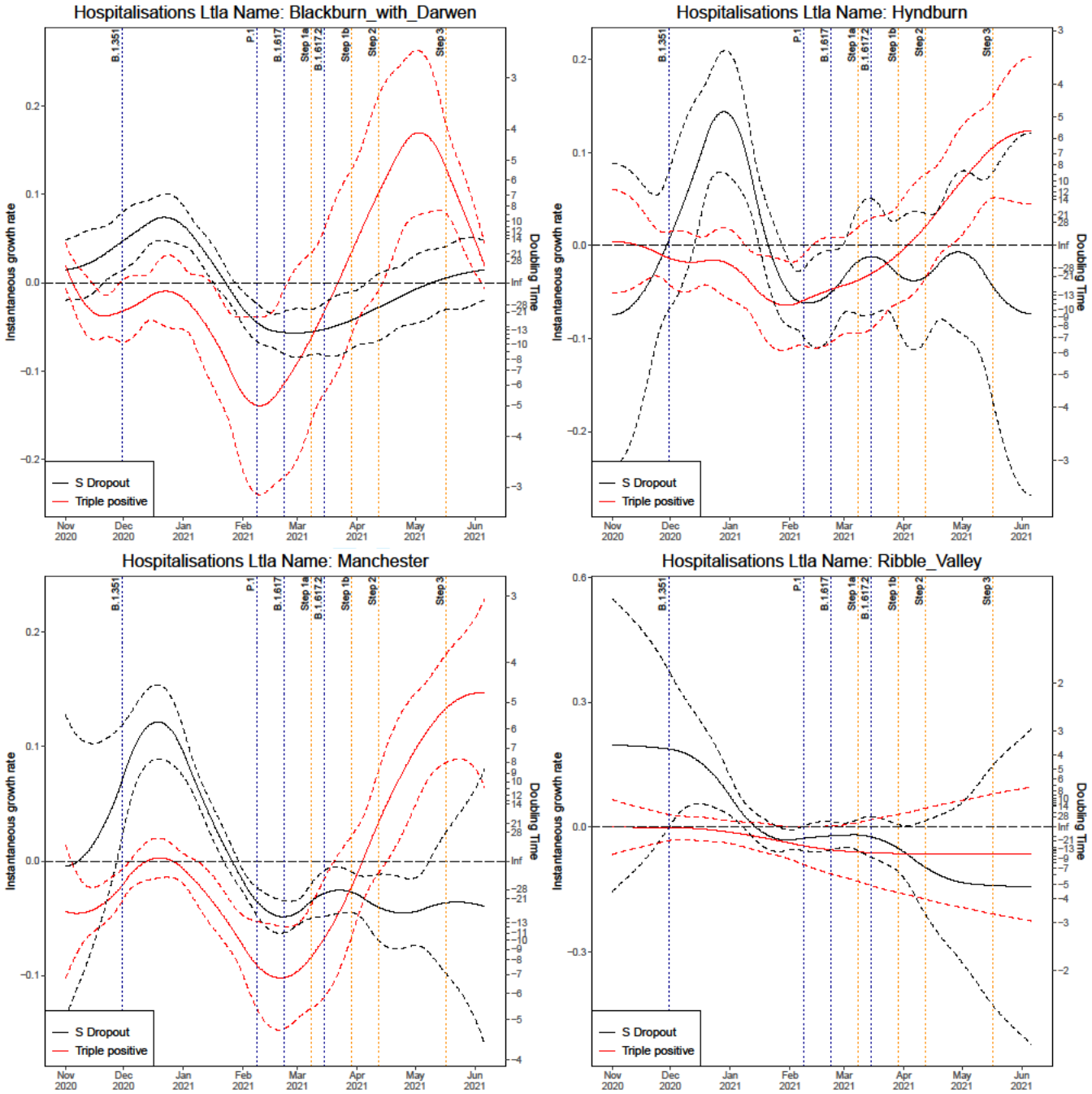


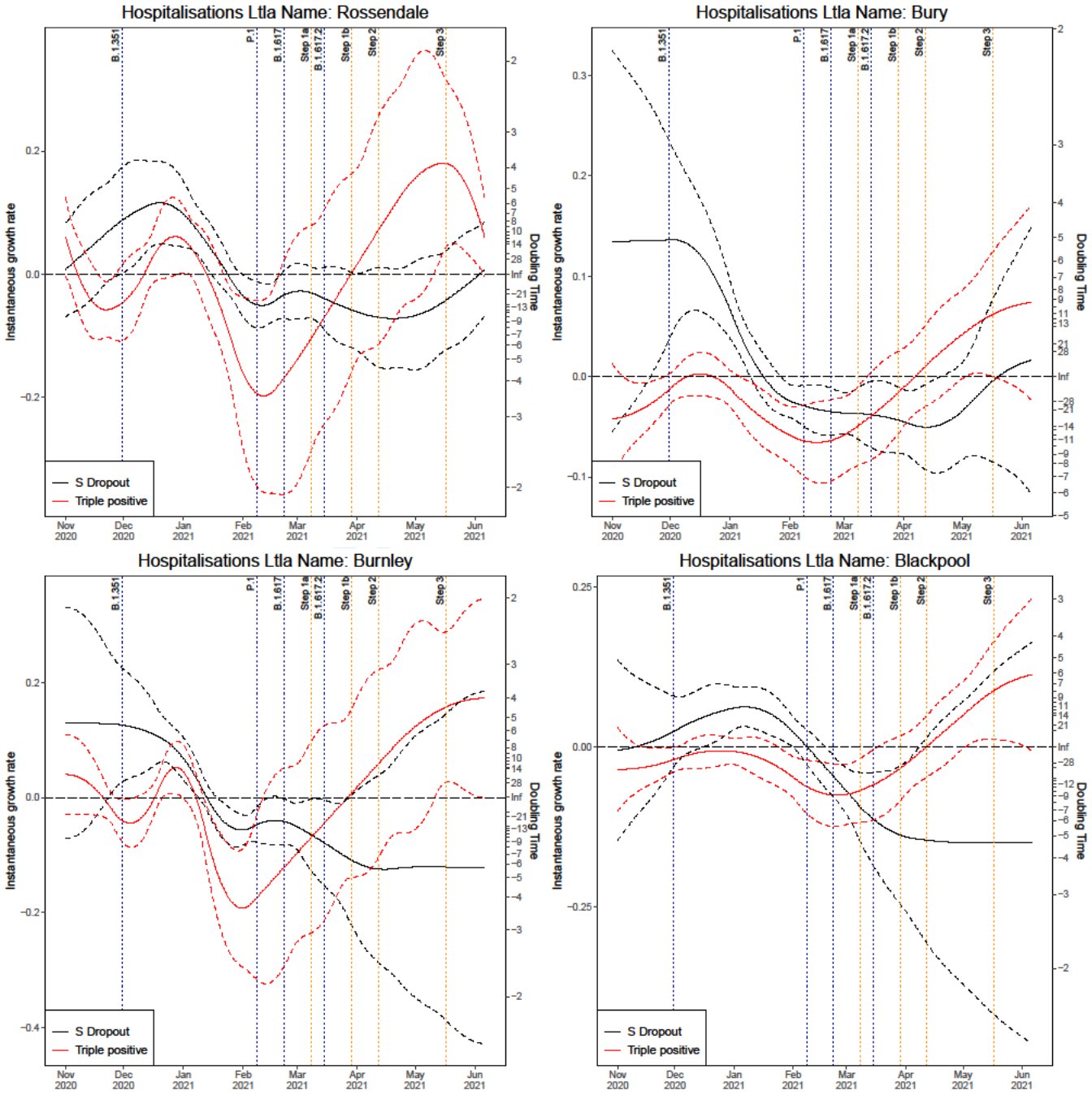


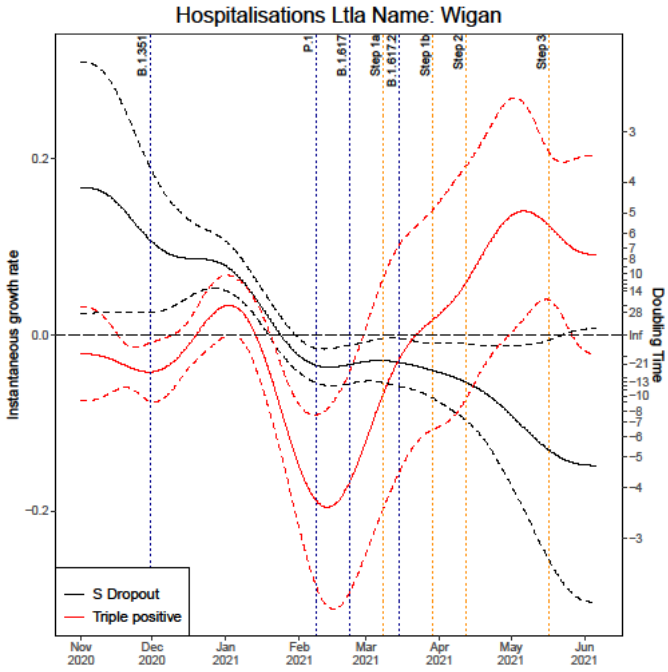


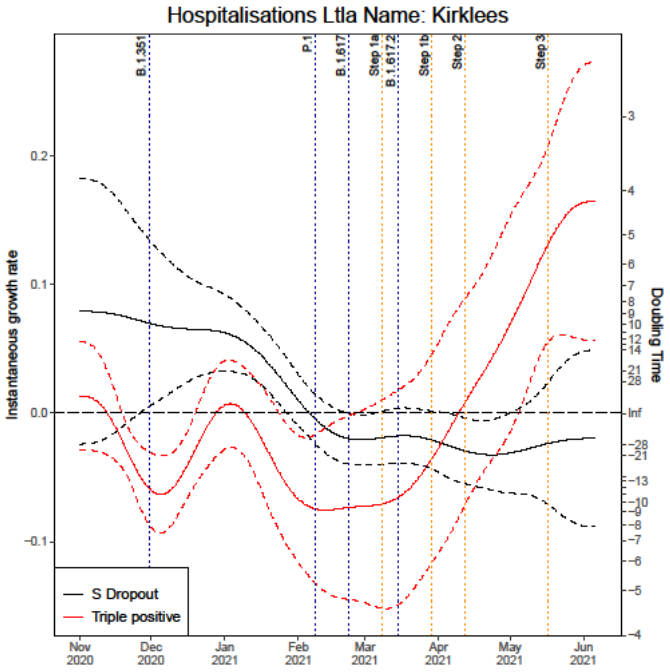


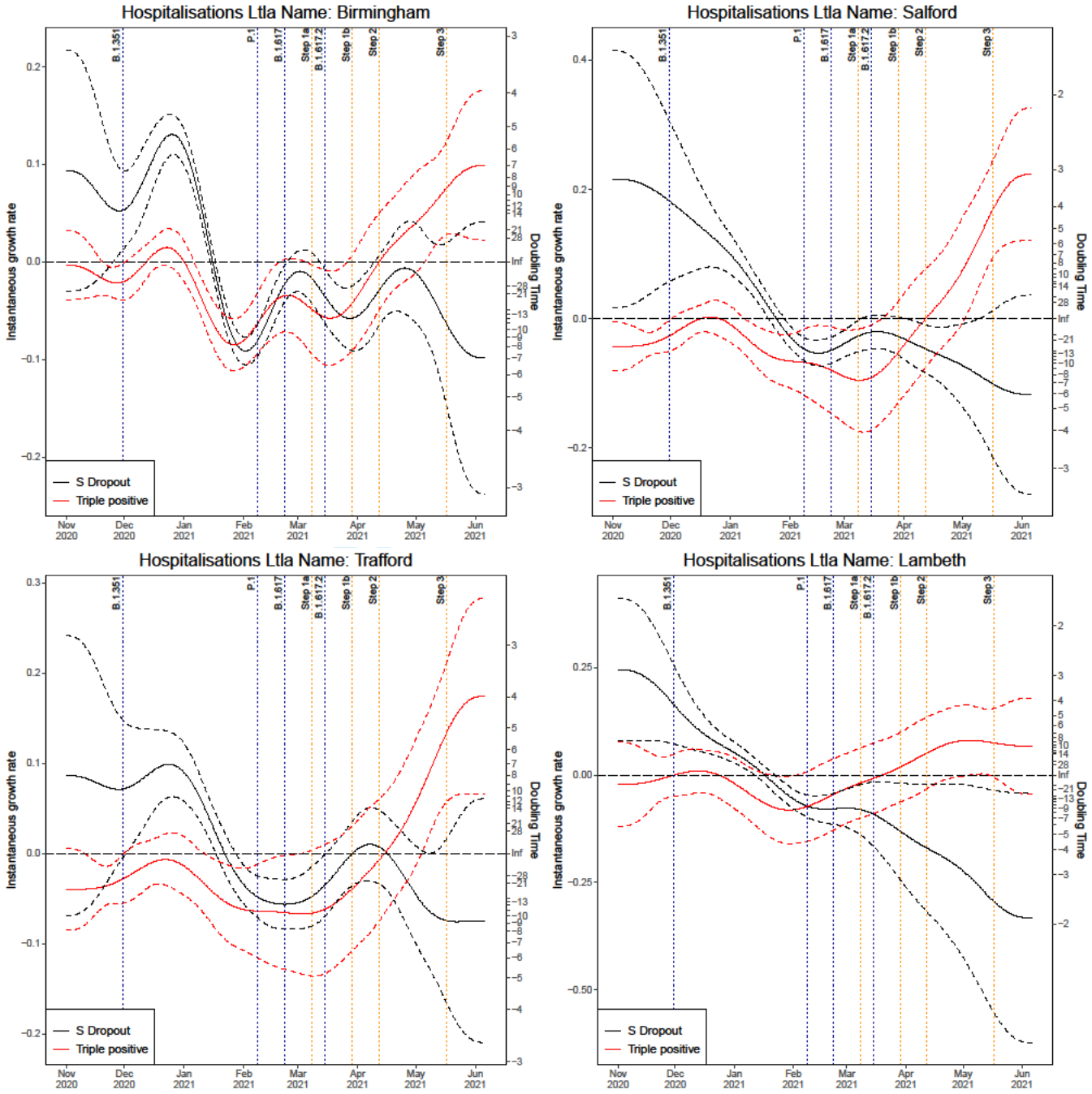
Appendix G: Hospitalisation LTLA











Appendix H

	SI mean	SI SD
	5.68	4.77
	2.97	3.29
	5.3	0.26
	7.2	
	4.55	3.3
	4.6	5.55
	5.1	5.3
	5.1	
	3.9	4.75
	6.5	
	4.9	4.4
	6.5	4.7
	6.3	4.2
	5.9	4.8
	3.9	4.24
	4.3	0.716
	7.5	3.4
	5.2	4.32
	4.1	0.882

	6.6	
	3.4	
	3.1	0.75
	5.5	3.9
	4.7	2.9
	5.8	
Mean	5.144	3.496
SD	1.22657	1.659
Min	2.97	0.26
Max	7.5	5.55

Appendix I

Time from Symptom Onset to Specimen Date

Start Date	End Date	N	Fit	Mean	SD	Alpha	Beta	Looic	Waic	Max Rhat	Min Rhat	Bad Pareto
01/01/2020	31/05/2020	557	weibull	3.66875	2.76983	1.34002	3.99321	2520.042	2519.159	1.00519	1.00089	0
01/06/2020	31/08/2020	4069	weibull	2.95717	2.39506	1.2423	3.17018	16859.34	16857.72	1.00025	0.999731	0
01/09/2020	30/09/2020	10000	weibull	2.6209	2.11027	1.24985	2.81378	39058.67	39055	1.00013	0.999542	0
01/10/2020	31/10/2020	10000	weibull	2.18193	2.05758	1.06113	2.2331	36226.77	36230.66	1.00124	0.99963	0
01/11/2020	30/11/2020	10000	weibull	1.97344	2.18076	0.906533	1.8826	34588.96	34602.01	1.00749	1.00303	0
01/12/2020	31/12/2020	10000	weibull	2.45101	2.17378	1.12993	2.56172	38238.76	38241.01	1.00077	1.00026	0
01/01/2021	31/01/2021	10000	weibull	2.07839	2.06587	1.00626	2.0836	35433.81	35439.34	1.00173	1.00054	0
01/02/2021	28/02/2021	10000	weibull	1.94196	2.00735	0.967796	1.91389	34250.36	34259.23	1.00184	0.999833	0
01/03/2021	31/03/2021	10000	weibull	1.96348	2.00263	0.98072	1.94674	34429.29	34437.08	1.00288	0.999865	0
01/04/2021	30/04/2021	10000	weibull	2.10586	2.06323	1.02091	2.12367	35648.06	35652.98	1.00075	0.999712	0
01/05/2021	31/05/2021	10000	weibull	2.01347	2.07044	0.972811	1.98905	34898.79	34906.32	1.00235	1.00046	0

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	3	State specific objectives, including any prespecified hypotheses	2
Methods			
Study design	4	Present key elements of study design early in the paper	2,3,4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2,3,4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	2,3,4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*2,3,4*
Bias	9	Describe any efforts to address potential sources of bias	2,3,4
Study size	10	Explain how the study size was arrived at	2,3,4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	2,3,4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	2,3,4
		(*b*) Describe any methods used to examine subgroups and interactions	2,3,4
		(*c*) Explain how missing data were addressed	
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	2,3,4

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	2,3,4
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	4-6
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	6-8
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	6-8
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	6-8
Generalisability	21	Discuss the generalisability (external validity) of the study results	6-8
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	8

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.