

Inferring community transmission of SARS-CoV-2 in the United Kingdom using the ONS COVID-19 Infection Survey

Ruth McCabe ^{a, b, c, *}, Gabriel Danelian ^c, Jasmina Panovska-Griffiths ^{c, d, e},
Christl A. Donnelly ^{a, b, e, f}

^a Department of Statistics, University of Oxford, UK

^b National Institute for Health and Care Research Health Protection Research Unit in Emerging and Zoonotic Infections, UK

^c United Kingdom Health Security Agency, UK

^d The Queen's College, University of Oxford, UK

^e The Pandemic Sciences Institute, University of Oxford, UK

^f MRC Centre for Global Infectious Disease Analysis, Imperial College London, UK

ARTICLE INFO

Article history:

Received 25 October 2023

Received in revised form 24 January 2024

Accepted 29 January 2024

Available online 30 January 2024

Handling Editor: Dr Yijun Lou

Keywords:

Effective reproduction number

Instantaneous growth rate

SARS-CoV-2

COVID-19

ONS COVID-19 Infection Survey

Surveillance

ABSTRACT

Key epidemiological parameters, including the effective reproduction number, $R(t)$, and the instantaneous growth rate, $r(t)$, generated from an ensemble of models, have been informing public health policy throughout the COVID-19 pandemic in the four nations of the United Kingdom of Great Britain and Northern Ireland (UK). However, estimation of these quantities became challenging with the scaling down of surveillance systems as part of the transition from the “emergency” to “endemic” phase of the pandemic.

The Office for National Statistics (ONS) COVID-19 Infection Survey (CIS) provided an opportunity to continue estimating these parameters in the absence of other data streams. We used a penalised spline model fitted to the publicly-available ONS CIS test positivity estimates to produce a smoothed estimate of the prevalence of SARS-CoV-2 positivity over time. The resulting fitted curve was used to estimate the “ONS-based” $R(t)$ and $r(t)$ across the four nations of the UK. Estimates produced under this model are compared to government-published estimates with particular consideration given to the contribution that this single data stream can offer in the estimation of these parameters.

Depending on the nation and parameter, we found that up to 77% of the variance in the government-published estimates can be explained by the ONS-based estimates, demonstrating the value of this singular data stream to track the epidemic in each of the four nations. We additionally find that the ONS-based estimates uncover epidemic trends earlier than the corresponding government-published estimates.

Our work shows that the ONS CIS can be used to generate key COVID-19 epidemiological parameters across the four UK nations, further underlining the enormous value of such population-level studies of infection. This is not intended as an alternative to ensemble modelling, rather it is intended as a potential solution to the aforementioned challenge faced by public health officials in the UK in early 2022.

© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author. Department of Statistics, 24 – 29 St Giles', Oxford, OX1 3LB, UK.

E-mail address: ruth.mccabe17@imperial.ac.uk (R. McCabe).

Peer review under responsibility of KeAi Communications Co., Ltd.

1. Introduction

Key epidemiological parameters, including the effective reproduction number, $R(t)$, and the instantaneous growth rate, $r(t)$, have been used to inform public health policy throughout the COVID-19 pandemic (Anderson et al., 2020; Government Publishes Latest R Number, 2020; Slides and Datasets to Accompany Coronavirus Press Conference: 22 October 2020, 2020; Slides to Accompany Coronavirus Press Conference: 11 May 2020, 2020). Estimation of these quantities by public health officials in the United Kingdom (UK) has relied on an ensemble of models which encompass a range of data sources and assumptions (Park et al., 2023). These parameters are traditionally estimated using case data as a proxy for the infection incidence curve (Cori, Ferguson, Fraser, & Cauchemez, 2013; Parag, 2021), but methods have also been developed to estimate these parameters from other sources, such as hospitalisations (Moore, Rosato, & Maskell, 2022) and genomic data (Vöhringer et al., 2021). The four nations of the UK (England, Scotland, Wales and Northern Ireland, herein ordered by population size) were recognised globally as having comprehensive SARS-CoV-2 testing surveillance systems (Clarke, Beaney, & Majeed, 2022; Dean, 2022; Tapper, 2022), comprising of widescale community testing (Coronavirus (COVID-19) in the UK, 2023), nationwide surveys of infection (Coronavirus (COVID-19) Infection Survey, UK, 2022; Real-Time Assessment of Community Transmission (REACT) Study, 2022), genomic data (COVID-19 Genomic Surveillance, 2023; The COVID-19 Genomics UK (COG-UK) consortium, 2020), and wastewater surveillance (Morvan et al., 2022), but these were largely scaled down from their peak capacity as part of the transition from an “emergency” to “endemic” state in the first half of 2022 (Coronavirus (COVID-19): Test and Protect - Transition Plan, 2022; COVID-19 Response: Living with COVID-19, 2022; COVID-19 Test, Trace and Protect Transition Plan, 2022; Together for a Safer Future: Wales' Long-Term Covid-19 Transition from Pandemic to Endemic, 2022). Consequently, estimation of $R(t)$ and $r(t)$, particularly using an ensemble model approach, became more challenging due to the reduction in available data streams.

The Office for National Statistics (ONS) COVID-19 Infection Survey (CIS) (Coronavirus (COVID-19) Infection Survey, UK, 2022) was a primary means by which to understand, quantify the impacts of and track SARS-CoV-2 transmission within the UK (Lythgoe et al., 2023; Pouwels et al., 2021; Rhodes et al., 2022; Shabnam et al., 2023). This COVID-19 testing study invited members of randomly selected private households across the UK to complete polymerase chain reaction (PCR) tests, regardless of symptoms or behaviour. The ONS CIS continued for one year after the cessation of community testing, until its “pause” in March 2023 (COVID-19 Infection Survey Participants Thanked for ‘Huge Contribution’ to Pandemic Response, 2023), and in October 2023 it was announced that a similar study would be undertaken for the upcoming 2023/24 winter period (UKHSA and ONS Launch New Winter COVID-19 Infection Study, 2023). Although the survey can estimate incidence of infection, this lags the main metric of interest, the percentage of the population testing positive for SARS-CoV-2 infection, which is used as a proxy for the prevalence of infection in each nation of the UK. These data, widely regarded as “gold-standard” due to the random sampling methods underpinning them, provided an opportunity to continue estimating $R(t)$ and $r(t)$ after the scaling down of many of the surveillance systems described above.

By adapting the methods deployed by another community surveillance survey with randomly-selected participants, the REal-time Assessment of Community Transmission (REACT) study (Eales, Ainslie, et al., 2022; Eales, Wang, et al., 2022; Elliott et al., 2021, 2022, Real-Time Assessment of Community Transmission (REACT) Study, 2022), this paper presents a model to estimate $R(t)$ and $r(t)$ directly from publicly-available ONS CIS estimates of test positivity in each nation of the UK. The estimates produced under this model are compared to government-published ensemble estimates, to assess the validity of this method to track the spread of SARS-CoV-2 in the absence of other surveillance data. In particular, we consider the level of contribution that this single data stream can offer in the estimation of these parameters. This methodology is then used to provide estimates of $R(t)$ and $r(t)$ in each nation of the UK until the initial pause of the ONS CIS in the first quarter of 2023.

2. Methods

2.1. Estimating $R(t)$ and $r(t)$ using the ONS COVID-19 Infection Survey (CIS)

2.1.1. The ONS CIS

The ONS CIS was the only long-term SARS-CoV-2 testing study in randomly selected households encompassing all four nations of the UK. In brief, private households were randomly selected across each nation, irrespective of factors such as members displaying symptoms or having contact with a known case, and household members aged over 2 years were invited to complete multiple PCR tests over time. The random sampling produced estimates that were unaffected by test-seeking behaviour and public availability of diagnostic tests. Further details regarding sampling can be found in (Coronavirus (COVID-19) Infection Survey: Methods and Further Information, 2023).

The primary outcome of interest was the estimated percentage of people testing positive for SARS-CoV-2, derived from the number of positive tests out of the total tests completed and post-stratified by key variables, such as gender and ethnicity. Until 29 July 2022, the raw numbers of positive and total tests were not publicly available, although they have since been reported retrospectively for the entire study period (Coronavirus (COVID-19) Infection Survey: England: Previous Versions, 2023; Coronavirus (COVID-19) Infection Survey: Northern Ireland: Previous Versions, 2023; Coronavirus (COVID-19) Infection Survey: Scotland: Previous Versions, 2023; Coronavirus (COVID-19) Infection Survey: Wales: Previous Versions, 2023). The publication

of corresponding estimates of incidence of infection, based on the primary outcome (prevalence of test positivity), was lagged by a couple of weeks, as this is substantially more complex to estimate, but ceased to be published from June 2022 ([Coronavirus \(COVID-19\) Infection Survey, UK: 15 July 2022, 2022](#)). Initially, the CIS had overlapping reporting windows of approximately 10–14 days but reporting settled into approximately weekly windows.

In this study, we focus exclusively on the publicly-available estimates of test positivity provided as point estimates with 95% credible intervals, derived from the post-stratification model ([Coronavirus \(COVID-19\) Infection Survey: Methods and Further Information, 2023](#)), due to these being consistently available to the public in real-time for the majority of the ONS CIS study period. Furthermore, these estimates are unaffected by issues concerning deductive disclosure, in which the number of positive tests cannot be disclosed due to their small number, as experienced by both Wales and Northern Ireland at the beginning of the study in 2020.

The survey commenced on different dates in each nation ([Table 1](#)) and there is some heterogeneity in the reporting windows across nations, but was “paused” on 13 March 2023 in England, Scotland and Wales and on 7 March 2023 in Northern Ireland (with the last publication (at time of writing) on 24 March 2023) ([Coronavirus \(COVID-19\) Infection Survey, UK: 24 March 2023, 2023](#)). In all settings, the midpoint of the reporting window is taken as the “date” of the observation, for the purposes of model fitting.

In October 2023, the UKHSA and the ONS announced a “Winter COVID-19 study” (WCIS) running from November 2023–March 2024, with similar aims to the ONS CIS of ascertaining prevalence of SARS-CoV-2 infection in communities across the UK ([UKHSA and ONS Launch New Winter COVID-19 Infection Study, 2023](#)). However, as these data are not available at the time of writing (October 2023) and their exact format is currently unknown, our study only covers the first, continuous period of the ONS CIS from 2020 to the end of March 2023. Nonetheless, the implications of our methodology for the newly announced WCIS are outlined in the Discussion.

2.1.2. Modelling $R(t)$ and $r(t)$

A two-step approach is used to estimate $R(t)$ and $r(t)$ from the ONS CIS, with these estimates denoted by $\hat{R}(t)$ and $\hat{r}(t)$, respectively, and referred to as “ONS-based” estimates. The methods are primarily adapted from those presented in Eales et al. ([Eales, Ainslie, et al., 2022](#)) and akin to those in ([Vöhringer et al., 2021](#); [Wallinga & Lipsitch, 2007](#); [Ward et al., 2021](#)). Specifically, a penalised spline model is fit to the publicly-available ONS CIS test positivity estimates to produce a smoothed estimate of the prevalence of SARS-CoV-2 positivity over time before the epidemiological parameters of interest are estimated directly from the resulting curve. Each step is described in turn below.

2.1.3. Step 1: fitting a spline to the ONS CIS test positivity data

The model used by the REACT study ([Eales, Ainslie, et al., 2022](#)) was taken as the basis for spline-fitting in this study and is explained here briefly. The REACT-1 data consist of the number of positive (Y_t) and total (N_t) tests on each day t of the study period, allowing test positivity of SARS-CoV-2 to be naively estimated as $p_t = \frac{Y_t}{N_t}$. Smoothed test positivity, defined as $\hat{p}(t)$, is estimated via a linear combination of K B-splines up to the 2nd degree, defined by a sequence of equidistant knots distributed throughout the period under consideration. Specifically:

$$\hat{p}(t) = \sum_{k=1}^K a_k B_{k,3}(t)$$

where $B_{k,3}(t)$ represents the k th 3rd order B-spline (see Supplementary Material) and a_k are the model coefficients, defined by a second-order random-walk prior distribution. The binomial likelihood is then parameterised by the observed number of total tests, N_t , and the estimated test positivity, $\hat{p}(t)$, at each time point.

Although the principal idea is the same, several adjustments were required to adapt the REACT spline model to ONS test positivity estimates. First, an additional sampling step was required to capture the uncertainty in the ONS test positivity estimates, which are provided in a different format to the REACT-1 data. Let μ_t denote the central ONS estimate of SARS-CoV-2

Table 1

Availability of estimates used in this study: test positivity from the ONS CIS survey and official estimates of the effective reproduction number and growth as published by the UK government. The study period for each nation is determined as the beginning date of the ONS CIS until the initial “pause” date of the ONS CIS (March 2023), while the period of overlap states the period for which both ONS-based estimates and government-published estimates are available for comparison.

Nation	ONS CIS coverage	Parameter	Government-published estimate coverage	Period of overlap
England	27/04/20–13/03/23	Reproduction number	29/05/20–23/12/22	29/05/20–23/12/22
		Growth rate	12/06/20–23/12/22	12/06/20–23/12/22
Scotland	03/10/20–13/03/23	Reproduction number	28/05/20–08/12/22	03/10/20–08/12/22
		Growth rate	18/06/20–08/12/22	03/10/20–08/12/22
Wales	27/07/20–13/03/23	Reproduction number	28/05/20–09/12/22	27/07/20–09/12/22
		Growth rate	12/06/20–27/12/21	27/07/20–27/12/21
Northern Ireland	11/09/20–07/03/23	Reproduction number	26/05/20–31/05/22	11/09/20–31/05/22
		Growth rate	09/06/20–17/11/20	11/09/20–17/11/20

test positivity at time t , with l_t and u_t denoting the corresponding lower and upper limits of the credible interval, respectively. The variance of the estimate can be estimated as $\sigma_t^2 \approx \left(\frac{u_t - l_t}{3.92}\right)^2$. For each time t , π_t , the test positivity, is considered a random variable following a Beta distribution parameterised using μ_t and σ_t^2 as shown:

$$\begin{aligned}\pi_t &\sim \text{Beta}(\alpha_t, \beta_t) \\ \alpha_t &= \mu_t \left(\frac{\mu_t(1 - \mu_t)}{\sigma_t^2} - 1 \right) > 0 \\ \beta_t &= \alpha_t \left(\frac{1 - \mu_t}{\mu_t} \right) > 0.\end{aligned}$$

The derivation of the expressions for α_t and β_t is shown in the Supplementary Material. For every iteration, i , our algorithm begins by sampling $\pi_t^{(i)}$ for every time point t ahead of the construction of the B-splines. Essentially, $\pi_t^{(i)}$ is then taken as the “data point” at time t being fitted to in iteration i ; without this additional parametric bootstrapping sampling we would fit exclusively to the point estimate and discard the information about uncertainty contained within the 95% credible intervals. The sampled value $\pi_t^{(i)}$ is then transformed onto the unconstrained logit scale:

$$\text{logit}(\pi_t) = \log\left(\frac{\pi_t}{1 - \pi_t}\right).$$

Second, the weekly timescale of the ONS CIS data, compared to the daily timescale of the REACT-1 study data, also necessitated adjustments to the model. In this instance, a first-order random walk was deemed a more appropriate prior distribution for the model coefficients due to the rapidly changing epidemic dynamics in these weekly data. Furthermore, knots could not be placed at the REACT-1-chosen value of every five days. Rather, a sensitivity analysis of the placing of knots was undertaken by considering scenarios in which the number of equidistant knots was equal to a percentage of the total data points, specifically 20%, 30% and 40%, balancing capturing sufficient information but without overfitting.

The final adaptation was to use a Normal likelihood function, again due to the different format of the data, based on the results of a simulation study, with details set out in the Supplementary Material. Bringing everything together:

$$\begin{aligned}\pi_t &\sim \text{Beta}(\alpha_t, \beta_t) \\ a_1 &\sim \text{Normal}(0, 1) \\ a_k &\sim \text{Normal}(a_{k-1}, \tau^2) \\ \tau^2 &\sim \text{InverseGamma}(0.0001, 0.0001) \\ \text{logit}(\hat{p}(t)) &= \sum_{k=1}^K a_k B_{k,3}(t) \\ \text{logit}(\pi_t) &\sim \text{Normal}(\text{logit}(\hat{p}(t)), \gamma^2) \\ \gamma^2 &\sim \text{InverseGamma}(0.0001, 0.0001)\end{aligned}$$

This model was fitted using a No-U-Turn sampler (Homan & Gelman, 2014), implemented in STAN, with 4 chains, each with 20,000 iterations and a burn-in of 2000 iterations. To ensure approximately-continuous estimates of $R(t)$ and $r(t)$, the coefficients from the model are used to estimate $\hat{p}(t)$ at a granular time step (hundredths of one day).

2.1.4. Step 2a: Estimating $R(t)$

The effective reproduction number, $R(t)$, is estimated using the renewal equation (Cori et al., 2013; Wallinga & Teunis, 2004):

$$\hat{R}(t) = \frac{\hat{p}(t)}{\int_{\tau=0}^{\infty} \hat{p}(t - \tau) g(\tau) d\tau}$$

where $g(\cdot)$ denotes the distribution of the generation time, defined as the time between the infections of infector-infectee pairs. The integral is approximated by summation.

The generation time distribution has been shown to change over time with the emergence of new variants (Hart, Miller, et al., 2022). As such, estimates of $g(\cdot)$ used in this study were extracted from the literature for four distinct time periods defined by epidemic waves: Wildtype, Alpha, Delta and Omicron. Each study distribution was parameterised using data from the UK and thus is assumed to be applicable to this setting. The transition dates between time periods were determined by the earliest date at which 50 % of the daily tests were attributable to the emerging variant according to Our World in Data (SARS-CoV-2 Variants in Analyzed Sequences, United Kingdom, 2023). The distributions are summarised in Table 2.

2.1.5. Step 2b: Estimating $r(t)$

The instantaneous growth rate, $r(t)$, is approximated as follows:

$$\hat{r}(t) = \frac{\frac{\partial \hat{p}(t)}{\partial t}}{\hat{p}(t)}.$$

The quantities of interest, $\hat{R}(t)$ and $\hat{r}(t)$, are estimated for all posterior realisations of the spline model, with results presented as the median and the 2.5% and 97.5% quantiles throughout.

2.2. Analysis of $\hat{R}(t)$ and $\hat{r}(t)$

2.2.1. Government-published estimates

Estimates of $R(t)$ and $r(t)$ for SARS-CoV-2 in each of the four nations were produced for and published by the government from early in the pandemic for either weekly or biweekly time periods (Coronavirus (COVID-19) in the UK, 2023; Park et al., 2023). The estimates were derived from an ensemble of (up to 14) independently run models as part of a cross-government and academic modelling hub that comprised the UK Health Security Agency (UKHSA) Epidemiological Ensemble team and Scientific Pandemic Influenza Group on Modelling, Operational sub-group (SPI-M-O) (Park et al., 2023; Reproduction Number (R) and Growth Rate: Methodology, 2021). These models encompassed a range of different assumptions and data streams as set out in (Park et al., 2023). Among the data streams used, officially reported cases and hospital admissions were the most common, but there were also two models which fit to the ONS CIS data (Abbott & Funk, 2022; Birrell, Blake, van Leeuwen, Gent, & De Angelis, 2021, 2022). Consequently, there is a degree of inbuilt dependency between the government-published estimates and the estimates obtained in this study, as demonstrated below.

Ensemble model outputs were combined into a single estimate with associated uncertainty, using an established method derived by the Defence Science and Technology Laboratory (DSTL) (Maishman et al., 2022). These estimates are publicly available for download and are presented as 90 % confidence intervals for the period, without any central estimate (Coronavirus (COVID-19) in the UK, 2023). We apply the weekly or biweekly estimates to each day in the corresponding time period, and approximate a central estimate by taking the mid-point of the upper and lower bounds, herein denoted by $\hat{R}(t)$ and $\hat{r}(t)$ for the effective reproduction number and instantaneous growth rate, respectively. These estimates are herein referred to as “government-published” estimates.

There is heterogeneity in the availability of the government-published estimates, depending on the parameter and nation under consideration (Table 2). For example, official estimates of the growth rate in Northern Ireland are only available from June to November 2020. All estimates for each nation ceased to be published by 23 December 2022 (The R value and growth rate, 2022).

2.2.2. Comparison of ONS-based and government-published estimates

The ONS-based estimates $\hat{R}(t)$ and $\hat{r}(t)$ are initially presented from the beginning of the survey period in each nation until the end of 2022, as this is when the government-published estimates ceased to be publicly available for both parameters and for all four nations (Table 1). The periods for which the ONS-based estimates can be compared to government-published estimates for each parameter and nation combination are set out in Table 1. Due to the short period of government published estimates for the growth rate in Northern Ireland, ONS-based estimates of this parameter cannot be compared in any meaningful way to the corresponding official estimates and are thus omitted from our analysis.

Table 2

Parameterisations of the generation time distribution used in the analysis. Variant time periods are defined by the earliest date at which 50 % of the daily tests were attributable to the emerging variant according to Our World in Data.(SARS-CoV-2 Variants in Analyzed Sequences, United Kingdom, 2023) Uncertainty was not characterised for the estimates obtained for the Omicron wave.

Variant	Time period	Parametric distribution	Mean	Standard deviation	Source
Wildtype	01/04/2020–25/12/2020	Lognormal	4.2 (3.3–5.3)	4.9 (3.0–8.3)	(Hart, Abbott, et al., 2022)
Alpha	26/12/2020–21/05/2021	Gamma	5.5 (4.7–6.5)	4.0 (3.3–5.3)	(Hart, Miller, et al., 2022)
Delta	22/05/2021–21/11/2021	Gamma	4.7 (4.1–5.6)	3.3 (3.0–4.3)	(Hart, Miller, et al., 2022)
Omicron	22/12/2021–13/03/2023	Gamma	3.3	3.5	(Abbott, Sherratt, Gerstung, & Funk, 2022)

We use three metrics to consider the relationship between the (median) ONS-based $[\hat{R}(t)]$ and $[\hat{r}(t)]$ and government-published $[\tilde{R}(t)]$ and $[\tilde{r}(t)]$ estimates of the parameters of interest, each of which are now presented. These are herein referred to as the “comparison metrics”.

First, we use linear regression models of the form:

$$\tilde{R}(t) = \beta_0 + \beta_1 \hat{R}(t) + \epsilon$$

and

$$\tilde{r}(t) = \gamma_0 + \gamma_1 \hat{r}(t) + \epsilon$$

to assess the level of variation within the government-published estimates explained by the ONS-based estimates. This is captured by:

$$R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{Total sum of squares}}.$$

Second, we consider the Spearman rank correlation, selected due to the lack of assumptions regarding the distributions of the data.

Finally, the parameters of interest in this paper are sometimes interpreted dichotomously by policymakers, for example, to assess whether the epidemic is either growing or shrinking. As such, the proportion of point estimates over the study period for which the modelled and official estimates are on the same side of the “growth thresholds”, 0 for the growth rate and 1 for the effective reproduction number, over time is also presented as the third and final metric for comparison. This quantity is herein referred to as the “agreement proportion”.

A sensitivity analysis was used to examine the potential of a time-lag between estimates, assessed via the three comparison metrics. Specifically, lags between plus and minus 20 days from the default values were considered. A positive lag of L days indicates that the ONS-based estimates of day D , $\hat{R}(D)$ and $\hat{r}(D)$, are compared with the government-published estimates at day $D + L$, $\tilde{R}(D + L)$ and $\tilde{r}(D + L)$, and vice versa. In the main plots, we present our ONS-based estimates alongside the government-published estimates for the time lag and knot value under which R^2 is maximised. A comparison of estimates without any time lags are presented in the Supplementary Material.

2.3. $\hat{R}(t)$ and $\hat{r}(t)$ in the first quarter of 2023

Finally, $\hat{R}(t)$ and $\hat{r}(t)$ are produced for the first quarter of 2023 by fitting the spline model to ONS CIS estimates from 1 November 2022 until the initial “pause” of the survey on 13 March 2023 in England, Scotland and Wales and on 7 March 2023 in Northern Ireland. These are presented for information but without comparison, due to the lack of availability of the government-published estimates in this period.

All data and code used are available from: <https://github.com/ruthmccabe/ons-test-positivity-model>.

3. Results

Figs. 1–4 present the results for England, Scotland, Wales and Northern Ireland, respectively. Across the board, we can demonstrate strong agreement between our ONS-based estimates and the government-published estimates for both parameters. Depending on nation and parameter, up to 77% of the variance in the government-published estimates can be explained by the ONS-based estimates, providing evidence of the suitability of this singular data stream to track the epidemic in each of the four nations (Table 3). Similarly, we observed high maximum values of the Spearman rank correlation, ranging between 0.72 and 0.87 (Table 3).

We found that the model in England produced similar values of the three comparison metrics for all numbers of knots considered (Fig. 1D–F). However, this was not the case in Scotland and Wales, where the number of knots played a more important role (Fig. 2D–F; Fig. 3D–F): for example, the maximum observed R^2 for $R(t)$ in Scotland rose by more than 40% from 0.53 to 0.74 when doubling the number of knots from 20% to 40% of total data points. In addition to requiring a greater number of knots, the spline fits in Scotland and Wales, and additionally in Northern Ireland, all have substantially greater uncertainty which is propagated through to $\tilde{R}(t)$ and $\tilde{r}(t)$. This is likely driven by the increased uncertainty arising from smaller sample sizes in the ONS CIS, which are proportional to the smaller population sizes in these nations compared to England.

Our sensitivity analysis has highlighted a time delay between the ONS-based estimates compared to the government-based estimates. In England, Scotland and Wales, the ONS-based estimates are correlated with later government-published estimates, suggesting that the ONS-based estimates can capture epidemic trends more quickly. (This effect can be seen clearly in Supplementary Figs. 11–13). The models in England and Scotland indicated a similar time delay of around 8 days for $R(t)$ and 10 days for $r(t)$. In Scotland, R^2 rose by almost 50% from 0.49 with no time delay to its maximum value (0.74)

under a delay of 8 days for the $R(t)$ under the model with 40% knots, emphasising the importance of considering such delays. The time delay which maximised the metrics of interest was much greater in Wales, sitting at around 2 weeks for both parameters.

For $R(t)$ and $r(t)$ in England and Scotland, and $R(t)$ in Wales, the agreement proportion sits close to 0.80 depending on the time delay and number of knots used (range 0.79–0.82). The majority of instances in which there is not agreement on whether the epidemic is growing or shrinking occurs for values close to the threshold (e.g. one estimate being slightly over the threshold while the other is slightly under and vice versa), rather than there being large differences between the two estimates. The is what drives the low agreement proportion of $r(t)$ in Wales, despite the corresponding R^2 and Spearman correlation values being relatively high.

Our ONS-based estimates for Northern Ireland have the weakest relationship with the government-published estimates. $\hat{R}(t)$ fluctuates substantially more so than $\hat{r}(t)$ throughout the period, in particular from November 2021 (around the time of the emergence of the Omicron variant). This relationship is reflected in all three comparison metrics, with a maximum of only 50% of the variance in the government-published estimates being explained by the ONS-based estimates. Furthermore, the ONS-based estimates are (weakly) correlated with an earlier government-based, in contrast to England, Scotland and Wales, meaning that in Northern Ireland the ONS-based estimates are slower to track the epidemic trends in this setting.

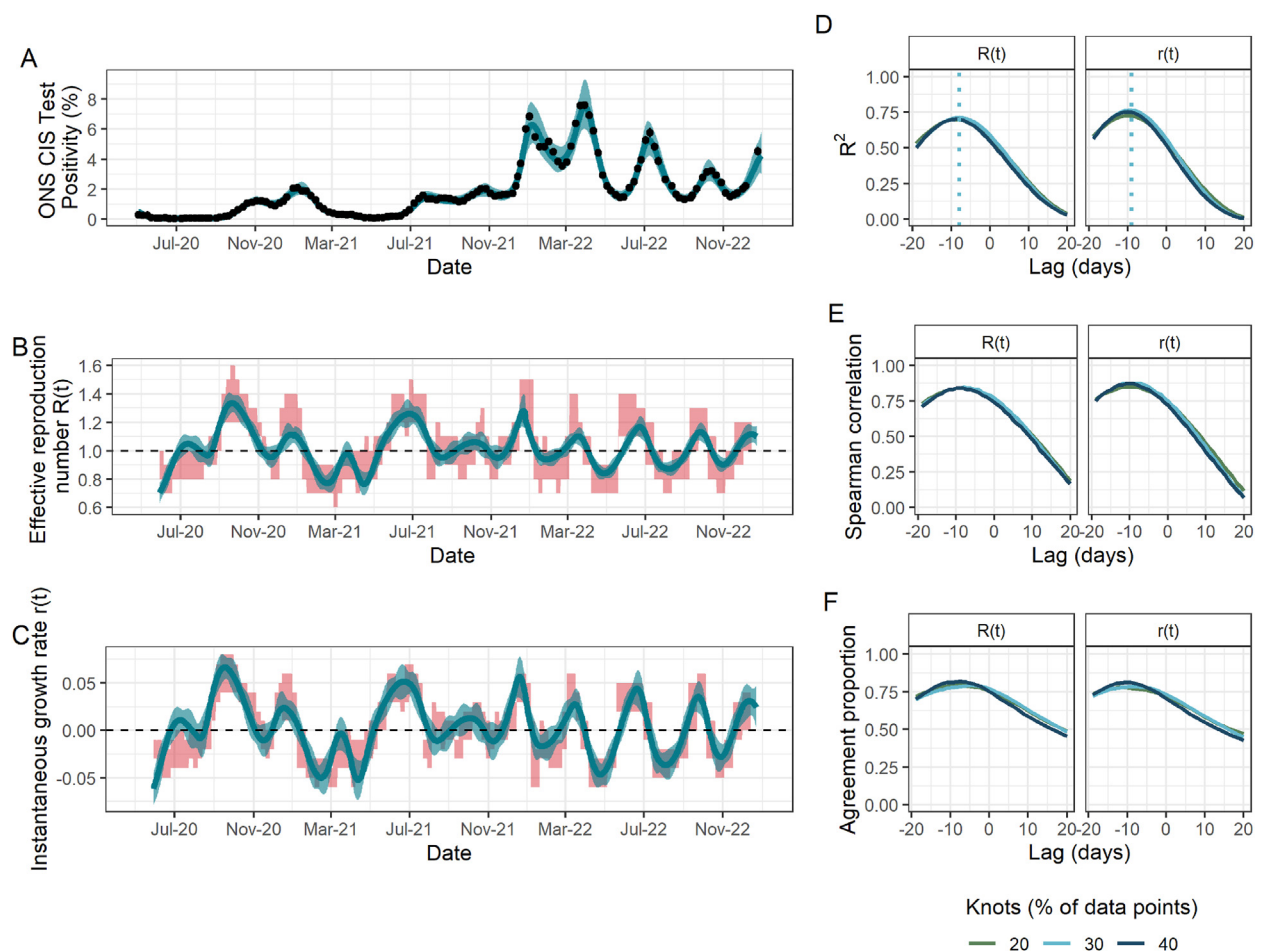


Fig. 1. Fit to ONS CIS data, resulting estimates of $R(t)$ and $r(t)$, and comparison metrics for England. (A) Spline model fit ($\hat{p}(t)$) (blue; median line with 95% credible intervals) to data (black points; ONS point estimate with 95% credible intervals) with the number of knots totalling 30% of the total data points. (B) The ONS-based estimated effective reproduction number ($\hat{R}(t)$) (blue; median line with 95% credible intervals) alongside the government-published estimates ($\bar{R}(t)$) (red; 90% confidence intervals), lagged by -8 days. The black dashed line highlights the epidemic growth threshold of 1. (C) The ONS-based estimated instantaneous growth rate ($\hat{r}(t)$) (blue; median line with 95% credible intervals) alongside the government-published estimates ($\bar{r}(t)$) (red; 90% confidence intervals), lagged by -9 days. The dashed line highlights the epidemic growth threshold of 0. (D)–(F) Results of the three metrics used to assess the agreement between ONS-based and government-published estimates under different lags and for models fitted with a different number of knots, taken as a percentage of the total data points. (D) R^2 . Dotted lines indicate the lag for which this metric is maximised for each parameter and is thus what is applied to the government-based estimates in panels (B) and (C). (E) Spearman rank correlation. (F) Agreement proportion.

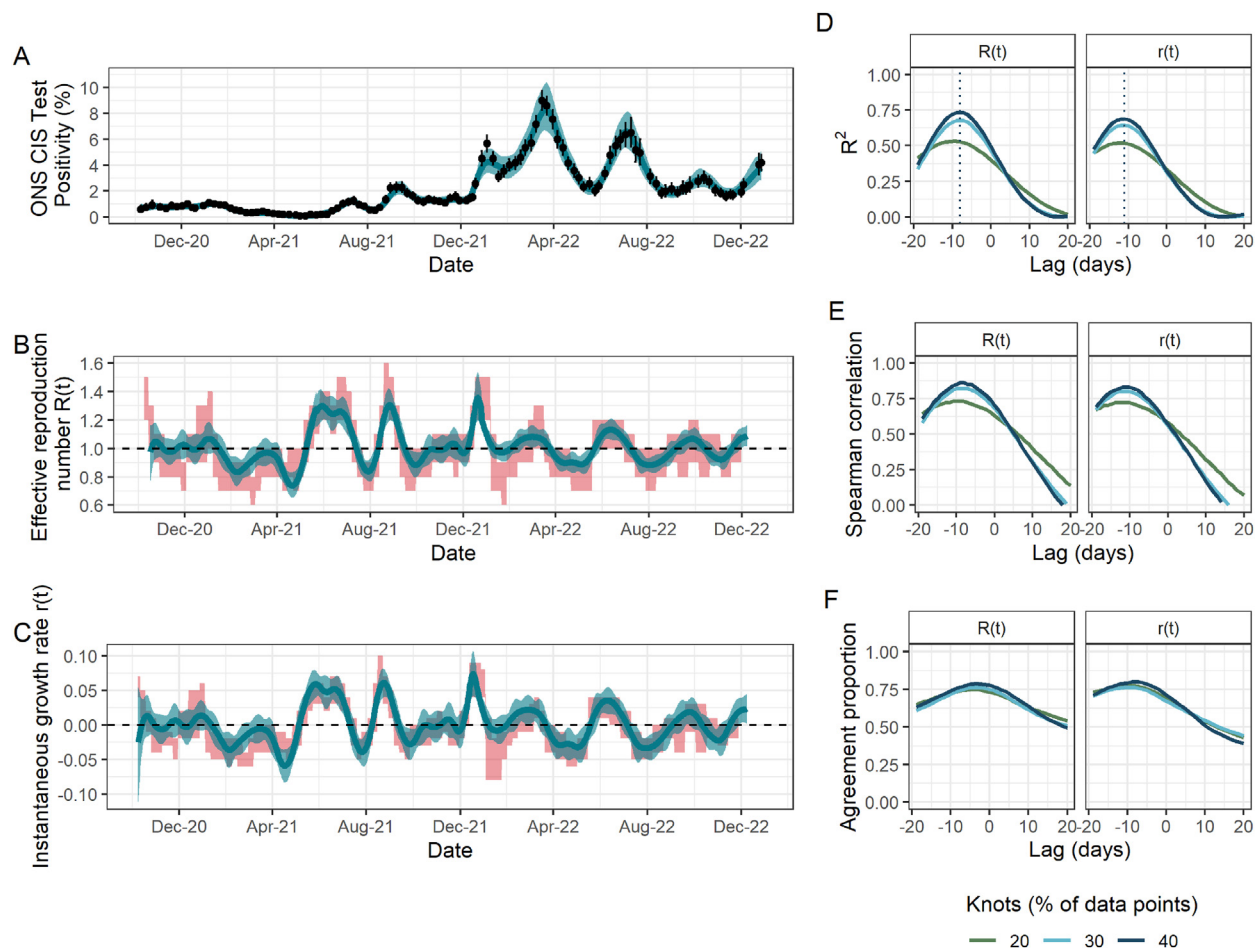


Fig. 2. Fit to ONS CIS data, resulting estimates of $R(t)$ and $r(t)$, and comparison metrics for Scotland. (A) Spline model fit ($\hat{p}(t)$) (blue; median line with 95% credible intervals) to data (black points; ONS point estimate with 95% credible intervals) with the number of knots totalling 40% of the total data points. (B) The ONS-based estimated effective reproduction number ($\hat{R}(t)$) (blue; median line with 95% credible intervals) alongside the government-published estimates ($\bar{R}(t)$) (red; 90% confidence intervals), lagged by -8 days. The black dashed line highlights the epidemic growth threshold of 1. (C) The ONS-based estimated instantaneous growth rate ($\hat{r}(t)$) (blue; median line with 95% credible intervals) alongside the government-published estimates ($\bar{r}(t)$) (red; 90% confidence intervals), lagged by -11 days. The dashed line highlights the epidemic growth threshold of 0. (D)–(F) Results of the three metrics used to assess the agreement between ONS-based and government-published estimates under different lags and for models fitted with a different number of knots, taken as a percentage of the total data points. (D) R^2 . Dotted lines indicate the lag for which this metric is maximised for each parameter and is thus what is applied to the government-based estimates in panels (B) and (C). (E) Spearman rank correlation. (F) Agreement proportion.

3.1. Estimates for 2023

Fig. 5 presents the spline fits and estimates of $\hat{R}(t)$ and $\hat{r}(t)$ from January–March 2023 in each nation, under varying numbers of knots in the spline model.

In each nation, ONS test positivity decreases to mid-January, which results in the corresponding estimates falling below the epidemic growth thresholds: $\hat{R}(t) < 1$ and $\hat{r}(t) < 0$. However, the degree to which the estimates indicate a shrinking epidemic is dependent on the number of knots used, with a lower number of knots resulting in the estimates being closer to the threshold and vice versa.

From mid-January, test positivity increases in all nations, peaking in mid-February in England, Scotland and Wales, and in early February in Northern Ireland. This results in epidemic growth, $\hat{R}(t) > 1$ and $\hat{r}(t) > 0$, for most of the month of February. In England and Scotland, by the end of the period, both $\hat{R}(t)$ and $\hat{r}(t)$ approach the epidemic threshold showing a stable epidemic, but in Wales and Northern Ireland the estimates imply that the epidemic was continuing to grow.

4. Discussion

We have demonstrated and validated a method to estimate the $R(t)$ and $r(t)$ of SARS-CoV-2 using the publicly-available ONS CIS data, which became a primary source characterising the ongoing epidemics in the four nations of the UK after the

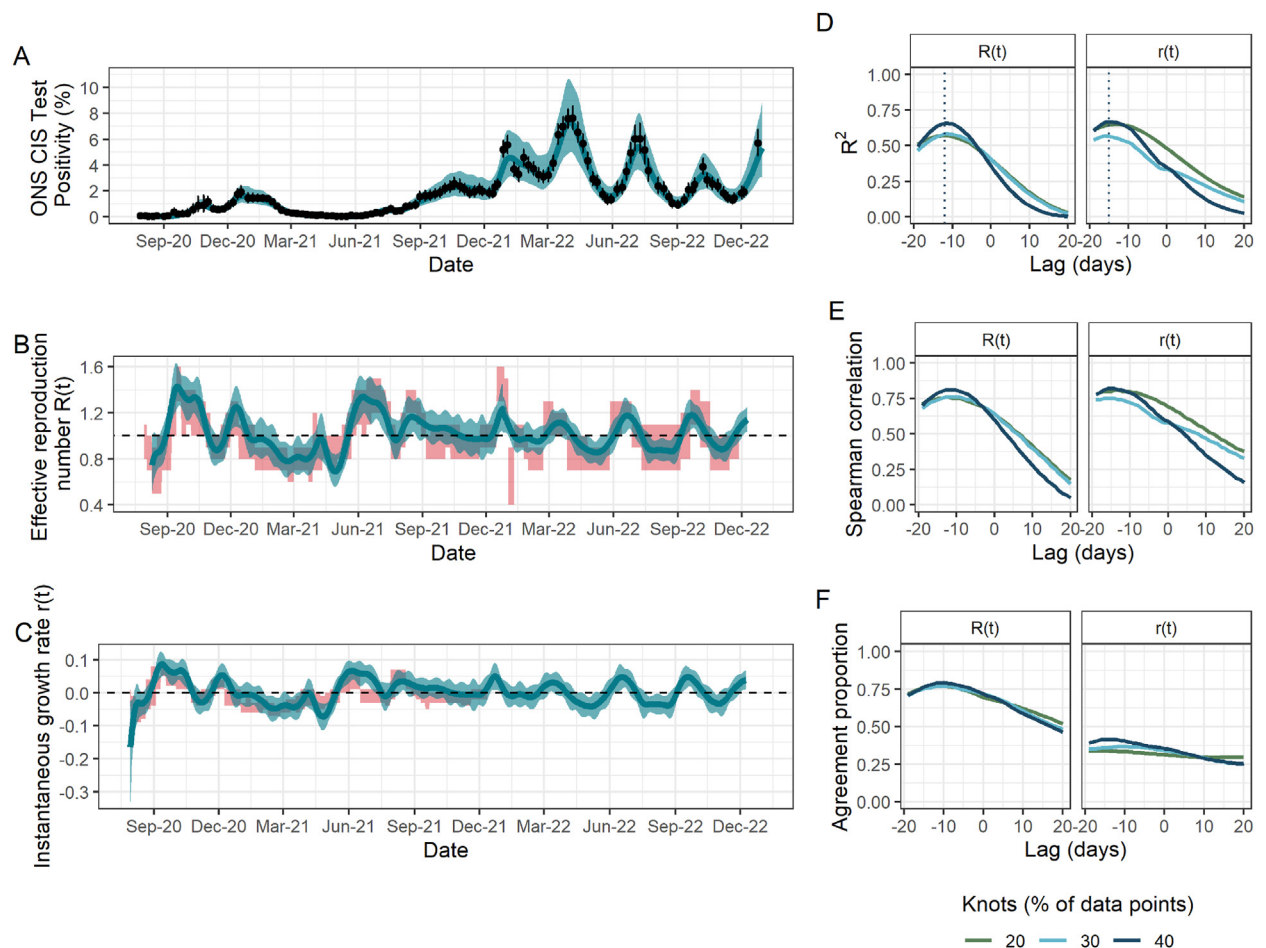


Fig. 3. Fit to ONS CIS data, resulting estimates of $R(t)$ and $r(t)$, and comparison metrics for Wales. (A) Spline model fit ($\hat{p}(t)$) (blue; median line with 95% credible intervals) to data (black points; ONS point estimate with 95% credible intervals) with the number of knots totalling 40% of the total data points. (B) The ONS-based estimated effective reproduction number ($\hat{R}(t)$) (blue; median line with 95% credible intervals) alongside the government-published estimates ($\bar{R}(t)$) (red; 90% confidence intervals), lagged by -12 days. The black dashed line highlights the epidemic growth threshold of 1. (C) The ONS-based estimated instantaneous growth rate ($\hat{r}(t)$) (blue; median line with 95% credible intervals) alongside the government-published estimates ($\bar{r}(t)$) (red; 90% confidence intervals), lagged by -15 days. The dashed line highlights the epidemic growth threshold of 0. (D)–(F) Results of the three metrics used to assess the agreement between ONS-based and government-published estimates under different lags and for models fitted with a different number of knots, taken as a percentage of the total data points. (D) R^2 . Dotted lines indicate the lag for which this metric is maximised for each parameter and is thus what is applied to the government-based estimates in panels (B) and (C). (E) Spearman rank correlation. (F) Agreement proportion.

scaling down of community testing in Spring 2022, until the survey's initial “pause” in March 2023. We have shown strong agreement between our ONS-based and government-published estimates across mid-2020 until the end of 2022, showing the suitability of this model applied to these data to track the trends in these key epidemic parameters. Specifically, we demonstrated that up to 77% of the variation in the government-published estimates could be explained by our ONS-based estimates, depending on the nation and parameter under consideration, which was complemented by high values of the Spearman rank correlation and agreement proportion. We have also found that most estimates under this model, except for Northern Ireland, led government-reported estimates by up to 2 weeks, a potentially advantageous gap in terms of producing timely real-time modelling estimates. These results are important for the WCIS announced in October 2023 ([UKHSA and ONS Launch New Winter COVID-19 Infection Study, 2023](#)), as the methodology deployed here could potentially be used on the data generated by this study to track community transmission in the UK over winter 2023/24. Furthermore, our study demonstrates the epidemiological value of population-level studies using random sampling, such as the ONS CIS and REACT, in terms of estimating key epidemiological parameters in addition to, for example, examining risk factors for severe infection ([Pouwels et al., 2021](#); [Rhodes et al., 2022](#)), the burden of long COVID-19 ([Shabnam et al., 2023](#)) and the evolutionary dynamics of SARS-CoV-2 variants over time ([Lythgoe et al., 2023](#)).

Our work is not intended as an alternative to ensemble modelling, which is advantageous in its ability to synthesise information from multiple sources ([Barai & Reich, 1999](#)). Although deployed in multiple different fields ([Di Napoli et al., 2020](#); [Grenouillet, Buisson, Casajus, & Lek, 2011](#); [Zhou, Lai, & Yu, 2010](#)), this is particularly important in the context of modelling the

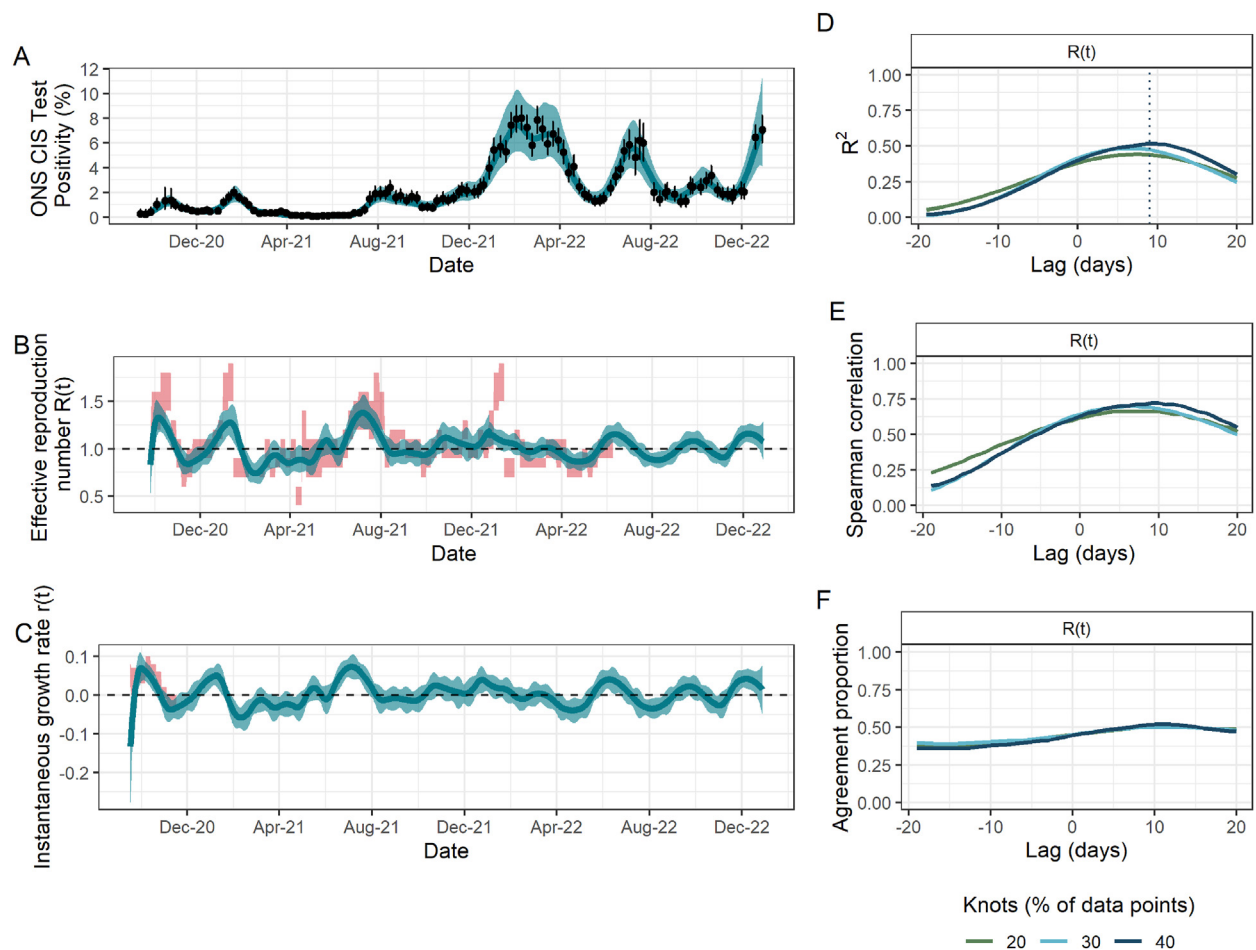


Fig. 4. Fit to ONS CIS data, resulting estimates of $R(t)$ and $r(t)$, and comparison metrics for Northern Ireland. (A) Spline model fit ($\hat{p}(t)$) (blue; median line with 95% credible intervals) to data (black points; ONS point estimate with 95% credible intervals) with the number of knots totalling 40% of the total data points. (B) The ONS-based estimated effective reproduction number ($\hat{R}(t)$) (blue; median line with 95% credible intervals) alongside the government-published estimates ($\bar{R}(t)$) (red; 90% confidence intervals), lagged by 9 days. The black dashed line highlights the epidemic growth threshold of 1. (C) The ONS-based estimated instantaneous growth rate ($\hat{r}(t)$) (blue; median line with 95% credible intervals) alongside the government-published estimates ($\bar{r}(t)$) (red; 90% confidence intervals), without any lag. The dashed line highlights the epidemic growth threshold of 0. (D)–(F) Results of the three metrics used to assess the agreement between ONS-based and government-published estimates of $R(t)$ under different lags and for models fitted with a different number of knots, taken as a percentage of the total data points. (D) R^2 . Dotted lines indicate the lag for which this metric is maximised for $R(t)$ and is thus what is applied to the government-based estimates in panels (B) and (C). (E) Spearman rank correlation. (F) Agreement proportion.

SARS-CoV-2 epidemic in the UK, due to the diversity of surveillance data streams available. As discussed by Park et al. (Park et al., 2023), the ensemble modelling approach has many strengths including increased prediction ability and greater robustness. We see our work completing this approach, by presenting a potential solution to the challenge faced by public health officials in the UK in early 2022 given the large scaling-down of the surveillance systems at the time. With less data available, some ensemble models could become less reliable, while the approach presented here would not have been affected given the continuation of the ONS CIS beyond this period.

One of the strengths of this study is its application to all nations of the UK separately, which is uncommon in the literature (Birrell et al., 2021; Danon, Brooks-Pollock, Bailey, & Keeling, 2021; Davies, Kucharski, Eggo, Gimma, & Edmunds, 2020; Keeling et al., 2022; Simpson et al., 2021). This work underlines the importance of setting-specific analysis of an epidemic and the differences that can arise even when these settings are geographically close and socioeconomically similar. While we were able to produce estimates which matched government-reported estimates closely, we found that both the smoothness of the spline and the time delay between the ONS-based and government-published estimates were dependent on the nation of the UK, with smaller ONS CIS sample sizes (proportional to population size) resulting in less smooth fits. As previously mentioned, this is likely attributable to the increased variability observed in the ONS CIS data among nations with smaller sample sizes. Moreover, each of the four nations have devolved governments which implemented slightly different public health and social measures at varying time points. This nuance would be difficult to capture accurately in a UK-level model.

Table 3

Maximum values of the three comparison metrics used to assess the relationship between the ONS-based and government-published estimates of $R(t)$ and $r(t)$ for each nation. The value of the metric is provided alongside the lag* (in days) and the percentage of knots** out of the total data points for which this maximum value is produced. For any instance in which there were multiple combinations producing the same maximum value, the lag closest to 0 and lowest percentage of knots was selected for presentation here.

		Effective reproduction number			Instantaneous growth rate		
		R^2	Spearman correlation	Agreement proportion	R^2	Spearman correlation	Agreement proportion
England	Value	0.71	0.85	0.81	0.77	0.87	0.82
	Lag*	–8	–6	–10	–9	–7	–7
	Knots**	30%	30%	40%	30%	30%	40%
Scotland	Value	0.74	0.86	0.79	0.69	0.83	0.80
	Lag*	–8	–8	–4	–11	–10	–7
	Knots**	40%	40%	40%	40%	40%	40%
Wales	Value	0.66	0.81	0.79	0.67	0.82	0.42
	Lag*	–12	–10	–10	–15	–14	–13
	Knots**	40%	40%	40%	40%	40%	40%
Northern Ireland	Value	0.52	0.72	0.52			
	Lag*	9	9	10			
	Knots**	40%	40%	40%			

The importance of setting-specific modelling was underlined in Northern Ireland. We found that, here, $\hat{R}(t)$ has the weakest relationship with $\tilde{R}(t)$ of all of the nations, and this relationship could not even be evaluated for $r(t)$ due to the lack of government-published estimates in this nation. In addition to the weaker relationship, the temporal relationship between the two sets of estimates were the opposite of that observed in the other three nations, with $\hat{R}(t)$ trailing $\tilde{R}(t)$. Northern Ireland has the smallest population size of the four nations of the UK (approximately 1.9 million ([Northern Ireland Population Mid-Year Estimate, 2022](#))), which inevitably increases the volatility of all surveillance data observed in this setting, and indeed this was demonstrated through the volatility of $\tilde{R}(t)$. In particular, this volatility may have contributed to the lack of published estimates $\tilde{r}(t)$, occurring when a consensus estimate from the ensemble models could not be generated due to the small number of reliable model outputs available.

Our model builds upon the methodology set out in ([Eales, Ainslie, et al., 2022](#)) by adapting the model to fit to the publicly-available ONS CIS estimates of test positivity. Of course, other methods also exist to estimate $R(t)$ and $r(t)$ using the ONS CIS outputs, for example, by fitting the model of Eales et al. ([Eales, Ainslie, et al., 2022](#)) to the (retrospectively-reported) numbers of positive and total tests. While theoretically it would be possible to use ONS-estimated incidence of infection within previously published frameworks such as ([Cori et al., 2013](#); [Parag, 2021](#); [Scott et al., 2021](#)), these figures were published up to 3 weeks later than the corresponding test positivity estimates and ceased to be published in June 2022, thus making these data unsuitable for ongoing real-time modelling of the epidemic. As an alternative, Abbott and Funk ([Abbott & Funk, 2022](#)) deconvolve ONS test positivity estimates into incidence, which they then model using a Gaussian process. However, in addition to the assumption of the generation time distribution, which is a necessity of the renewal model, the probability density function of the time from infection until PCR positivity is also required. Similarly, the ONS CIS is among the data streams that Birrell et al. ([Birrell et al., 2021, 2022](#)) fit to as part of their age- and NHS-England-region-stratified Susceptible-Exposed-Infected-Recovered (SEIR) model. However, this model is specific to England, and requires additional parameters to calibrate the complex mechanistic model to the observed data, as is the case with mathematical models of this structure. By contrast, our method only has two key parameters which require tuning to the setting of interest: firstly, the generation time distribution, which is a common assumption in models with an element of mechanistic transmission ([Cori et al., 2013](#); [Parag, 2021](#); [Scott et al., 2021](#)) and secondly, the smoothness of the spline, controlled by the number of knots. For the latter, we conducted a sensitivity analysis to assess the impact of the number of knots and show that there are often multiple viable options, although we found that data sets with smaller sample sizes were better fitted by models with more knots. Another important strength of our model is that we presented a method to fit directly to the publicly-available estimates of test positivity, which was the format in which the survey results were presented for the vast majority of the study period (until 29 July 2022). Furthermore, by fitting exclusively to these publicly-available estimates, anyone may reproduce our results and/or adapt the model to similar data in other contexts, for example in different countries or for other pathogens, or even to other data streams presented in this format if appropriate. Furthermore, the generation time distribution can be updated easily over time in our model, as was required due to the emergence of new variants ([Hart, Miller, et al., 2022](#)).

It is important to be aware of the limitations of this work. First, our estimates suffer from boundary effects attributable to the ONS CIS survey starting mid-way through a wave of infection. This is particularly noticeable in Wales and Northern Ireland (also with smaller sample sizes) and may contribute to the lower levels of agreement observed between the ONS-based and government-published estimates for each comparison metric. Second, as discussed, there is not a standard way to select the optimal level of smoothing and thus was determined in this instance via sensitivity analysis. Eales et al. ([Eales, Ainslie, et al., 2022](#)) have taken a similar approach in their selection the number of knots. Third, as $R(t)$ and $r(t)$ are not directly observable quantities, there are no “true” values by which to validate our model. Instead, we have focussed on comparison with the government-published estimates and considered the amount of information that the ONS-based estimates alone could yield. This implicitly treats the government-published estimates as reliable and accurate estimates of the true values of $R(t)$ and $r(t)$.

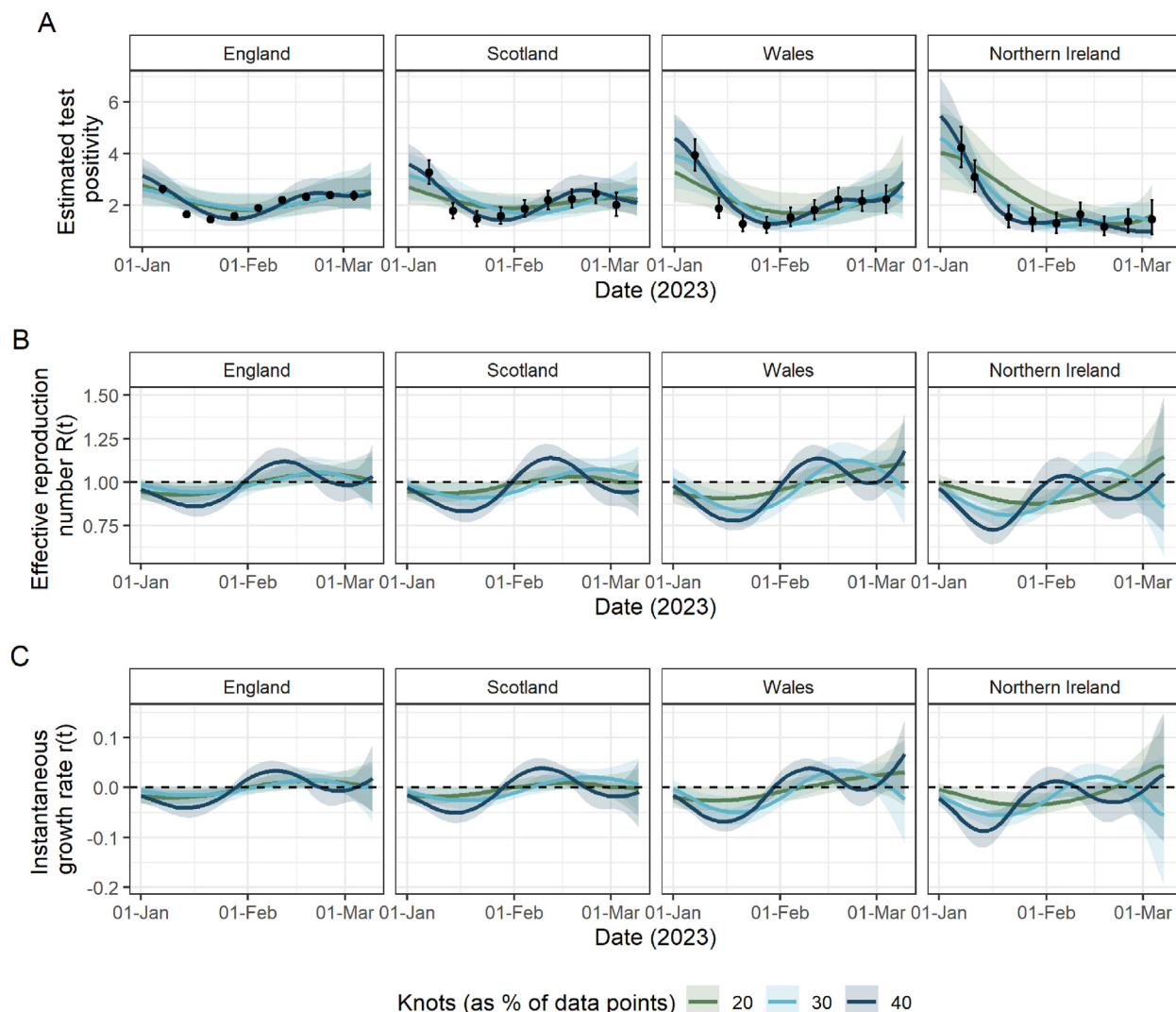


Fig. 5. Fit to ONS CIS data and resulting estimates of $R(t)$ and $r(t)$ for England, Scotland, Wales and Northern Ireland from January 2023–March 2023, at which point the ONS CIS was initially “paused”. Throughout, lines represent median values and shaded areas are the corresponding 95% credible intervals (A) Spline model fits ($\hat{p}(t)$) to data (black points; ONS point estimate with 95% credible intervals) with the number of knots totalling 20% (green), 30% (light blue) and 40% (purple) of the total data points. (B) The estimated effective reproduction number ($\hat{R}(t)$) resulting from the spline fits in (A). The dashed line highlights the epidemic growth threshold of 1. (C) The estimated instantaneous growth rate ($\hat{r}(t)$) resulting from the spline fits in (A). The dashed line highlights the epidemic growth threshold of 0.

Fourth, although our results suggest that our ONS-based estimates provide more timely insights into epidemic trends than the corresponding government-based estimates, this does not consider the time delays between the collection and publication of the ONS CIS data. Although this delay would not change the date on which trends occur, it would change the date when the trends would be able to be estimated and so real-time modelling gains may not be fully realised. This could be overcome by more timely access to the ONS CIS data, but this may compromise the data being publicly available and could also make estimation more logistically challenging. Although not a limitation of our work per se, it is also important to highlight that the ONS CIS was expensive to run (approximately £945m to the end of December 2022 ([COVID-19 Infection Survey Cost, 2023](#))), which undoubtedly raises questions regarding the long-term implementation of similar schemes in the future.

4.1. Conclusions

In this study, we have demonstrated the utility and validity of this model for estimating $R(t)$ and $r(t)$ using the ONS CIS test positivity data, which was a primary source measuring the ongoing epidemics in the four nations of the UK. Our model

provides a reliable means by which to track the ongoing epidemics in each of the four nations of the UK after the scaling down of SARS-CoV-2 surveillance, which was largely reduced due to the transition from “emergency” to “endemic” state in Spring 2022. Our study highlights the critical role that studies such as the ONS CIS play as part of an effective and data-driven epidemic response. Although the ONS CIS was “paused” in March 2023, the WCIS (announced in October 2023) should provide a new surveillance data stream, similar to the ONS CIS data, for the winter period 2023/24. Consequently, this model is uniquely placed to provide a reliable method for tracking the spread of the UK epidemic throughout this potentially challenging winter period.

Data availability

All data and code used are available from: <https://github.com/ruthmccabe/ons-test-positivity-model>.

CRediT authorship contribution statement

Ruth McCabe: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Gabriel Danelian:** Writing – review & editing. **Jasmina Panovska-Griffiths:** Writing – review & editing, Supervision. **Christl A. Donnelly:** Writing – review & editing, Supervision, Methodology.

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.

Acknowledgements

This work was supported by the NIHR HPRU in Emerging and Zoonotic Infections, a partnership between the United Kingdom Health Security Agency (UKHSA), University of Oxford, University of Liverpool and Liverpool School of Tropical Medicine [grant number NIHR200907 supporting RM and CAD]; and the MRC Centre for Global Infectious Disease Analysis [grant number MR/X020258/1], funded by the UK Medical Research Council (MRC). This UK funded award is carried out in the frame of the Global Health EDCTP3 Joint Undertaking. RM was also supported by the UKHSA and the Isaac Newton Institute (INI) Knowledge Transfer Network (KTN) in funding and coordinating a 3-month placement at the UK Health Security Agency, respectively.

Disclaimer: “The views expressed are those of the authors and not necessarily those of the United Kingdom (UK) Department of Health and Social Care, EU, FCDO, MRC, National Health Service, NIHR, PHE or UKHSA. The funding bodies had no role in the design of the study, analysis and interpretation of data and in writing the manuscript.”

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.idm.2024.01.011>.

References

- Abbott, S., & Funk, S. (2022). Estimating epidemiological quantities from repeated cross-sectional prevalence measurements. *medRxiv*. <https://doi.org/10.1101/2022.03.29.22273101>
- Abbott, S., Sherratt, K., Gerstung, M., & Funk, S. (2022). Estimation of the test to test distribution as a proxy for generation interval distribution for the Omicron variant in England. *medRxiv*. <https://doi.org/10.1101/2022.01.08.22268920>, 2022.01.08.22268920.
- Anderson, R., Donnelly, C. A., Hollingsworth, D., Keeling, M., Vegvari, C., Baggaley, R., et al. (2020). *Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: Methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation*. The Royal Society.
- Barai, S. V., & Reich, Y. (1999). Ensemble modelling or selecting the best model: Many could be better than one. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 13(5), 377–386. <https://doi.org/10.1017/S0890060499135029>
- Birrell, P., Blake, J., van Leeuwen, E., De Angelis, D., & MRC Biostatistics Unit COVID-19 Working Group. (2022). *COVID-19: Nowcast and forecast*. MRC Biostatistics Unit. <https://joshuablake.co.uk/public-RTM-reports/iframe.html>.
- Birrell, P., Blake, J., van Leeuwen, E., Gent, N., & De Angelis, D. (2021). Real-time nowcasting and forecasting of COVID-19 dynamics in England: The first wave. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829), Article 20200279. <https://doi.org/10.1098/rstb.2020.0279>
- Clarke, J., Beaney, T., & Majeed, A. (2022). UK scales back routine covid-19 surveillance. *BMJ*, 376, o562. <https://doi.org/10.1136/bmj.o562>
- Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9), 1505–1512. <https://doi.org/10.1093/aje/kwt133>
- Coronavirus (COVID-19) in the UK. (2023). Gov.uk. <https://coronavirus.data.gov.uk/>.
- Coronavirus (COVID-19) Infection Survey, UK. (2022). Office for National Statistics. <https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/covid19infectionsurvey>.
- Coronavirus (COVID-19) Infection Survey, UK: 15 July 2022. (2022). Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurvey/pilot/15july2022>.
- Coronavirus (COVID-19) Infection Survey, UK: 24 March 2023. (2023). Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurvey/pilot/24march2023>.

- Coronavirus (COVID-19) Infection Survey: England: Previous versions. (2023). Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata/2022>.
- Coronavirus (COVID-19) Infection Survey: Methods and further information. (2023). Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveyspilotmethodsandfurtherinformation>.
- Coronavirus (COVID-19) Infection Survey: Northern Ireland: Previous versions. (2023). Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/covid19infectionsurveysnorthernireland/2022>.
- Coronavirus (COVID-19) Infection Survey: Scotland: Previous versions. (2023). Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/covid19infectionsurveyscotland/2022>.
- Coronavirus (COVID-19) Infection Survey: Wales: Previous versions. (2023). Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/covid19infectionsurveywales/2022>.
- Coronavirus (COVID-19): Test and Protect - Transition Plan. (2022). Scottish Government. <https://www.gov.scot/publications/test-protect-transition-plan/covid-19-infection-survey-cost>.
- COVID-19 Infection Survey cost. (2023). Office for National Statistics. <https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/covid19infectionsurveycost>.
- COVID-19 Infection Survey participants thanked for 'huge contribution' to pandemic response. (2023). UK Health Security Agency. <https://www.gov.uk/government/news/covid-19-infection-survey-participants-thanked-for-huge-contribution-to-pandemic-response>.
- COVID-19 Response. (2022). Living with COVID-19. Cabinet Office. <https://www.gov.uk/government/publications/covid-19-response-living-with-covid-19/covid-19-response-living-with-covid-19>.
- COVID-19 test, Trace and Protect transition plan. (2022). Northern Ireland: Department of Health. <https://www.health-ni.gov.uk/sites/default/files/publications/health/Test-Trace-Transition-Plan.pdf>.
- COVID-19 genomic surveillance. (2023). Wellcome Sanger Institute. <https://covid19.sanger.ac.uk/lineages/raw>.
- Danon, L., Brooks-Pollock, E., Bailey, M., & Keeling, M. (2021). A spatial model of COVID-19 transmission in England and Wales: Early spread, peak timing and the impact of seasonality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829), Article 20200272. <https://doi.org/10.1098/rstb.2020.0272>
- Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., & Edmunds, W. J. (2020). Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: A modelling study. *The Lancet Public Health*, 5(7), e375–e385. [https://doi.org/10.1016/S2468-2667\(20\)30133-X](https://doi.org/10.1016/S2468-2667(20)30133-X)
- Dean, N. E. (2022). Tracking COVID-19 infections: Time for change. *Nature*, 602(185). <https://doi.org/10.1038/d41586-022-00336-8>
- Di Napoli, M., Carotenuto, F., Cevasso, A., Confuorto, P., Di Martire, D., Firpo, M., et al. (2020). Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides*, 17(8), 1897–1914. <https://doi.org/10.1007/s10346-020-01392-9>
- Eales, O., Ainslie, K. E. C., Walters, C. E., Wang, H., Atchison, C., Ashby, D., et al. (2022). Appropriately smoothing prevalence data to inform estimates of growth rate and reproduction number. *Epidemics*, 40. <https://doi.org/10.1016/j.epidem.2022.100604>
- Eales, O., Wang, H., Haw, D., Ainslie, K. E. C., Walters, C. E., Atchison, C., et al. (2022). Trends in SARS-CoV-2 infection prevalence during England's roadmap out of lockdown, January to July 2021. *PLoS Computational Biology*, 18(11), Article e1010724. <https://doi.org/10.1371/journal.pcbi.1010724>
- Elliott, P., Eales, O., Steyn, N., Tang, D., Bodinier, B., Wang, H., et al. (2022). Twin peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 epidemics in England. *Science*, 376(6600). <https://doi.org/10.1126/science.abq4411>. eabq4411.
- Elliott, P., Haw, D., Wang, H., Eales, O., Walters, C. E., Ainslie, K. E. C., et al. (2021). Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant. *Science*, 374(6574). <https://doi.org/10.1126/science.abl9551>. eabl9551.
- Government publishes latest R number. (2020). Gov.uk. <https://www.gov.uk/government/news/government-publishes-latest-r-number>.
- Grenouillet, G., Buisson, L., Casajus, N., & Lek, S. (2011). Ensemble modelling of species distribution: The effects of geographical and environmental ranges. *Ecography*, 34(1), 9–17. <https://doi.org/10.1111/j.1600-0587.2010.06152.x>
- Hart, W. S., Abbott, S., Endo, A., Hellewell, J., Miller, E., Andrews, N., et al. (2022). Inference of the SARS-CoV-2 generation time using UK household data. *Elife*, 11, Article e70767. <https://doi.org/10.7554/eLife.70767>
- Hart, W. S., Miller, E., Andrews, N. J., Waight, P., Maini, P. K., Funk, S., et al. (2022). Generation time of the alpha and delta SARS-CoV-2 variants: An epidemiological analysis. *The Lancet Infectious Diseases*, 22(5), 603–610. [https://doi.org/10.1016/S1473-3099\(22\)00001-9](https://doi.org/10.1016/S1473-3099(22)00001-9)
- Homan, M. D., & Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Keeling, M. J., Dyson, L., Guyver-Fletcher, G., Holmes, A., Semple, M. G., Tildesley, M. J., et al. (2022). Fitting to the UK COVID-19 outbreak, short-term forecasts and estimating the reproductive number. *Statistical Methods in Medical Research*, 31(9), 1716–1737. <https://doi.org/10.1177/09622802211070257>
- Lythgoe, K. A., Golubchik, T., Hall, M., House, T., Cahuantzi, R., MacIntyre-Cockett, G., et al. (2023). Lineage replacement and evolution captured by 3 years of the United Kingdom Coronavirus (COVID-19) infection survey. *Proceedings. Biological Sciences*, 290(2009), Article 20231284. <https://doi.org/10.1098/rspb.2023.1284>
- Maishman, T., Schaap, S., Silk, D. S., Nevitt, S. J., Woods, D. C., & Bowman, V. E. (2022). Statistical methods used to combine the effective reproduction number, R(t), and other related measures of COVID-19 in the UK. *Statistical Methods in Medical Research*, 31(9), 1757–1777. <https://doi.org/10.1177/0962280221109506>
- Moore, R. E., Rosato, C., & Maskell, S. (2022). Refining epidemiological forecasts with simple scoring rules. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 380(2233), Article 20210305. <https://doi.org/10.1098/rsta.2021.0305>
- Morvan, M., Jacomo, A. L., Souque, C., Wade, M. J., Hoffmann, T., Pouwels, K., et al. (2022). An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence. *Nature Communications*, 13(1), 4313. <https://doi.org/10.1038/s41467-022-31753-y>
- Northern Ireland population mid-year estimate. (2022). Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/timeseries/nipop/pop>.
- Parag, K. V. (2021). Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Computational Biology*, 17(9), Article e1009347. <https://doi.org/10.1371/journal.pcbi.1009347>
- Park, J., Bevan, L., Sanchez-Marroquin, A., Danelian, G., Bayley, T., Manley, H., et al. (2023). Combining models to generate a consensus effective reproduction number R for the COVID-19 epidemic status in England. *medRxiv*. <https://doi.org/10.1101/2023.02.27.23286501>
- 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., Pouwels, K. B., et al. (2021). Community prevalence of SARS-CoV-2 in England from april to november, 2020: Results from the ONS Coronavirus infection survey. *The Lancet Public Health*, 6(1), e30–e38. [https://doi.org/10.1016/S2468-2667\(20\)30282-6](https://doi.org/10.1016/S2468-2667(20)30282-6)
- The R value and growth rate. (2022). Gov.Uk. <https://www.gov.uk/guidance/the-r-value-and-growth-rate>.
- REal-time assessment of community transmission (REACT) study. (2022). Imperial College London. <https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/>.
- Reproduction number (R) and growth rate: Methodology. (2021). Gov.Uk. <https://www.gov.uk/government/publications/reproduction-number-r-and-growth-rate-methodology/reproduction-number-r-and-growth-rate-methodology>.
- Rhodes, S., Wilkinson, J., Pearce, N., Mueller, W., Cherrie, M., Stocking, K., et al. (2022). Occupational differences in SARS-CoV-2 infection: Analysis of the UK ONS COVID-19 infection survey. *Journal of Epidemiology & Community Health*, 76(10), 841–846. <https://doi.org/10.1136/jech-2022-219101>
- SARS-CoV-2 variants in analyzed sequences, United Kingdom. Our World in Data, (2023). <https://ourworldindata.org/grapher/covid-variants-area?country=-GBR>.

- Scott, J. A., Gandy, A., Mishra, S., Bhatt, S., Flaxman, S., Unwin, H. J. T., et al. (2021). *Epidemia: An R package for semi-mechanistic bayesian modelling of infectious diseases using point processes*. ArXiv. <https://doi.org/10.48550/arxiv.2110.12461>
- Shabnam, S., Razieh, C., Dambha-Miller, H., Yates, T., Gillies, C., Chudasama, Y. V., et al. (2023). Socioeconomic inequalities of long COVID: A retrospective population-based cohort study in the United Kingdom. *Journal of the Royal Society of Medicine*, 116(8), 263–273. <https://doi.org/10.1177/01410768231168377>
- Simpson, C. R., Robertson, C., Vasileiou, E., Moore, E., McCowan, C., Agrawal, U., et al. (2021). Temporal trends and forecasting of COVID-19 hospitalisations and deaths in Scotland using a national real-time patient-level data platform: A statistical modelling study. *The Lancet Digital Health*, 3(8), e517–e525. [https://doi.org/10.1016/S2589-7500\(21\)00105-9](https://doi.org/10.1016/S2589-7500(21)00105-9)
- Slides and datasets to accompany coronavirus press conference: 22 October 2020. (2020). Gov.uk. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/928812/Press_conference_slides_-_Thursday_22_October_.pdf
- Slides to accompany coronavirus press conference: 11 May 2020. (2020). Gov.uk. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/884352/slides_-_11_05_2020.pdf
- Tapper, J. (2022). Dismay as funding for UK's 'world-beating' Covid trackers is axed. *The Observer*. <https://www.theguardian.com/world/2022/mar/12/dismay-as-funding-for-uks-world-beating-covid-trackers-is-axed>
- The COVID-19 Genomics UK (COG-UK) consortium. (2020). An integrated national scale SARS-CoV-2 genomic surveillance network. *The Lancet Microbe*, 1(3), e99–e100. [https://doi.org/10.1016/S2666-5247\(20\)30054-9](https://doi.org/10.1016/S2666-5247(20)30054-9)
- Together for a safer future: Wales' long-term Covid-19 transition from pandemic to endemic. (2022). Welsh Government. <https://gov.wales/sites/default/files/publications/2022-05/wales-long-term-covid-19-transition-from-pandemic-to-endemic.pdf>
- UKHSA and ONS launch new winter COVID-19 infection study. (2023). UK Health Security Agency. <https://ukhsa-newsroom.prgloo.com/news/ukhsa-and-ons-launch-new-winter-covid-19-infection-study>
- Vöhringer, H. S., Sanderson, T., Sinnott, M., De Maio, N., Nguyen, T., Goater, R., et al., Team, T. W. S. I. C.-19 S.. (2021). Genomic reconstruction of the SARS-CoV-2 epidemic in England. *Nature*, 600(7889), 506–511. <https://doi.org/10.1038/s41586-021-04069-y>
- Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings. Biological Sciences*, 274(1609), 599–604. <https://doi.org/10.1098/rspb.2006.3754>
- Wallinga, J., & Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, 160(6), 509–516. <https://doi.org/10.1093/aje/kwh255>
- Ward, H., Atchison, C., Whitaker, M., Ainslie, K. E. C., Elliott, J., Okell, L., et al. (2021). SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nature Communications*, 12(1), 905. <https://doi.org/10.1038/s41467-021-21237-w>
- Zhou, L., Lai, K. K., & Yu, L. (2010). Least squares support vector machines ensemble models for credit scoring. *Expert Systems with Applications*, 37(1), 127–133. <https://doi.org/10.1016/j.eswa.2009.05.024>