



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Appropriately smoothing prevalence data to inform estimates of growth rate and reproduction number

Oliver Eales^{a,b,*}, Kylie E.C. Ainslie^{a,b,c}, Caroline E. Walters^{a,b}, Haowei Wang^{a,b}, Christina Atchison^a, Deborah Ashby^a, Christl A. Donnelly^{a,b,d}, Graham Cooke^{e,f,g}, Wendy Barclay^e, Helen Ward^{a,f,g}, Ara Darzi^{f,g,h}, Paul Elliott^{a,f,g,i,j,k}, Steven Riley^{a,b,*}

^a School of Public Health, Imperial College London, London, United Kingdom

^b MRC Centre for Global Infectious Disease Analysis and Abdul Latif Jameel Institute for Disease and Emergency Analytics, Imperial College London, London, United Kingdom

^c Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

^d Department of Statistics, University of Oxford, Oxford, United Kingdom

^e Department of Infectious Disease, Imperial College London, London, United Kingdom

^f Imperial College Healthcare NHS Trust, Imperial College London, London, United Kingdom

^g National Institute for Health Research, Imperial Biomedical Research Centre, Imperial College London, London, United Kingdom

^h Institute of Global Health Innovation, Imperial College London, London, United Kingdom

ⁱ MRC Centre for Environment and Health, Imperial College London, London, United Kingdom

^j Health Data Research (HDR), Imperial College London, London, United Kingdom

^k UK Dementia Research Institute, Imperial College London, London, United Kingdom

ARTICLE INFO

Dataset link: <https://github.com/mrc-ide/react-1-study-materials>, <https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/real-time-assessment-of-community-transmission-findings/>, <https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/react-1-study-materials/>

Keywords:

SARS-CoV-2

COVID-19

Bayesian P-spline

Cross-sectional study

Reproduction number

ABSTRACT

The time-varying reproduction number (R_t) can change rapidly over the course of a pandemic due to changing restrictions, behaviours, and levels of population immunity. Many methods exist that allow the estimation of R_t from case data. However, these are not easily adapted to point prevalence data nor can they infer R_t across periods of missing data. We developed a Bayesian P-spline model suitable for fitting to a wide range of epidemic time-series, including point-prevalence data. We demonstrate the utility of the model by fitting to periodic daily SARS-CoV-2 swab-positivity data in England from the first 7 rounds (May 2020–December 2020) of the REal-time Assessment of Community Transmission-1 (REACT-1) study. Estimates of R_t over the period of two subsequent rounds (6–8 weeks) and single rounds (2–3 weeks) inferred using the Bayesian P-spline model were broadly consistent with estimates from a simple exponential model, with overlapping credible intervals. However, there were sometimes substantial differences in point estimates. The Bayesian P-spline model was further able to infer changes in R_t over shorter periods tracking a temporary increase above one during late-May 2020, a gradual increase in R_t over the summer of 2020 as restrictions were eased, and a reduction in R_t during England's second national lockdown followed by an increase as the Alpha variant surged. The model is robust against both under-fitting and over-fitting and is able to interpolate between periods of available data; it is a particularly versatile model when growth rate can change over small timescales, as in the current SARS-CoV-2 pandemic. This work highlights the importance of pairing robust methods with representative samples to track pandemics.

1. Introduction

Since the beginning of the COVID-19 pandemic, governments and policy makers have sought to strike a delicate balance between

controlling the spread of the SARS-CoV-2 virus and allowing society to function as close to pre-pandemic levels as possible. Non-pharmaceutical interventions (NPIs) have often been introduced to curtail the spread of the virus by reducing rates of transmission,

* Correspondence to: Imperial College London, St Mary's Hospital, Praed St, London W2 1NY.

E-mail addresses: o.eales18@imperial.ac.uk (O. Eales), k.ainslie@imperial.ac.uk (K.E.C. Ainslie), caroline.walters@healthlumen.com (C.E. Walters), haowei.wang18@imperial.ac.uk (H. Wang), christina.atchison11@imperial.ac.uk (C. Atchison), deborah.ashby@imperial.ac.uk (D. Ashby), c.donnelly@imperial.ac.uk (C.A. Donnelly), g.cooke@imperial.ac.uk (G. Cooke), w.barclay@imperial.ac.uk (W. Barclay), h.ward@imperial.ac.uk (H. Ward), a.darzi@imperial.ac.uk (A. Darzi), p.elliott@imperial.ac.uk (P. Elliott), s.riley@imperial.ac.uk (S. Riley).

<https://doi.org/10.1016/j.epidem.2022.100604>

Received 28 January 2022; Received in revised form 31 May 2022; Accepted 17 June 2022

Available online 22 June 2022

1755-4365/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

with varying levels of stringency from national lockdowns and curfews (Prime, 2021; Premier, 2022; van Algemene Zaken, 2021), to reduced opening hours for hospitality (Cabinet, 2020a). Despite their success in controlling the virus, NPIs have also been associated with negative impacts on the economy and on other facets of people's health, for example mental health (O'Connor et al., 2021). Accurate and timely measurements of infection prevalence and its rate of growth are required to effectively inform individuals and governments so that interventions can be timely and proportionate.

Throughout the pandemic there has been a large amount of data collected in order to track the SARS-CoV-2 virus spread and prevalence of COVID-19 disease, with numerous data streams in England alone (Elliott et al., 2021; Drew et al., 2020; Hillary et al., 2021; Pouwels et al., 2021; UK, 2022). However, sources of pandemic data are prone to biases which may lead to effective sample populations unrepresentative of the underlying population. For example, mass testing, a bedrock of pandemic response, can be heavily influenced by the test supply, the propensity of individuals to seek tests and whether tests are only used on symptomatic individuals (Ricoça Peixoto et al., 2020).

The REal-time Assessment of Community Transmission-1 (REACT-1) study (Riley et al., 2021) is a series of repeat cross-sectional study that aims to accurately measure the community prevalence of SARS-CoV-2 in England over time. There has been a round of the study approximately monthly since May 2020 (Elliott et al., 2021). Due to the random sampling procedure it employs, it is relatively unbiased compared to many other sources of data collection, allowing accurate estimates of prevalence to be obtained over each round.

Prior pandemics (Poletti et al., 2011) and standard epidemic theory suggest that prevalence should increase or decrease exponentially over substantial periods of time. For this reason, typical time-series analysis of epidemic data might include fitting an exponential model to the data under a constant growth rate assumption and from this inferring the reproduction number, R (Mills et al., 2004). This is valid at small timescales, but due to depletion of susceptibles, NPIs, and changing behaviour over time, the assumption of constant growth rate is invalid over a longer duration. Various methods exist that allow the course of an epidemic to be studied over a longer period. The progression of the instantaneous reproduction number R_t over time has been estimated from incidence data with only assumptions of the serial interval required (Cori et al., 2013; Wallinga and Teunis, 2004). These methods can have significant variation over small timescales (White and Pagano, 2008) and so smoothing over time windows is often performed. However, this can make the models sensitive to assumptions on the degree of smoothing (Gostic et al., 2020). Even those methods that are able to avoid these limitations (Parag, 2021) are limited in their ability to deal with times-series data where a significant number of daily data is missing.

Splines are a versatile tool for modelling non-linear functions when the exact form of the function is unknown. However, underlying assumptions can lead to either over- or under-fitting. Penalised splines (P-splines) (Eilers and Marx, 1996) seek to avoid over-fitting through the inclusion of discrete penalties on the basis coefficients, though this penalty has no exact interpretation in terms of the function's shape. Their Bayesian counterpart however (Lang and Brezger, 2004) offers a statistically robust method of capturing variation in the data whilst also preventing over-fitting through the inclusion of appropriate prior distributions that act on the functional form of the spline. Bayesian P-splines have been used to model the spatial variation in the risk of cholera infection (Osei et al., 2012), the effects of age, space and time on cancer mortality (Goicoa et al., 2019), and to model the relation between respiratory mortality and daily fine particle exposure (Fang et al., 2019). Their use in epidemic time-series data however, has been limited (van de Kasstele et al., 2019), perhaps because major programming packages only allow for prior distributions that penalise changes in the main response function which would correspond to a

prior distribution of constant prevalence which is problematic given expected exponential trends.

Here we develop a Bayesian P-spline model that may be widely applicable to epidemic time-series data. The model, in the absence of data, penalises changes in the growth rate, reflecting the prior knowledge we have on an epidemic system. We demonstrate the utility of the model by fitting to the daily swab-positivity data from the first 7 rounds of the REACT-1 study. The data is semi-continuous over time with well sampled periods of between 18 and 32 days for each round, with non-sampled periods between rounds.

2. Materials and methods

2.1. REACT-1 data

The complete study protocol of REACT-1 has been described previously (Riley et al., 2020). In brief, letters are sent randomly to named individuals from the list of GP patients in England held by the National Health Service, stratified by lower-tier local authorities (LTLAs, $n = 315$). Those who agree to participate in the study self-administer throat and nose swabs (administered by parent or guardian for children aged 5–12 years). These are then collected and sent on a cold chain for analysis by reverse-transcript polymerase-chain-reaction (RT-PCR), with E- and N-gene targets, in a single commercial laboratory (round 1 also included some swabs tested in Public Health England laboratories). The REACT-1 study received research ethics approval from the South Central-Berkshire Research Ethics Committee (IRAS ID: 283787). The sample size of the study has ranged from 120,000 to 160,000 across the 7 rounds of data collection from May 2020 to December 2020 (Supplementary Table 1).

The cycle threshold (Ct) value of the E- and N-gene targets in the RT-PCR test is used as a proxy for intensity of the viral load (where Ct value is inversely proportional to viral load). A swab was defined as positive if (1) both the E- and N-gene targets were positive or (2) the N-gene target was positive with a Ct value less than 37. To test the sensitivity of our analyses we use different definitions of swab positivity: asymptomatic positives are defined as those who match our original positivity criteria but did not self-report having had any symptoms in the month prior to their swab test; double positives were defined as those with positive E- and N-gene targets; and lower threshold positives were defined in a similar way to our original positivity definition, using a Ct cut off value of 35 for the swabs with only N-gene target positive.

Participants provided information on the day they completed the swab test and the day it was collected is also recorded. However, one or both dates are missing for some participants. In participants where both the date of swab test and date of collection are known, we observe that the two dates are highly consistent (86.9% are the same date, 9.6% are one day apart). Thus, we define a new composite variable that is equal to the date an individual completed the swab test, if available and within the range of dates for which they could plausibly have received the test (70.1% of individuals), or, if the swab completion date is not available or is inconsistent, then we use the date of collection as a proxy (29.1% of individuals). For all rounds, over 90% of swabs had either a date of completion or a date of collection and at least 95% of positive swab tests had one of these dates associated (Supplementary Table 1). Swab tests without a date variable could not be included in the temporal analysis and so were excluded. Individual swab-test results were weighted using rim weighting (Sharot, 1986) by: sex, deciles of the indices of multiple deprivation (IMD), LTLA counts and ethnic group.

2.2. Simple exponential model

As a baseline model to study how prevalence varies over short periods of time, we fit a simple exponential model with a constant

growth rate:

$$\pi(t) = A e^{r(t-t_0)} \quad (1)$$

where $\pi(t)$ is the prevalence at time t , A is the prevalence at time t_0 (the first date in the data the model is fit to) and r is the growth rate. We fit the exponential model to multiple subsets of the data. First, to explore the rate of growth across the period covered by two rounds, we fit the model to the data from each pair of subsequent rounds. Secondly, we explore within round growth rates by fitting the model to the 7 individual rounds of data. The model is implemented in STAN and parameter posteriors are sampled using a No-U-Turn Sampler (Hoffman and Gelman, 2011) assuming a weighted Binomial likelihood, and assuming uninformative constant priors for A and r .

The average reproduction number for the period described by each model was estimated assuming a generation time following a gamma distribution with shape parameter $\alpha = 2.29$ and rate parameter $\beta = 0.36$ (Bi et al., 2020). The reproduction number was then calculated from the estimate of the growth rate (Wallinga and Lipsitch, 2007), r , using the equation:

$$R = \left(1 + \frac{r}{\beta}\right)^\alpha \quad (2)$$

which is valid for $r > -\beta$.

2.3. Bayesian P-spline model

2.3.1. The model

A general model for prevalence using smoothed functions can be written as

$$g(\pi(t)) = s(t) \quad (3)$$

where $\pi(t)$ is, as before, the prevalence at time t , $s(t)$ is the value of a general smooth function at time t , and g is the link function, which we take as the logit function for the REACT-1 binomial data. We define our smooth function as a linear combination of N B-splines (see below) of order n :

$$s(t) = \sum_{i=1}^N b_i B_{i,n}(t). \quad (4)$$

The B-splines are defined by a non-decreasing sequence of knots $t_1 \dots t_q$ and the polynomial degree, p ($n - 1$) (de Boor, 1978). The first order ($p = 0$) B-splines are defined as

$$B_{i,1}(t) = \begin{cases} 1 & \text{if } t_i < t < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

and then higher order B-splines are defined recursively

$$B_{i,n}(t) = w_{i,n} B_{i,n-1}(t) + (1 - w_{i+1,n}) B_{i+1,n-1} \quad (6)$$

where,

$$w_{i,n} = \begin{cases} \frac{t-t_i}{t_{i+n-1}-t_i} & \text{if } t_i \neq t_{i+n-1} \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

To model how prevalence varies over time we define a family of 4th order B-spline functions over the duration of the study. The locations of knots are placed at regular intervals over the entire duration of the study so that there are approximately 5 days between each pair of adjacent knots (approximately as the length of data is not always divisible by 5). Additionally, the model is extended a further 3 knots from both the beginning and end of the study so that the model output for the period of interest is not affected by the irregular B-splines defined at the boundaries. These extensions of the model outside of the duration of the study are not included in any visualisation after fitting. Models using different knot sizes were investigated (Supplementary Figure 3), and a knot size of 5 days was chosen for the model used

in the main analyses. The knot size of 5 days was chosen to ensure that the model was not under-fitted, whilst also not allowing the fitting of the model to become too computationally expensive.

In order to limit over-fitting of the model to noise in the data, we define a second-order random-walk prior distribution (Lang and Brezger, 2004) on the coefficients:

$$b_i = 2b_{i-1} - b_{i-2} + u_i, \quad (8)$$

where

$$u_i \sim \mathcal{N}(0, \rho^2). \quad (9)$$

This prior distribution penalises any change from the first derivative of the smoothing function, which, for a logit link function, is approximately equal to the growth rate at low prevalence, and for a log link function is exactly equal to the growth rate. The amount to which changes of the first derivative of the smooth function are penalised is controlled by another parameter of the model, ρ , that we give a loose but proper inverse gamma prior distribution $\rho \sim IG(0.0001, 0.0001)$ (Supplementary Figure 4). The first two coefficients of the model, b_1 and b_2 , are given an uninformative constant prior distribution. A first-order random-walk prior distribution was also considered: $b_i = b_{i-1} + u_i$, where u_i is as before. However, a comparison of both models fit to simulated data of an exponentially increasing prevalence (Fig. 1) showed that a first-order random-walk prior distribution, in the absence of data, favours a constant prevalence, which is unnatural for an epidemic system.

The model is implemented in STAN and parameter posteriors are sampled using a No-U-Turn Sampler (Hoffman and Gelman, 2011) assuming a weighted Binomial likelihood. Four chains are run and the performance of the model fitting is checked by looking at measures of the bulk-ESS (checking it is greater than 100), tail-ESS (checking it is greater than 100) and individual parameters' potential scale reduction statistic, \hat{R} (checking they are less than 1.05) (Vehtari et al., 2019). An analogous Bayesian P-spline model is also fitted to Public Health England Pillar 1 and 2 case (by specimen date) data (UK, 2022). Instead of prevalence, the model fits to the daily number of cases using a log link function. The model is fitted assuming a Negative-Binomial likelihood, with the additional over-dispersion parameter that is required given an uninformative constant prior distribution.

When fitting the model to different subsets of the data, despite the target knot size being 5 days the actual knot size was not always exactly 5 days. In order to compare the estimated value of ρ between models using slightly different knot sizes we defined the standardised ρ , $\hat{\rho}$. The value of ρ should increase linearly with knot size for small knot sizes (for large knot sizes linearity will break down due to under-fitting). Therefore, we defined $\hat{\rho}$ as the value that ρ would have for a knot size of exactly 5 days assuming linearity.

2.3.2. Calculating average growth rate over a finite period

In order to compare the fit of the Bayesian P-spline model to the more simple exponential model it is necessary to calculate the average growth rate over finite periods of time of the model output. These can then be compared to the average growth rates calculated for the same periods of time by the exponential model. For any set of sampled parameters from the posterior we can calculate the prevalence at any time during the study period. We then assume that the prevalence at a time t_2 is linked to the prevalence at time t_1 by a constant growth rate \bar{r} such that

$$\pi(t_2) = \pi(t_1) e^{\bar{r}(t_2-t_1)}. \quad (10)$$

Then we can calculate the average growth rate between two times for any set of parameters sampled from the posterior through the equation:

$$\bar{r} = \frac{\log\left(\frac{\pi(t_2)}{\pi(t_1)}\right)}{t_2 - t_1}. \quad (11)$$

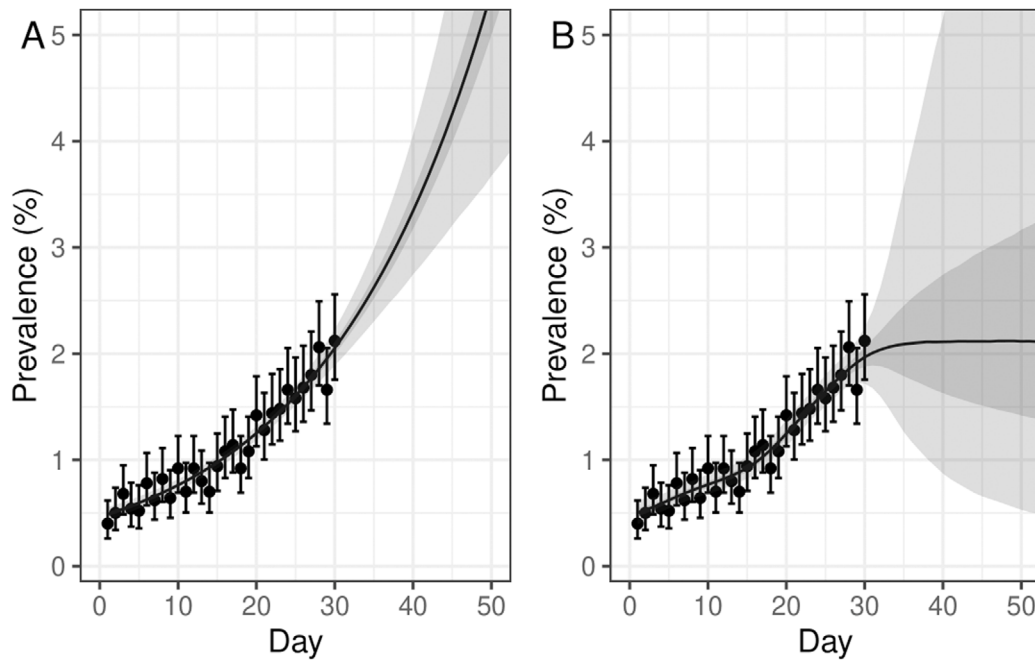


Fig. 1. Comparison of Bayesian P-spline models with second-order and first order random-walk prior distributions. Simulated Binomial data for an exponentially increasing prevalence, $\pi(t) = 0.5 \times \exp(0.05 \times t)$, with 5000 tests performed a day. Daily estimates of prevalence (points) and 95% confidence intervals are shown. (A) Model fit for a Bayesian P-spline model with a second-order random-walk prior distribution. (B) Model fit for a Bayesian P-spline model with a first order random-walk prior distribution. (A, B) Central estimates (solid line) are shown with 50% (dark shaded region) and 95% (light shaded region) credible intervals are also shown. Models have been defined up to 60 days, despite the data only going up to 30 days, in order to demonstrate the effect of the model's prior distribution during periods of no data.

By calculating this value over the entire posterior of parameter values we can calculate the posterior probability for the average growth rate between two time points of our model. This average growth rate can then be converted into an estimate for the average reproduction number over the period using Eq. (2). These estimates are analogous to the previous estimates found for the exponential model and so can readily be compared.

2.3.3. Calculating instantaneous growth rate

The rate of change of prevalence can be written in terms of a time varying growth rate, $r(t)$

$$\frac{d\pi(t)}{dt} = r(t)\pi(t). \tag{12}$$

This then has the solution,

$$\pi(t) = e^{\int r(t)dt}. \tag{13}$$

Equating the two equations (Eqs. (3) and (13)) for prevalence, $\pi(t)$ we have

$$\int r(t)dt = \log(g^{-1}(s(t))). \tag{14}$$

This can be solved for the instantaneous growth rate, $r(t)$, through differentiation with respect to time and Leibniz's rule giving:

$$r(t) = \frac{\frac{ds(t)}{dt}}{e^{s(t)+1}} \tag{15}$$

for the case in which g is the logit function. The smooth function, $s(t)$ and its first derivative with respect to time are defined over the entire period of the study. We calculate $r(t)$ over the duration of the study for all combinations of parameters in our sampled posterior. This gives us the posterior distribution of $r(t)$ for the entire study period. For the case in which g is the log function $r(t)$ is simply the first derivative with respect to time of $s(t)$.

2.3.4. Calculating reproduction number from prevalence estimates

If the rate of secondary infections at time τ since infection is given by $\eta(\tau)$ then we can write the reproduction number as

$$R = \int_0^\infty \eta(\tau)d\tau. \tag{16}$$

Normalising $\eta(\tau)$ gives us the generation time $g(\tau)$

$$g(\tau) = \frac{\eta(\tau)}{\int_0^\infty \eta(\tau)d\tau} = \frac{\eta(\tau)}{R}. \tag{17}$$

Writing the incidence at time t , $I(t)$, in terms of the incidence at times less than t we have,

$$I(t) = \int_0^\infty I(t-\tau)\eta(\tau) d\tau \tag{18}$$

$$= \int_0^\infty I(t-\tau)g(\tau)R(t-\tau)d\tau. \tag{19}$$

The prevalence at time t , $\pi(t)$ can similarly be linked to incidence by a function, $f(\tau)$, that describes the probability of someone infected τ days ago testing swab-positive:

$$\pi(t) = \int_0^\infty I(t-\tau)f(\tau) d\tau \tag{20}$$

Or substituting Eq. (19) into Eq. (20) we get

$$\pi(t) = \int_0^\infty \int_0^\infty I(t-\tau-\eta) f(\tau) R(t-\tau-\eta) g(\eta) d\tau d\eta \tag{21}$$

If we assume that R is approximately constant over the time frame in which $f(\tau)$ goes to 0 then we get:

$$\pi(t) = \int_0^\infty \pi(t-\tau)g(\tau)R(t-\tau)d\tau. \tag{22}$$

If we further make the assumption that R has been constant over the duration of time in which $g(\tau)$ goes to zero then we can rearrange the expression (Wallinga and Lipsitch, 2007) to get

$$R_t = \frac{\pi(t)}{\int_0^\infty \pi(t-\tau)g(\tau) d\tau}. \tag{23}$$

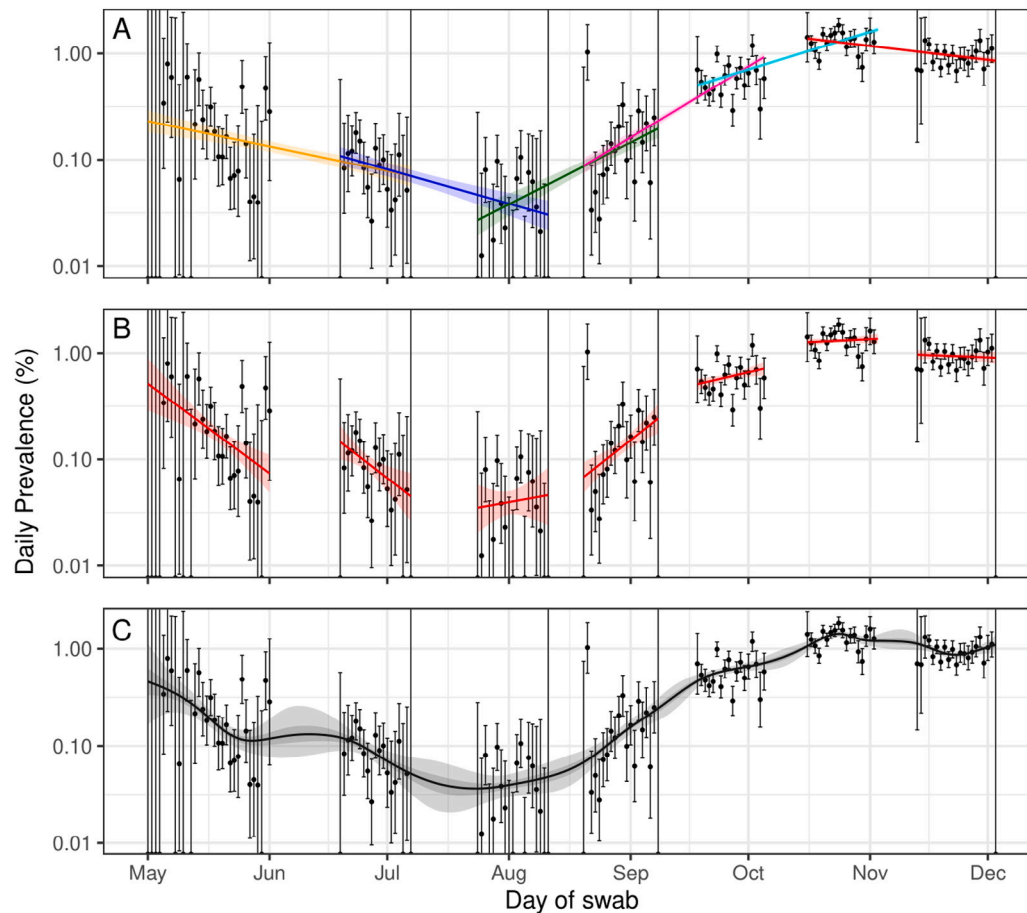


Fig. 2. Exponential and Bayesian P-spline model fits. A: Exponential models fit to rounds 1 and 2 (yellow), rounds 2 and 3 (blue), rounds 3 and 4 (green), rounds 4 and 5 (pink), rounds 5 and 6 (cyan), and rounds 6 and 7 (red) of REACT-1 with shaded regions showing the 95% credible intervals. B: Exponential models fit to individual rounds (red) of REACT-1 with shaded regions showing the 95% credible intervals. C: Fit of the Bayesian P-spline model to all 7 rounds of REACT-1. Shaded regions show central 50% (dark grey) and 95% (light grey) credible intervals. Daily prevalence estimates (points) are shown with 95% credible intervals (error bars). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This equation relies on the assumption that R has remained constant over a significant duration of time and so is valid for periods in which R is constant. During periods in which R is not constant the estimates will instead only be approximate reflecting the average R over the fixed period.

The fit of the Bayesian P-spline model gives us the posterior distribution of $\pi(t)$ for the entire duration of the study. We again take the generation time $g(\tau)$ to be an inverse gamma distribution with scale parameter 2.29, and rate parameter 0.36 (Bi et al., 2020). The rolling 14 day average of R_t can then be calculated from the above equation by integrating over the previous 14 days at each point in time. Note that 14 days was chosen so that the generation time distribution had declined to a negligible amount. By calculating R_t for the entire posterior of our prevalence estimates we can obtain an appropriate credible interval for our estimates of R_t .

3. Results

3.1. Comparison of exponential model and Bayesian P-spline model

The growth rates of prevalence between rounds were assessed by fitting the simple exponential model to each pair of subsequent rounds (Fig. 2a). The Bayesian P-spline model was fitted to all 7 rounds (Fig. 2c) and, from this continuous estimate of prevalence, estimates of the average growth rates over the same subsequent rounds were also obtained. The Bayesian P-spline model was also fit to an increasing number of rounds (from round 1 only to all 7 rounds) and the average

growth rate for the final two rounds of each fit was estimated. Estimates of R and doubling/halving times were then calculated from all estimated growth rates (Supplementary Table 2). Good agreement was found between estimates obtained from the simple exponential models and those obtained from the Bayesian P-spline models with overlapping credible intervals (Fig. 3a). The estimates obtained for rounds 6 and 7 showed the greatest level of disagreement with R estimated at 0.94 (0.92, 0.95) for the exponential model, but at 1.00 (0.97, 1.04) for the Bayesian P-spline model.

Similarly estimates of growth rates, R , and doubling/halving times were estimated for each individual round using the simple exponential model (Fig. 2b), the Bayesian P-spline model, and the Bayesian P-spline model fit to an increasing number of rounds (Supplementary Table 2). As before, there was good agreement between the estimated values of R with overlapping credible intervals for all rounds (Fig. 3b). However, there were clear differences in the posterior distributions of the estimates and their median values. Estimates of R obtained from the Bayesian P-spline model (fit to all 7 rounds) were higher than corresponding estimates from the exponential model over the periods of round 1 and round 2. There was a greater level of agreement between the two estimates for the periods of round 3, round 4, round 5, round 6 and round 7. Estimates of R obtained from the Bayesian P-spline model fit to subsets of the data (including all rounds up to the round for which R was estimated) showed similar point estimates and probability density to the Bayesian P-spline model fit to all data, with the exceptions of round 1 and round 3. This was most likely due to the fact that not enough data had been included to accurately estimate

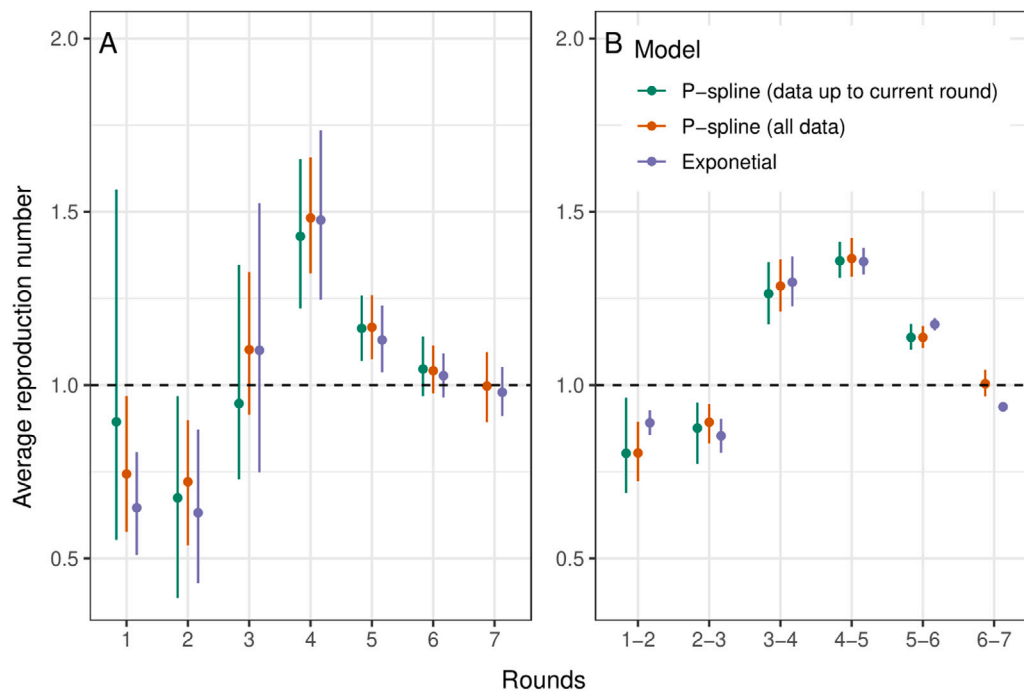


Fig. 3. Estimated R for fixed periods. Comparison of R calculated for fixed periods using the exponential model (purple), and the Bayesian P-spline model fit using the first 7 rounds of REACT-1 (orange) and using all rounds up to the round for which R is being estimated (green). A: Estimates over the period of pairs of subsequent rounds. B: Estimates over the period of each individual round. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the parameter ρ which controls the degree to which the growth rate changes over time.

3.2. Continuous estimates of growth rate and R

Instantaneous growth rate was estimated from the prevalence estimates of the Bayesian P-spline model (Fig. 4). Changes in growth rate at a time frame less than the duration of a round (a few weeks) were detected as well as changes in growth rate in the periods between rounds. There was a significant level of variation in growth rate during the periods of round 1 (1 May–1 June 2020) and round 2 (19 June–7 July 2020), a period over which England's lockdown restrictions were just beginning to ease. In the period between rounds 1 and 2 the instantaneous growth rate temporally became positive, indicative of a growing epidemic. The growth rate was approximately constant during the periods of round 3 (24 July–11 August 2020), 4 (20 August–8 September 2020) and 5 (18 September–5 October 2020), but in between these rounds there was a significant level of change with the growth rate increasing from round 3 to 4 and then decreasing significantly into round 5. During the period after round 5 to the end of round 7 (3 December 2020) there was significant variation in growth rate with it increasing and decreasing three times. We were also able to estimate the date at which prevalence began to increase, heralding the second wave of the pandemic. The day of minimum prevalence (Fig. 5) was found to be 24 July 2020 (13 July 2020, 11 August 2020).

Previously we estimated the reproduction number from the Bayesian P-spline model through a direct conversion of the average growth rate for a fixed period of time. However, using the continuous estimates for prevalence we were able to use a more appropriate method (see Methods) to calculate the rolling two week average instantaneous reproduction number, R_t , for the entire study (Fig. 6). As expected R_t behaved similarly to the instantaneous growth rate over the study period. There was a brief period in early-June 2020 in which R_t increased above 1 before decreasing below 1 going into late June 2020. R_t then once again increased, becoming greater than 1 towards the end of July 2020 where it remained until late October 2020 when it temporarily decreased below 1 until the end of November 2020. At

the end of the study period (3 December 2020) it had a value of 1.16 (0.95, 1.42) with a 93% probability of being greater than 1.

Sensitivity analyses were performed by fitting the Bayesian P-spline model to subsets of the positive samples. Models were fit to subsets of data only including positive samples in which: the individuals did not report having symptoms (asymptomatic individuals), both gene targets were detected as positive, and where samples would still be defined as positive using a lower N-gene Ct value of 35 to define positivity. The model estimates for growth rate and R_t exhibited mainly similar patterns to the model fit using all positive samples (Supplementary Figures 1 and 2), with any divergences having overlapping credible intervals. Modelled estimates of growth rate and R_t obtained from fitting to positives defined with a more stringent definition of positivity (N-gene Ct value less than 35), showed far less variation over time, though exhibited broadly similar patterns.

3.3. Assessing presence of under-fitting

The second-order random-walk prior distribution used in the Bayesian P-spline model acts to smooth the estimates of prevalence over time and thus limit over-fitting. However, if the estimates are smoothed too heavily then statistically significant features of the data will not be fitted to. This under-fitting of the Bayesian P-spline model was assessed by fitting the model multiple times with different knot sizes. The model parameter ρ that controls the second-order random-walk prior distribution should, if there is no under-fitting, increase linearly (as should its standard deviation) as the knot size increases (distance between knots increases linearly). At small knot sizes the value of ρ and its standard deviation increases linearly with the size of knots used (Fig. 7). However, for knot sizes larger than approximately 6 days the linearity of the relationship breaks down. The linear section of the graph included the knot size of 5 days which we have used for the main analysis and so no under-fitting was found to have occurred. The fit of the model for knot sizes of 2.5, 5 and 10 days were also inspected visually; no drastic difference was seen for the models with knot sizes of 2.5 and 5 days, but the model with a knot size of 10 days showed less

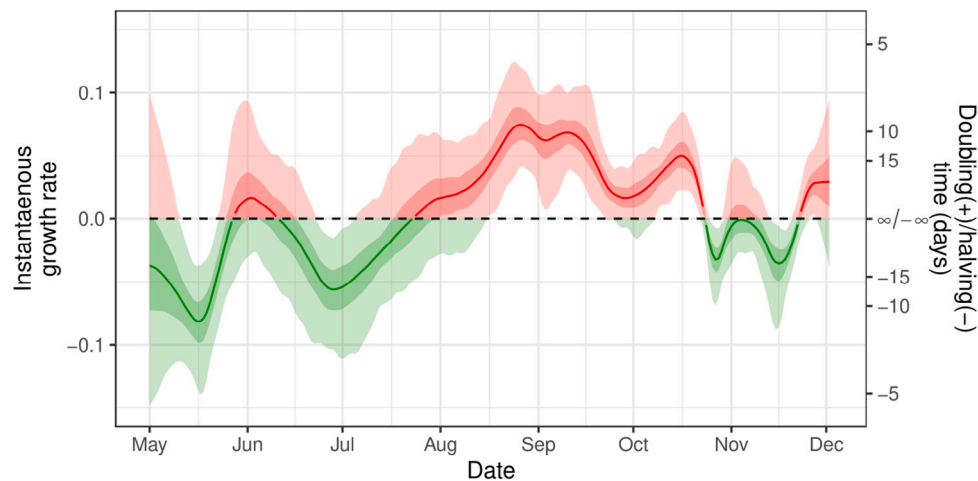


Fig. 4. Instantaneous growth rate. The instantaneous growth rate over the study period as inferred from the Bayesian P-spline model. The Y axis on the right shows the corresponding doubling/halving time corresponding to the growth rate on the left Y axis. The dotted line shows where growth rate = 0 and so the point of transition between epidemic growth and decline. The light shaded regions show the 95% credible interval and the dark shaded regions show the 50% credible interval. Red highlights regions of the credible interval with values of growth rate greater than 0. Green highlights regions of the credible interval with values of growth rate less than 0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

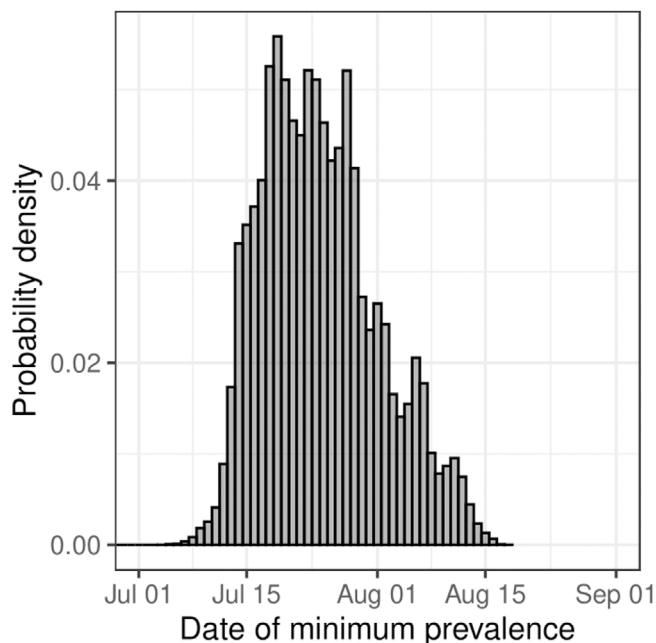


Fig. 5. Date of minimum prevalence. Posterior probability distribution for the date at which prevalence was at its lowest in England from 1 May 2020 to 3 December 2020 estimated from the Bayesian P-spline model.

variation in prevalence over time especially from late-October to mid-November 2020 most likely due to being in the under-fitting regime (Supplementary Figure 3).

3.4. Effect of increasing data

We investigated how the fit of the Bayesian P-spline model changed as data over a longer period of time was included (Fig. 8). Inclusion of new data can not only influence the fit of the model at the most recent points in time, but through changing the estimate of the parameter ρ can influence the model output over the entire period. The fit of the model to just round 1 of the data showed a significant level of variation with large uncertainty towards the beginning of the round and a possible uptick towards the end of the round. When round 2 and then

round 3 were included, the model fit to round 1 changed substantially, with a flatter prevalence during the final week of round 1 and an approximately exponential decrease in prevalence during the first two weeks. This was due to much smaller estimates of the standardised ρ , $\hat{\rho}$, (decreasing from 0.77 to 0.39 to 0.19 as rounds 2 and 3 were included respectively). As rounds 4, 5, 6 and 7 were included in the model fitting, the estimated value of $\hat{\rho}$ varied only slightly, reaching a low of 0.16 when round 5 was included and a high of 0.26 as round 6 was included. The estimated value of $\hat{\rho}$ including all 7 rounds was 0.22. The effect of $\hat{\rho}$ decreasing is most easily observed in the fit of the model at the beginning of round 1 (becoming closer to an exponential decrease) and between round 1 and round 2 (shrinking uncertainty in the increase in prevalence during the period).

3.5. Model fit to public case data

In order to compare the trends in publicly available case data with REACT-1 prevalence a Bayesian P-spline model was fit to the Public Health England Pillar 1 and Pillar 2 case (by specimen date) data (UK, 2022) allowing smoothed estimates of the daily number of cases to be obtained (Fig. 9). Broadly similar patterns to the modelled prevalence of REACT-1 were observed, but some differences were detected. The expected number of cases decreased steadily from early May 2020 reaching a minimum on 7 July (30 June, 16 July) 2020, earlier than the date in which prevalence as measured by REACT-1 reached a minimum (though with overlapping credible intervals). Further, estimates of instantaneous growth rate and R_t inferred from the model fit to case data showed far less variation over time than their REACT-1 model counterparts. Similarly to before, a small uptick in the growth rate was observed in June 2020, but the increase was not statistically significant and did not go above zero, the threshold for epidemic growth. The estimated R_t was below 1 until early July 2020; it then remained greater than 1 until early November 2020, plateauing at approximately 1.35 for the length of September 2020. R_t decreased steadily during the month of October 2020 in contrast to the trends observed in REACT-1 which measured a rapid decrease in R_t in mid-September 2020 and a temporary rise in mid-October 2020. During November and early December 2020 R_t , as measured from the case data, remained below 1, with no suggestion of an increasing epidemic on 3 December 2020 as was the case for the model fit to REACT-1 data.

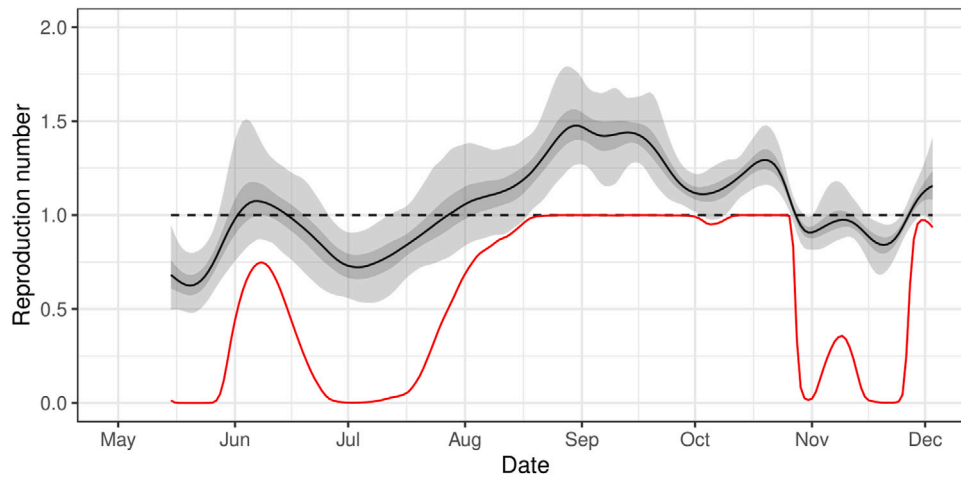


Fig. 6. Continuous estimate of Reproduction number The rolling two week average instantaneous reproduction number, R_t , over the duration of the study calculated from the prevalence estimates of the Bayesian P-spline model. The central estimate is shown (solid black line) with 50% (dark grey shaded region) and 95% (light grey shaded region) credible intervals. The dotted line shows where $R_t = 1$ and so the point of transition between epidemic growth and decline. The red line shows the probability that R_t is greater than 1 over time. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

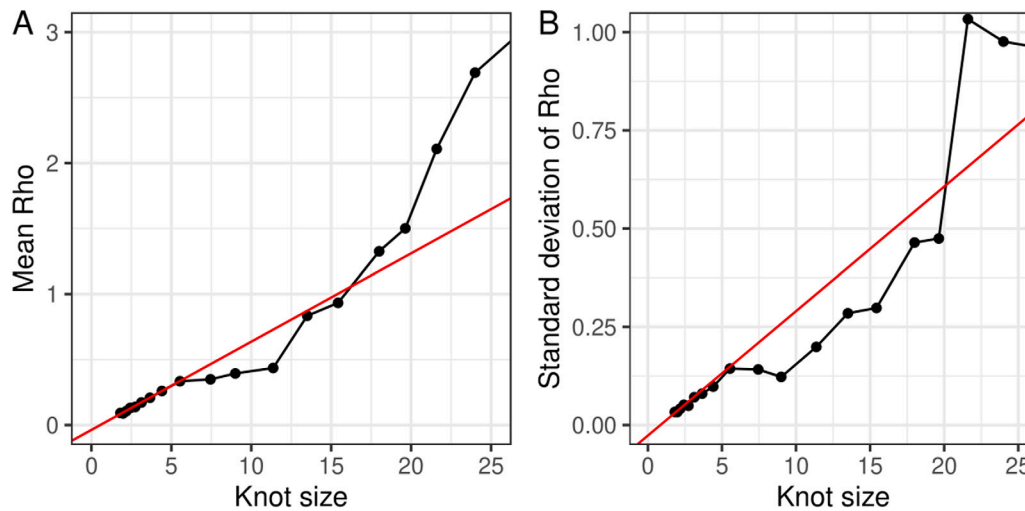


Fig. 7. Effect of knot size on the parameter ρ . The estimated value of the parameter ρ and its standard deviation from models using different knot sizes. The x-axis shows the knot size used in each model fit, the smaller the value the greater the number/density of knots. The red line shows a best fit line to the first four points in each graph and reflects the parameter values that would be expected if the model did not under fit when at a low density of knots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

We have developed a Bayesian P-spline model widely applicable to either prevalence or incidence epidemic time-series data and demonstrated its utility by fitting it to the first 7 rounds of the REACT-1 study. This period covered the end of the first national lockdown in England (Prime, 2020a) and the gradual easing of restrictions. The Bayesian P-spline model allows smooth estimates of prevalence over time (even in the periods of no data between rounds) that limits over-fitting through the inclusion of a second-order random-walk prior distribution, and concurrently avoids under-fitting through the inclusion of a high enough density of basis splines. From these prevalence estimates, average growth rates over a period of time (including daily for instantaneous growth rate), and average reproduction numbers over a period of time can straightforwardly be calculated. We see that average behaviour over a single round is broadly consistent with estimates based on a simple exponential model with overlapping credible intervals, though point estimates can differ substantially. The model has effectively included a prior distribution based on the previous and subsequent rounds' estimated prevalence and trend. However, the

Bayesian P-spline model is not limited to looking at average behaviour over the duration of a single round (a few weeks) and allows trends at finer timescales to also be quantified.

The REACT-1 testing procedure has its own set of limitations (Riley et al., 2021). However, it is relatively unbiased compared to traditional testing systems (Ricocha Peixoto et al., 2020) and thus allows a representative picture of the true nature of the epidemic within the population of England to be obtained. From modelled prevalence estimates we calculated that the date of minimum prevalence came in mid-July to early-August 2020. This estimate was later than estimates obtained using Public Health England Pillar 1 and Pillar 2 data (UK, 2022) but can be explained by increasing testing capacity obscuring underlying trends (Ricocha Peixoto et al., 2020), especially as Pillar 2 mass testing of the community only began on July 14 2020. Prevalence was seen to increase temporarily in late-May/early-June 2020 with a high probability of the instantaneous growth rate being positive for two to three weeks. This is possibly explained by the easing of restrictions at the beginning of June 2020 (Prime, 2020a) leading to a temporary increase in contact rates. The same temporary increase of growth rate was initially observed in the Office for National Statistics' Covid-19

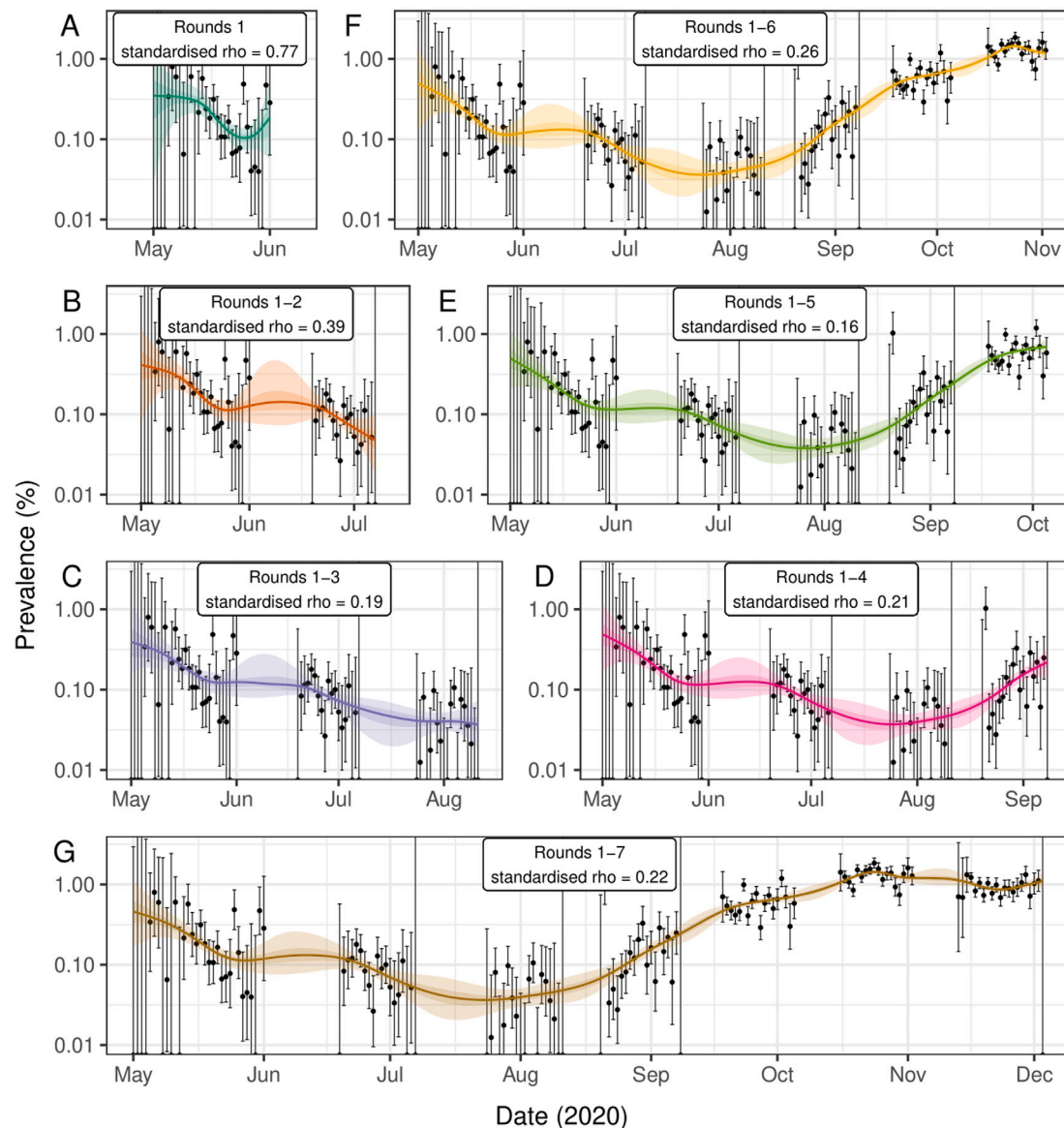


Fig. 8. Comparison of models fitted to subsets of the data. Bayesian P-spline model prevalence estimates for the REACT-1 data as each new round is included in the model. (A) Model fit to round 1 only (dark green). (B) model fit to rounds 1 and 2 (orange). (C) Model fit to rounds 1 to 3 (purple). (D) Model fit to rounds 1 to 4 (pink). (E) Model fit to rounds 1 to 5 (light green). (F) Model fit to rounds 1 to 6 (yellow). (G) Model fit to rounds 1 to 7 (brown). The standardised ρ value is given for each model and is standardised for a knot size of 5 days (the target knot size for each model). Central model estimate is shown (solid line) with central 50% (dark shaded region) and 95% (light shaded region) credible intervals. Daily estimates of prevalence (points) are shown with 95% credible intervals (error bars). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Infection Survey (ONS CIS) (Pouwels et al., 2020), but this increase was not present in the ONS CIS's subsequent modelled prevalence estimates (Kara, 2020).

Trends in prevalence over time were assessed by calculating the rolling 2-week average R_t . However, the REACT-1 study measures prevalence and not incidence, and so estimates of R_t were made under the assumption that R_t was constant for the duration an infected individual remains swab-positive. Individuals have been found to remain positive for a significant period of time (Eales et al., 2022) and so estimates of R_t are likely biased over periods in which R_t was changing. It is worth noting however that incidence is never truly measured and estimates of R_t using data obtained from other sources (such as mass testing) likely exhibit similar biases.

We additionally assessed trends over time by calculating the instantaneous growth rate. Growth rate was found to increase temporarily above 0 in late May 2020 returning to a relative minimum at the end of June 2020. Growth rate then gradually increased over the next

two months, likely reflecting the slow easing of restrictions including hospitality business (such as restaurants) opening from the 4 July 2020 (Department, 2020), with financial incentives offered to encourage people to eat out in August 2020 (HM, 2022). We observed a plateau in our estimate of growth rate at around 0.07 (~10 day doubling time) at the end of August 2020 that decreased to approximately 0.02 by the end of September. This might suggest that NPI's introduced in September including the "rule of 6" (limiting social interactions to no more than 6 people) on the 14 September 2020 (Cabinet, 2020a), and 10pm closures of hospitality businesses from the 24 September 2020 (Cabinet, 2020b) were only mildly effective at reducing transmission. During October 2020, growth rate began to increase, peaking at approximately 0.05 (~15 day doubling time) before decreasing temporarily in late-October, perhaps due to the week-long school "half-term" holiday in late-October temporarily reducing contacts. During November 2020 and the implementation of a second national lockdown in England (Prime, 2020b) growth rate stayed below 0, until in late

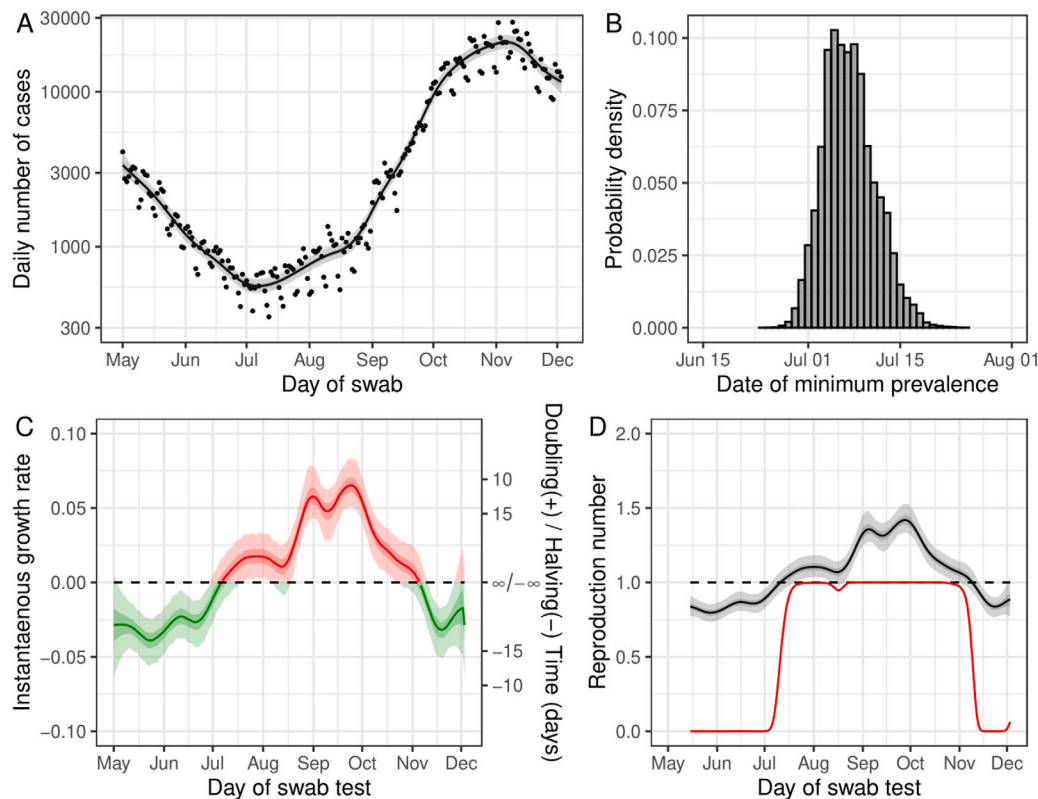


Fig. 9. Model fit to publicly available case data. (A) Fit of the Bayesian P-spline model to publicly available case data. The central estimate (solid line) is shown with 50% (dark grey shaded region) and 95% (light grey shaded region) credible intervals and with raw data (points). (B) Posterior probability distribution for the date at which case numbers were at its lowest in England, estimated from the Bayesian P-spline model. (C) The instantaneous growth rate over the study period as inferred from the Bayesian P-spline model. The Y axis on the right shows the corresponding doubling/halving time corresponding to the growth rate on the left Y axis. The dotted line shows where growth rate = 0 and so the point of transition between epidemic growth (red region) and decline (green region). The central estimate (solid line) is shown with 95% (light shaded region) and 50% (dark shaded region) credible intervals. (D) The rolling two week average instantaneous Reproduction number over the duration of the study calculated from the prevalence estimates of the Bayesian P-spline model. The central estimate (solid line) is shown with 95% (light grey shaded region) and 50% (light grey shaded region) credible intervals. Also shown is the probability that the reproduction number is greater than 1 over time (red line). The dashed line shows $R_t = 1$, the points of transition between epidemic growth and decline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

November growth rate increased to approximately 0.03. This was potentially due to the introduction and subsequent growth of the more transmissible Alpha variant in England (Volz et al., 2021). It is worth noting that our estimates of growth rate from prevalence will have lagged the actual growth rate of incidence. However, the trends in time we report here appeared to be 2–4 weeks ahead of estimates published by the Scientific Advisory Group for Emergencies, likely because they were based on a variety of different data sources (UK Health, 2022).

The model applied to epidemic time-series data by the ONS CIS is comparable to ours (Pouwels et al., 2020) though relies on thin-plate regression splines instead of penalised-splines. As in all Bayesian models it is important that the chosen prior distribution does not incorrectly influence the posterior distribution. We demonstrated how the use of a first-order random-walk prior distribution would improperly favour a constant prevalence in the absence of data. We avoided this problem by instead penalising the 1st derivative of the link function, which favours a constant growth rate. Although this is the most applicable prior distribution for the model, penalising the 1st derivative of the link function has its own limitations. Firstly, changes in growth rate are likely to be smoothed over time, whereas many policy changes are introduced overnight and so a step change in growth rate is probably more realistic. Additionally, during an epidemic with no changes in policy or behaviour, growth rate is only approximately constant over small timescales due to a decreasing proportion of susceptible individuals over longer timescales. We do not expect this to have been a problem for the current data set as a previous study estimated that only 5.56 percent of people tested positive for IgG antibodies to SARS-CoV-2

during October 27–November 10 2020 (Ward et al., 2021), suggesting there had been a limited number of infections prior to this period.

The Bayesian P-spline model has clear advantages in estimating the underlying temporal patterns of prevalence when there are large gaps in time-series data, interpolating between periods of missing data. It can further be applied to more general epidemic time-series as we have demonstrated with the publicly available case data. It is a particularly useful tool when the growth rate can change over smaller timescales, as in the current pandemic, as the model has been shown to be sensitive to these changes whereas a more simplistic exponential model can miss these fine details. Furthermore, with the large amounts of data available on variants of infection (COVID-19, 2020; Volz et al., 2021; Eales et al., 2021; du Plessis et al., 2021) the model could easily be applied to the relative proportion of two variants in order to test for a changing fitness advantage over time. Given that growth rate and R_t can be estimated from the prevalence estimates of the model it also has potential use as a forecasting tool, though more work would have to be completed looking at the sensitivity of the final growth rate estimates and optimising the window over which growth rate, and R_t , are estimated in order to improve predictive power.

5. Conclusion

In summary, we have developed a versatile model for use with epidemic time series data and applied it to the first 7 rounds of the REACT-1 study. The REACT-1 data, due to its random sampling procedure, is relatively unbiased compared to other sources of epidemic data in England over this same period. Through the application of the

Bayesian P-spline model to the REACT-1 data we have been able to infer the state of the SARS-CoV-2 pandemic within England from 1 May 2020 to 3 December 2020, a period that saw numerous changes in restrictions and testing procedures. The trends we report contrast with the reported trends based on other data sets due to potential sources of bias. This study highlights the importance of not only obtaining relatively unbiased data, such as the REACT-1 data, but also in pairing it with statistically robust methods in order to effectively track the within-country dynamics of a pandemic such as the current SARS-CoV-2 pandemic.

CRediT authorship contribution statement

Oliver Eales: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Kylie E.C. Ainslie:** Data curation, Writing – review & editing. **Caroline E. Walters:** Data curation, Writing – review & editing. **Haowei Wang:** Data curation, Writing – review & editing. **Christina Atchison:** Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Deborah Ashby:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Christl A. Donnelly:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Graham Cooke:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Wendy Barclay:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Helen Ward:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Ara Darzi:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Paul Elliott:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Steven Riley:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Access to individual level REACT-1 data is restricted due to ethical and security considerations. Summary statistics and data, including the daily number of positive tests and daily total number of tests, are available at <https://github.com/mrc-ide/reactidd/tree/master/inst/extdata>. Additional summary statistics and results from the REACT-1 programme are also available at <https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/real-time-assessment-of-community-transmission-findings/>. REACT-1 study materials are available for each round at <https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/react-1-study-materials/>.

Acknowledgements

SR, CAD acknowledge support: MRC Centre for Global Infectious Disease Analysis, United Kingdom, National Institute for Health Research (NIHR), United Kingdom Health Protection Research Unit (HPRU), Wellcome Trust (200861/Z/16/Z, 200187/Z/15/Z), and Centres for Disease Control and Prevention, USA (US, U01CK0005-01-02). GC is supported by an NIHR Professorship. PE is Director of the MRC Centre for Environment and Health (MR/L01341X/1, MR/S019669/1).

PE acknowledges support from Health Data Research UK (HDR UK); the NIHR Imperial Biomedical Research Centre, United Kingdom; NIHR HPRUs in Chemical and Radiation Threats and Hazards, and Environmental Exposures and Health; the British Heart Foundation Centre for Research Excellence at Imperial College London (RE/18/4/34215); and the UK Dementia Research Institute at Imperial (MC_PC_17114). We thank The Huo Family Foundation, United Kingdom for their support of our work on COVID-19. We thank key collaborators on this work – Ipsos MORI: Kelly Beaver, Sam Clemens, Gary Welch, Nicholas Gilby, and Kelly Ward; Institute of Global Health Innovation at Imperial College: Gianluca Fontana, Dr Hutan Ashrafian, Sutha Satkunarajah and Lenny Naar; North West London Pathology and Public Health England for help in calibration of the laboratory analyses; NHS Digital for access to the NHS register; and the Department of Health and Social Care for logistic support. SR acknowledges helpful discussion with attendees of meetings of the UK Government Office for Science (GO-Science) Scientific Pandemic Influenza – Modelling (SPI-M) committee.

Funding

The study was funded by the Department of Health and Social Care in England.

Code availability

All code is available in the reactidd R package available at <https://github.com/mrc-ide/reactidd/>.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.epidem.2022.100604>. Supplementary tables are available in the supporting document ‘SupplementaryTables.xlsx’. Supplementary figures are available in the supporting document ‘SupplementaryFigures.docx’.

References

- van Algemene Zaken, M., 2021. Night-time curfew as of saturday 23 january. Accessed: 2022-1-27. <https://www.government.nl/latest/news/2021/01/22/night-time-curfew-as-of-saturday-23-january>.
- Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S.A., Zhang, T., et al., 2020. Epidemiology and transmission of COVID-19 in shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv.
- de Boor, C., 1978. A Practical Guide To Spline, Vol. 27. American Mathematical Society.
- Cabinet Office, 2020a. Coronavirus (COVID-19): What has changed – 22 september. Accessed: 2022-1-28. <https://www.gov.uk/government/news/coronavirus-covid-19-what-has-changed-22-september>.
- Cabinet Office, 2020b. Coronavirus (COVID-19): What has changed – 9 September. Accessed: 2022-1-28. <https://www.gov.uk/government/news/coronavirus-covid-19-what-has-changed-9-september>.
- Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S., 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* 178 (9), 1505–1512.
- COVID-19 Genomics U.K. (COG-UK) consortiumcontact@cogconsortium.uk, 2020. An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* 1 (3), e99–e100.
- Department for Business and Energy & Industrial Strategy, 2020. Pubs, restaurants and hairdressers to reopen from 4 July. Accessed: 2022-1-28. <https://www.gov.uk/government/news/pubs-restaurants-and-hairdressers-to-reopen-from-4-july>.
- Drew, D.A., Nguyen, L.H., Steves, C.J., Menni, C., Freydin, M., Varsavsky, T., Sudre, C.H., Cardoso, M.J., Ourselin, S., Wolf, J., Spector, T.D., Chan, A.T., Consortium, C., 2020. Rapid implementation of machine technology for real-time epidemiology of COVID-19. *Science* 368 (6497), 1362–1367.
- Eales, O., Page, A.J., Tang, S.N., Walters, C.E., Wang, H., Haw, D., Trotter, A.J., Viet, T.L., Foster-Nyarko, E., Prosolek, S., Atchison, C., Ashby, D., Cooke, G., Barclay, W., Donnelly, C.A., O’Grady, J., Volz, E., Consortium, T.C.-G.U.C.-U., Darzi, A., Ward, H., Elliott, P., Riley, S., 2021. SARS-CoV-2 lineage dynamics in England from january to march 2021 inferred from representative community samples. <http://dx.doi.org/10.1101/2021.05.08.21256867>, medRxiv.

- Eales, O., Walters, C.E., Wang, H., Haw, D., Ainslie, K.E.C., Atchison, C.J., Page, A.J., Prosolek, S., Trotter, A.J., Le Viet, T., Alikhan, N.-F., Jackson, L.M., Ludden, C., Ashby, D., Donnelly, C.A., Cooke, G., Barclay, W., Ward, H., Darzi, A., Elliott, P., Riley, S., Consortium, C.-G.U., 2022. Characterising the persistence of RT-PCR positivity and incidence in a community survey of SARS-CoV-2. *Wellcome Open Res.* 7, 102.
- Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. *Stat. Sci.* 11 (2), 89–121.
- Elliott, P., Haw, D., Wang, H., Eales, O., Walters, C.E., Ainslie, K.E.C., Atchison, C., Fronterre, C., Diggle, P.J., Page, A.J., Trotter, A.J., Prosolek, S.J., Consortium11, C.-G.U.C.-U., Ashby, D., Donnelly, C.A., Barclay, W., Taylor, G., Cooke, G., Ward, H., Darzi, A., Riley, S., Robson, S.C., Loman, N.J., Connor, T.R., Golubchik, T., Martinez Nunez, R.T., Ludden, C., Corden, S., Johnston, I., Bonsall, D., Smith, C.P., Awan, A.R., Bucca, G., Torok, M.E., Saeed, K., Prieto, J.A., Jackson, D.K., Hamilton, W.L., Snell, L.B., Moore, C., Harrison, E.M., Goncalves, S., Jackson, L.M., Goodfellow, I.G., Fairley, D.J., Loose, M.W., Watkins, J., Livett, R., Moses, S., Amato, R., Nicholls, S., Bull, M., Smith, D.L., Barrett, J., Aanensen, D.M., Curran, M.D., Parmar, S., Aggarwal, D., Shepherd, J.G., Parker, M.D., Glaysher, S., Bashton, M., Underwood, A.P., Pacchiari, N., Lovesson, K.F., Carabelli, A.M., Templeton, K.E., Langford, C.F., Sillitoe, J., de Silva, T.I., Wang, D., Kwiatkowski, D., Rambaut, A., O'Grady, J., Cottrell, S., Holden, M.T.G., Thomson, E.C., Osman, H., Andersson, M., Chauhan, A.J., Hassan-Ibrahim, M.O., Lawnczak, M., Gupta, R.K., Alderton, A., Chand, M., Constantinidou, C., Unnikrishnan, M., Darby, A.C., Hiscox, J.A., Paterson, S., Martincorena, I., Robertson, D.L., Volz, E.M., Page, A.J., Pybus, O.G., Bassett, A.R., Ariani, C.V., Spencer Chapman, M.H., Li, K.K., Shah, R.N., Jesudason, N.G., Taha, Y., McHugh, M.P., Dewar, R., Jahun, A.S., McMurray, C., Pandey, S., McKenna, J.P., Nelson, A., Young, G.R., McCann, C.M., Elliott, S., Lowe, H., Temperton, B., Roy, S., Price, A., Rey, S., Wyles, M., Rooke, S., Shaaban, S., de Cesare, M., Letchford, L., Silveira, S., Pelosi, E., Wilson-Davies, E., Hosmillo, M., O'Toole, A., Hesketh, A.R., Stark, R., du Plessis, L., Ruis, C., Adams, H., Bourgeois, Y., Michell, S.L., Grammatopoulos, D., Edgeworth, J., Preucor, J., Todd, J.A., Fraser, C., Buck, D., John, M., Kay, G.L., Palmer, S., Peacock, S.J., Heyburn, D., Weldon, D., Robinson, E., McNally, A., Muir, P., Vipond, I.B., Boyes, J., Sivaprakasam, V., Saluja, T., Dervisevic, S., Meader, E.J., Park, N.R., Oliver, K., Jeffries, A.R., Ott, S., da Silva Filipe, A., Simpson, D.A., Williams, C., Masoli, J.A., Knight, B.A., Jones, C.R., Koshy, C., Ash, A., Casey, A., Bosworth, A., Ratcliffe, L., Xu-McCrae, L., Pymont, H.M., Hutchings, S., Berry, L., Jones, K., Halstead, F., Davis, T., Holmes, C., Iturriza-Gomara, M., Lucaci, A.O., Randell, P.A., Cox, A., Madona, P., Harris, K.A., Brown, J.R., Mahungu, T.W., Irish-Tavares, D., Haque, T., Hart, J., Witele, F., Fenton, M.L., Liggett, S., Graham, C., Swindells, E., Collins, J., Eltringham, G., Campbell, S., McClure, P.C., Clark, G., Sloan, T.J., Jones, C., Lynch, J., Warne, B., Leonard, S., Durham, J., Williams, T., Haldenby, S.T., Storey, N., Alikhan, N.-F., Holmes, N., Moore, C., Carlisle, M., Perry, M., Craine, N., Lyons, R.A., Beckett, A.H., Goudarzi, S., Fearn, C., Cook, K., Dent, H., Paul, H., Davies, R., Blane, B., Girgis, S.T., Beale, M.A., Bellis, K.L., Dorman, M.J., Drury, E., Kane, L., Kay, S., McGuigan, S., Nelson, R., Prestwood, L., Rajatileka, S., Batra, R., Williams, R.J., Kristiansen, M., Green, A., Justice, A., Mahananda, A.I.K., Samaraweera, B., Hadjirin, N.F., Quick, J., Poplawski, R., Kermack, L.M., Reynolds, N., Hall, G., Chaudhry, Y., Pinckert, M.L., Georgiana, I., Moll, R.J., Thornton, A., Myers, R., Stockton, J., Williams, C.A., Yew, W.C., Trotter, A.J., Trebes, A., MacIntyre-Cockett, G., Birchley, A., Adams, A., Plimmer, A., Gatica-Wilcox, B., McKerr, C., Hilvers, E., Jones, H., Asad, H., Coombes, J., Evans, J.M., Fina, L., Gilbert, L., Graham, L., Cronin, M., Kumziene-Summerhayes, S., Taylor, S., Jones, S., Groves, D.C., Zhang, P., Gallis, M., Louka, S.F., Starinskij, I., Keatley, J.-P., Singer, J.B., de Oliveira Martins, L., Yeats, C.A., Abudahab, K., Taylor, B.E., Menegazzo, M., Danesh, J., Hogsden, W., Eldirdiri, S., Kenyon, A., Mason, J., Robinson, T.I., Holmes, A., Price, J., Hartley, J.A., Curran, T., Mather, A.E., Shankar, G., Jones, R., Howe, R., Morgan, S., Wastenge, E., Chapman, M.R., Mookerjee, S., Stanley, R., Smith, W., Peto, T., Eyre, D., Crook, D., Vernet, G., Kitchen, C., Gulliver, H., Merrick, I., Guest, M., Munn, R., Bradley, D.T., Wyatt, T., Beaver, C., Foulser, L., Palmer, S., Churcher, C.M., Brooks, E., Smith, K.S., Galai, K., McManus, G.M., Bolt, F., Coll, F., Meadows, L., Attwood, S.W., Davies, A., De Lacy, E., Downing, F., Edwards, S., Scarlett, G.P., Jeremiah, S., Smith, N., Leek, D., Sridhar, S., Forrest, S., Cormie, C., Gill, H.K., Dias, J., Higginson, E.E., Maes, M., Young, J., Wantoch, M., Jamrozny, D., Lo, S., Patel, M., Hill, V., Bewshea, C.M., Ellard, S., Auckland, C., Harrison, I., Bishop, C., Chalker, V., Richter, A., Beggs, A., Best, A., Percival, B., Mirza, J., Megram, O., Mayhew, M., Crawford, L., Ashcroft, F., Moles-Garcia, E., Cumley, N., Hopes, R., Asamaphan, P., Niebel, M.O., Gunson, R.N., Bradley, A., Maclean, A., Mollett, G., Blacow, R., Bird, P., Helmer, T., Fallon, K., Tang, J., Hale, A.D., Macfarlane-Smith, L.R., Harper, K.L., Carden, H., Machin, N.W., Jackson, K.A., Ahmad, S.S.Y., George, R.P., Turtle, L., O'Toole, E., Watts, J., Breen, C., Cowell, A., Alcolea-Medina, A., Charalampous, T., Patel, A., Levett, L.J., Heaney, J., Rowan, A., Taylor, G.P., Shah, D., Atkinson, L., Lee, J.C., Westhorpe, A.P., Jannoo, R., Lowe, H.L., Karamani, A., Ensell, L., Chatterton, W., Pusok, M., Dadrach, A., Symmonds, A., Sluga, G., Molnar, Z., Baker, P., Bonner, S., Essex, S., Barton, E., Padgett, D., Scott, G., Greenaway, J., Payne, B.A., Burton-Fanning, S., Waugh, S., Raviprakash, V., Sheriff, N., Blakey, V., Williams, L.-A., Moore, J., Stonehouse, S., Smith, L., Davidson, R.K., Bedford, L., Coupland, L., Wright, V., Chappell, J.G., Tsoileridis, T., Ball, J., Khakh, M., Fleming, V.M., Lister, M.M., Howson-Wells, H.C., Berry, L., Boswell, T., Joseph, A., Willingham, I., Duckworth, N., Walsh, S., Wise, E., Moore, N., Mori, M., Cortes, N., Kidd, S., Bmbs, R.W., Gifford, L., Bicknell, K., Wylie, S., Lloyd, A., Impey, R., Malone, C.S., Cogger, B.J., Levene, N., Monaghan, L., Keeley, A.J., Partridge, D.G., Raza, M., Evans, C., Johnson, K., Betteridge, E., Farr, B.W., Goodwin, S., Quail, M.A., Scott, C., Shirley, L., Thurston, S.A., Rajan, D., Bronner, I.F., Aigrain, L., Redshaw, N.M., Lensing, S.V., McCarthy, S., Makunin, A., Balcazar, C.E., Gallagher, M.D., Williamson, K.A., Stanton, T.D., Michelsen, M.L., Warwick-Dugdale, J., Manley, R., Farbos, A., Harrison, J.W., Sambles, C.M., Studholme, D.J., Lackenby, A., Mbisa, T., Platt, S., Miah, S., Bibby, D., Manso, C., Hubb, J., Dabrera, G., Ramsay, M., Bradshaw, D., Schaefer, U., Groves, N., Gallagher, E., Lee, D., Williams, D., Ellaby, N., Hartman, H., Manesis, N., Patel, V., Ledesma, J., Twohig, K.A., Allara, E., Pearson, C., Cheng, J.K.J., Bridgewater, H.E., Frost, L.R., Taylor-Joyce, G., Brown, P.E., Tong, L., Broos, A., Mair, D., Nichols, J., Carmichael, S.N., Smollett, K.L., Nomikou, K., Aranday-Cortes, E., Johnson, N., Nickbakhsh, S., Vamos, E.E., Hughes, M., Rainbow, L., Eccles, R., Nelson, C., Whitehead, M., Gregory, R., Gemmell, M., Wierzbicki, C., Webster, H.J., Fisher, C.L., Signell, A.W., Betancor, G., Wilson, H.D., Nebbia, G., Flaviani, F., Cerda, A.C., Merrill, T.V., Wilson, R.E., Cotic, M., Bayzid, N., Thompson, T., Acheson, E., Rushton, S., O'Brien, S., Baker, D.J., Rudder, S., Aydin, A., Sang, F., Debebe, J., Francois, S., Vasylyeva, T.I., Zamudio, M.E., Gutierrez, B., Marchbank, A., Maksimovic, J., Spellman, K., McCluggage, K., Morgan, M., Beer, R., Affi, S., Workman, T., Fuller, W., Bresner, C., Anygal, A., Green, L.R., Parsons, P.J., Tucker, R.M., Brown, R., Whiteley, M., Bonfield, J., Puethe, C., Whitwham, A., Liddle, J., Rowe, W., Siveroni, I., Le-Viet, T., Gaskin, A., Johnson, R., 2021. Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the delta variant. *Science* eabl9551.
- Fang, X., Fang, B., Wang, C., Xia, T., Bottai, M., Fang, F., Cao, Y., 2019. Comparison of frequentist and Bayesian generalized additive models for assessing the association between daily exposure to fine particles and respiratory mortality: A simulation study. *Int. J. Environ. Res. Public Health* 16 (5).
- Goicoa, T., Adin, A., Etxeberria, J., Militino, A.F., Ugarte, M.D., 2019. Flexible Bayesian P-splines for smoothing age-specific spatio-temporal mortality patterns. *Stat. Methods Med. Res.* 28 (2), 384–403.
- Gostic, K.M., McGough, L., Baskerville, E.B., Abbott, S., Joshi, K., Tedijanto, C., Kahn, R., Niehus, R., Hay, J.A., De Salazar, P.M., Hellewell, J., Meakin, S., Munday, J.D., Bosse, N.I., Sherratt, K., Thompson, R.N., White, L.F., Huisman, J.S., Scire, J., Bonhoeffer, S., Stadler, T., Wallinga, J., Funk, S., Lipsitch, M., Cobey, S., 2020. Practical considerations for measuring the effective reproductive number. *Rt. PLoS Comput. Biol.* 16 (12), e1008409.
- Hillary, L.S., Farkas, K., Maher, K.H., Lucaci, A., Thorpe, J., Distaso, M.A., Gaze, W.H., Paterson, S., Burke, T., Connor, T.R., McDonald, J.E., Malham, S.K., Jones, D.L., 2021. Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK. *Water Res.* 200, 117214.
- HM Revenue and Customs, 2022. Get a discount with the eat out to help out scheme. Accessed: 2022-1-27. <https://www.gov.uk/guidance/get-a-discount-with-the-eat-out-to-help-out-scheme>.
- Hoffman, M.D., Gelman, A., 2011. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *arXiv:1111.4246*. stat.CO.
- Kara Steel And, 2020. Coronavirus (COVID-19) Infection Survey pilot - Office for National Statistics. Office for National Statistics, Accessed: 2022-1-28. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscommunity19infectionsurveys/pilot/englandwalesandnorthernireland2october2020>.
- van de Kastele, J., Eilers, P.H.C., Wallinga, J., 2019. Nowcasting the number of new symptomatic cases during infectious disease outbreaks using constrained P-spline smoothing. *Epidemiology* 30 (5), 737–745.
- Lang, S., Brezger, A., 2004. Bayesian P-Splines. *J. Comput. Graph. Stat.* 13 (1), 183–212.
- Mills, C.E., Robins, J.M., Lipsitch, M., 2004. Transmissibility of 1918 pandemic influenza. *Nature* 432 (7019), 904–906.
- O'Connor, R.C., Wetherall, K., Cleare, S., McClelland, H., Melson, A.J., Niedzwiedz, C.L., Platt, S., Scowcroft, E., Watson, B., Zortea, T., Ferguson, E., Robb, K.A., 2021. Mental health and well-being during the COVID-19 pandemic: longitudinal analyses of adults in the UK COVID-19 mental health & wellbeing study. *Br. J. Psychiatry* 218.
- Osei, F.B., Duker, A.A., Stein, A., 2012. Bayesian structured additive regression modeling of epidemic data: application to cholera. *BMC Med. Res. Methodol.* 12, 118.
- Parag, K.V., 2021. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Comput. Biol.* 17 (9), e1009347.
- du Plessis, L., McCrone, J.T., Zarebski, A.E., Hill, V., Ruis, C., Gutierrez, B., Raghwan, J., Ashworth, J., Colquhoun, R., Connor, T.R., Faria, N.R., Jackson, B., Loman, N.J., O'Toole, A., Nicholls, S.M., Parag, K.V., Scher, E., Vasylyeva, T.I., Volz, E.M., Watts, A., Bogoch, I.I., Khan, K., Consortium, C.-G.U.C.-U., Aanensen, D.M., Kraemer, M.U.G., Rambaut, A., Pybus, O.G., 2021. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* 371 (6530), 708–712.
- Poletti, P., Ajelli, M., Merler, S., 2011. The effect of risk perception on the 2009 H1N1 pandemic influenza dynamics. *PLoS One* 6 (2), e16460.

- Pouwels, K.B., House, T., Pritchard, E., Robotham, J.V., Birrell, P.J., Gelman, A., Vihta, K.-D., Bowers, N., Boreham, I., Thomas, H., Lewis, J., Bell, I., Bell, J.I., Newton, J.N., Farrar, J., Diamond, I., Benton, P., Walker, A.S., Team, C.-I.S., 2021. Community prevalence of SARS-CoV-2 in England from april to november, 2020: results from the ONS coronavirus infection survey. *Lancet Public Health* 6 (1), e30–e38.
- Pouwels, K.B., House, T., Robotham, J.V., Birrell, P.J., Gelman, A., Bowers, N., Boreham, I., Thomas, H., Lewis, J., Bell, I., Bell, J.I., Newton, J.N., Farrar, J., Diamond, I., Benton, P., Walker, A.S., the COVID-19 Infection Survey team, 2020. Community prevalence of SARS-CoV-2 in England: Results from the ONS coronavirus infection survey pilot.
- Premier of Victoria's Office, 2022. Extended melbourne lockdown to keep victorians safe. Accessed: 2022-1-27. <https://www.premier.vic.gov.au/extended-melbourne-lockdown-keep-victorians-safe-0>.
- Prime Minister's Office, 2020a. PM: Six people can meet outside under new measures to ease lockdown. Accessed: 2022-1-28. <https://www.gov.uk/government/news/pm-six-people-can-meet-outside-under-new-measures-to-ease-lockdown>.
- Prime Minister's Office, 2020b. Prime minister announces new national restrictions. Accessed: 2022-1-23. <https://www.gov.uk/government/news/prime-minister-announces-new-national-restrictions>.
- Prime Minister's Office, 2021. Prime minister announces national lockdown. Accessed: 2022-1-27. <https://www.gov.uk/government/news/prime-minister-announces-national-lockdown>.
- Ricoca Peixoto, V., Nunes, C., Abrantes, A., 2020. Epidemic surveillance of Covid-19: Considering uncertainty and Under-Ascertainment. *Port. J. Public Health* 38 (1), 23–29.
- Riley, S., Ainslie, K.E.C., Eales, O., Walters, C.E., Wang, H., Atchison, C., Fronte, C., Diggle, P.J., Ashby, D., Donnelly, C.A., Cooke, G., Barclay, W., Ward, H., Darzi, A., Elliott, P., 2021. Resurgence of SARS-CoV-2: detection by community viral surveillance. *Science* 372, 990–995.
- Riley, S., Atchison, C., Ashby, D., Donnelly, C.A., Barclay, W., Cooke, G., Ward, H., Darzi, A., Elliott, P., Group, R.S., et al., 2020. REal-time assessment of community transmission (REACT) of SARS-CoV-2 virus: Study protocol. *Wellcome Open Res.* 5 (200), 200.
- Sharot, T., 1986. Weighting survey results. *J. Mark. Res. Soc.* 28 (3), 269–284.
- UK Government, 2022. Official UK coronavirus dashboard. Accessed: 2022-1-28. <https://coronavirus.data.gov.uk/details/cases?areaType=nation&areaName=England>.
- UK Health Security Agency, 2022. The R number and growth rate in the UK. Accessed: 2022-1-28. <https://www.gov.uk/guidance/the-r-number-in-the-uk>.
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P.-C., 2019. Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of MCMC. *arXiv:1903.08008*. [stat.CO](https://arxiv.org/abs/1903.08008).
- Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, A., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C.V., Boyd, O., Loman, N.J., McCrone, J.T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D.K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D.P., consortium, C.-G.U.C.-U., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., Ferguson, N.M., 2021. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 593, 266–269.
- Wallinga, J., Lipsitch, M., 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* 274 (1609), 599–604.
- Wallinga, J., Teunis, P., 2004. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* 160 (6), 509–516.
- Ward, H., Atchison, C., Whitaker, M., Donnelly, C.A., Riley, S., Ashby, D., Darzi, A., Barclay, W.S., Cooke, G., Elliott, P., for the REACT study team, 2021. Increasing SARS-CoV-2 antibody prevalence in England at the start of the second wave: REACT-2 round 4 cross-sectional study in 160,000 adults. <http://dx.doi.org/10.1101/2021.07.21.21260926>, medRxiv.
- White, L.F., Pagano, M., 2008. Transmissibility of the influenza virus in the 1918 pandemic. *PLoS One* 3 (1), e1498.