



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.

## Journal Pre-proof

Extending EpiEstim to estimate the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study

Sangeeta Bhatia, Jack Wardle, Rebecca K. Nash, Pierre Nouvellet, Anne Cori



PII: S1755-4365(23)00028-2

DOI: <https://doi.org/10.1016/j.epidem.2023.100692>

Reference: EPIDEM 100692

To appear in: *Epidemics*

Received date: 20 December 2022

Revised date: 20 April 2023

Accepted date: 29 May 2023

Please cite this article as: S. Bhatia, J. Wardle, R.K. Nash et al., Extending EpiEstim to estimate the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study. *Epidemics* (2023), doi: <https://doi.org/10.1016/j.epidem.2023.100692>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Extending EpiEstim to estimate the transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a case-study**

Sangeeta Bhatia<sup>1, 2\*</sup>, Jack Wardle<sup>1\*</sup>, Rebecca K Nash<sup>1\*</sup>, Pierre Nouvellet<sup>1, 3</sup>, Anne Cori<sup>1\*\*</sup>

**1** MRC Centre for Global Infectious Disease Analysis, Jameel Institute, School of Public Health, Imperial College London.

**2** NIHR Health Protection Research Unit in Modelling and Health Economics, Modelling & Economics Unit, UK Health Security Agency, London, UK.

**3** School of Life Sciences, University of Sussex, Brighton, UK.

\* equal contribution

\*\* corresponding author (a.cor@imperial.ac.uk)

## Abstract

The evolution of SARS-CoV-2 has demonstrated that emerging variants can set back the global COVID-19 response. The ability to rapidly assess the threat of new variants is critical for timely optimisation of control strategies.

We present a novel method to estimate the effective transmission advantage of a new variant compared to a reference variant combining information across multiple locations and over time. Through an extensive simulation study designed to mimic real-time epidemic contexts, we show that our method performs well across a range of scenarios and provide guidance on its optimal use and interpretation of results. We also provide an open-source software implementation of our method. The computational speed of our tool enables users to rapidly explore spatial and temporal variations in the estimated transmission advantage.

We estimate that the SARS-CoV-2 Alpha variant is 1.46 (95% Credible Interval 1.44-1.47) and 1.29, (95% CrI 1.29-1.30) times more transmissible than the wild type, using data from England and France respectively. We further estimate that Delta is 1.77 (95% CrI: 1.69-1.85) times more transmissible than Alpha (England data).

Our approach can be used as an important first step towards quantifying the threat of emerging or co-circulating variants of infectious pathogens in real-time.

## Introduction

The SARS-CoV-2 pandemic has highlighted the potentially dramatic influence that emerging novel pathogen variants can have on transmission dynamics and on the control measures needed to mitigate the epidemic burden. The emergence of the Alpha variant of SARS-CoV-2 in September 2020, and of the Delta variant in December 2020 drastically altered the trajectory of the COVID-19 epidemic in several countries leading to renewed imposition of public health measures such as lockdowns [1, 2]. The continued high level of transmission of SARS-CoV-2 globally makes the emergence of new variants very likely. As of April 2023, the World Health Organization has classified five variants of SARS-CoV-2 as “variants of concern” or VOCs (i.e. Alpha, Beta, Gamma, Delta, and Omicron), because of their increased transmissibility, severity, and/or immune escape properties compared to the circulating SARS-CoV-2 variants [3]. Rapidly quantifying characteristics of such emerging variants is critical to anticipate their potential impact and adjust interventions accordingly. Estimates of the transmission advantage of a new variant over previously circulating variants can help re-evaluate key metrics which depend on the reproduction number, e.g. the herd-immunity threshold, and adjust short-term projections of the epidemic trajectory. They can also help to disentangle the impact on transmission levels of the newly introduced variant from other factors, such as concurrent changes in control measures [4]. Shortly after the emergence of the Alpha variant in England in September 2020 [5], a number of studies aimed to estimate its transmission potential, compared to the previously circulating non-VOC lineages [6–14]. More recently, several papers have evaluated the transmissibility of the Delta variant compared to Alpha [4, 15–19], and of one or more VOCs [9, 18, 20–25]. All of these studies have developed new approaches to estimate the transmission advantages of the Alpha and the Delta variants, often synthesising evidence from multiple data sources including genomic data. The time and expertise required to design and implement such approaches, with methods tailored to the specificity of each dataset and context, often limit their widescale and real-time use.

In this study, we present a new Bayesian inference method, MV-EpiEstim (for Multi-Variant EpiEstim), to estimate in real-time the transmission advantage of a new variant of a pathogen compared to a reference variant, using simple data consisting of the time series of incidence of cases of each variant in one or more locations. The aims of this work are to: 1) develop a method and tool for such analyses; 2) assess how well it can estimate the transmission advantage across a range of simulated and real-life epidemic scenarios.

In the rest of the manuscript, we refer to different “variants” but the method can be equally applied to different strains. We present the method for one reference and one new variant, but the method naturally extends to more than one new variant. Our work builds on a previously published methodology [26, 27] to estimate the instantaneous reproduction number  $R_t$  (defined as the average number of secondary cases that an individual infected at time  $t$  would generate if conditions remained the same as at time  $t$ ).

We assume that locally, the transmissibility of all variants follows the same temporal pattern, i.e. the reproduction number of the new variant is the same as that of the reference variant, albeit with a multiplicative factor. We refer to this multiplicative factor as the “effective transmission advantage” of the new variant, compared to the reference variant. We further assume that the effective transmission advantage remains constant over a user-defined time-window and across all locations under consideration. Note that both the time-window and the set of locations over which the effective transmission advantage is assumed to be constant can be varied by the user.

We provide an open source implementation of our method in the R package EpiEstim [28]. The approach, which we validate on an extensive simulation study, is computationally efficient as it takes advantage of an analytical formulation of marginal posterior densities of both the instantaneous reproduction number for the reference variant, and the transmission advantage of the new variant.

We illustrate the use of our tool by retrospectively estimating the effective transmission advantage of SARS-CoV-2 VOCs (Alpha, Beta/Gamma, and Delta) over the previously circulating variants using data from England and France. In addition, we perform a literature review to summarise other existing approaches and tools available for estimating the transmission advantage of new variants from incidence data. We show that the estimates from our method are consistent with those from other studies, and that our fast ready-to-use tool allows timely estimation and easy exploration of changes in the transmission advantage over time and space. Our inference framework and open source software should allow rapid quantification and monitoring of the effective transmission advantage of future new variants in real-time.

## Methods

We extend the methodology from Cori et al. [26] and Thompson et al. [27] to develop an inference framework for jointly estimating the transmissibility (instantaneous reproduction number  $R_t$ ) of a reference variant and the effective transmission advantage of novel variants, compared to the reference. For simplicity, we present the method for two variants only (a reference and a new variant). The method is applicable to, and has been implemented for, estimating the transmission advantages of multiple variants over a single reference.

**Assumptions.** Our method relies on daily incidence data of the reference and the variant. Where data from more than one location are used, we assume that the epidemics in each location are independent and closed except for cases of each variant on the first day who are assumed to be imported. The effective reproduction number is defined as the ratio of newly infected cases to the total infectiousness (due to past cases) in a location. For more details, see [26, 27].

**Notation.** We use the following notations:

- Indexes  $t$  for time,  $l$  for location and  $v$  for variant, with  $v = 0$  denoting the reference variant and  $v = 1$  the new variant,

- $n_l$  the number of locations considered
- $T$  the number of days of observation
- $I_t^{l,v}$  denotes the number of incident cases of variant  $v$  at time  $t$  in location  $l$ ,
- $R_t^{l,v}$  denotes the instantaneous reproduction number for variant  $v$  at time  $t$  in location  $l$ . For simplicity we use  $R_t^l$  to denote the instantaneous reproduction number for the reference variant in location  $l$  i.e.  $R_t^l = R_t^{l,0}$ .
- $w^v$  is the probability mass function of the discrete serial interval for variant  $v$ , assumed the same across all locations, but potentially different between variants ( $w_s^v$  is the probability that the serial interval lasts  $s$  days,  $s = 1, \dots, SI_{max}$ ; and we assume  $w_0^v = 0$ ).
- $\Lambda_t^{l,v} = \sum_{s=1}^t I_{t-s}^{l,v} w_s^v$  is the overall infectiousness for variant  $v$  at time  $t$  and in location  $l$  due to past incident cases of that variant in that location.
- For simplicity we introduce the generic notation  $\mathbf{X}_t^{L,V} = \{X_t^{l,v}\}_{l=1, \dots, n_l; v=0,1}$  for the variable  $X$  at time  $t$  across all locations and both variants.

We assume that  $R_t^{l,1} = \epsilon R_t^l$ , i.e. the reproduction number of the new variant is proportional to that for the reference variant; the proportional factor  $\epsilon$  is the effective transmission advantage (if  $\epsilon > 1$ , or disadvantage if  $\epsilon < 1$ ) of the new variant compared to the reference variant, assumed constant over a time-window and across a set of locations defined by the user.

We explored values of  $\epsilon > 1$  in all simulation scenarios as values of  $\epsilon < 1$  correspond to swapping the reference and new variant.

We assume the number of secondary infections generated by each case is Poisson distributed. Under these assumptions, the likelihood of the time series of incident cases of the reference and the new variants can be written as

$$\begin{aligned}
 \mathcal{L}_t &= P(\mathbf{I}_t^{L,V} | \mathbf{I}_0^{L,V}, \dots, \mathbf{I}_{t-1}^{L,V}, \mathbf{w}^V, \mathbf{R}_t^{L,V}, \epsilon) \\
 &= \prod_{l=1}^{n_l} \prod_{v=0}^1 \frac{(R_t^{l,v})^{I_t^{l,v}} e^{-R_t^{l,v} \Lambda_t^{l,v}}}{I_t^{l,v}!} \\
 &\propto \prod_{l=1}^{n_l} \left( (R_t^l)^{\sum_{v=0}^1 I_t^{l,v}} \epsilon^{I_t^{l,1}} e^{-R_t^l (\Lambda_t^{l,0} + \epsilon \Lambda_t^{l,1})} \right) \\
 &\propto \epsilon^{\sum_{l=1}^{n_l} I_t^{l,1}} e^{-\sum_{l=1}^{n_l} R_t^l (\Lambda_t^{l,0} + \epsilon \Lambda_t^{l,1})} \prod_{l=1}^{n_l} (R_t^l)^{\sum_{v=0}^1 I_t^{l,v}}
 \end{aligned}$$

We assume Gamma priors for each  $R_t^l$ , with same shape  $a$  and scale  $b$  across times and locations, and for  $\epsilon$ , with shape  $c$  and scale  $d$ . The joint posterior

distribution of parameters given the observations is (assuming the serial interval distributions for both variants  $\mathbf{w}^V$  are known):

$$\begin{aligned} & P\left(\epsilon, \mathbf{R}_1^L, \dots, \mathbf{R}_T^L | \mathbf{I}_0^{L,V}, \dots, \mathbf{I}_T^{L,V}, \mathbf{w}^V\right) \\ & \propto \prod_{t=1}^T \left( \mathcal{L}_t \prod_{l=1}^{n_l} (R_t^l)^{a-1} e^{-\frac{R_t^l}{b}} \right) \epsilon^{c-1} e^{-\frac{\epsilon}{d}} \\ & \propto \left( \prod_{t=1}^T \prod_{l=1}^{n_l} (R_t^l)^{a-1 + \sum_{v=0}^1 I_t^{l,v}} \right) \epsilon^{c-1 + \sum_{t=1}^T \sum_{l=1}^{n_l} I_t^{l,1}} \\ & \quad \times e^{-\frac{\epsilon}{d} - \sum_{t=1}^T \sum_{l=1}^{n_l} R_t^l \left( \frac{1}{b} + \Lambda_t^{l,0} + \epsilon \Lambda_t^{l,1} \right)} \end{aligned}$$

The marginal posterior distribution for  $\epsilon$  given the data (i.e. the incidence for all variants, at all locations and for all time steps) and given the reproduction number for the reference variant in all locations and at all time steps is given by:

$$\begin{aligned} & P\left(\epsilon | \mathbf{I}_0^{L,V}, \dots, \mathbf{I}_T^{L,V}, \mathbf{R}_1^L, \dots, \mathbf{R}_T^L, \mathbf{w}^V\right) \\ & \propto \epsilon^{c-1 + \sum_{t=1}^T \sum_{l=1}^{n_l} I_t^{l,1}} e^{-\epsilon \left( \frac{1}{d} + \sum_{t=1}^T \sum_{l=1}^{n_l} R_t^l \Lambda_t^{l,1} \right)} \end{aligned}$$

Therefore, the marginal posterior distribution of  $\epsilon$  given the data and other parameters is a Gamma distribution with shape  $c + \sum_{t=1}^T \sum_{l=1}^{n_l} I_t^{l,1}$  and scale  $\frac{1}{\frac{1}{d} + \sum_{t=1}^T \sum_{l=1}^{n_l} R_t^l \Lambda_t^{l,1}}$ .

Similarly, the marginal posterior distribution for  $R_t^l$  at time step  $t$  and in location  $l$  given the data,  $\epsilon$ , and the reproduction number at other locations and time steps, is given by:

$$\begin{aligned} & P(R_t^l | \mathbf{I}_0^{L,V}, \dots, \mathbf{I}_T^{L,V}, \epsilon, \\ & \quad \mathbf{R}_1^L, \dots, \mathbf{R}_{t-1}^L, \mathbf{R}_{t+1}^L, \dots, \mathbf{R}_T^L, \mathbf{R}_t^1, \dots, \mathbf{R}_t^{l-1}, \mathbf{R}_t^{l+1}, \dots, \mathbf{R}_t^{n_l}, \mathbf{w}^V) \\ & \propto (R_t^l)^{a-1 + \sum_{v=0}^1 I_t^{l,v}} e^{-R_t^l \left( \frac{1}{b} + \Lambda_t^{l,0} + \epsilon \Lambda_t^{l,1} \right)} \end{aligned}$$

Therefore, the marginal posterior distribution of  $R_t^l$  given the data and other parameters is a Gamma distribution with shape  $a + \sum_{v=0}^1 I_t^{l,v}$  and scale  $\frac{1}{\frac{1}{b} + \Lambda_t^{l,0} + \epsilon \Lambda_t^{l,1}}$ .

**Precision in  $\epsilon$  estimates.** An analytical formulation of the marginal posterior distribution of  $\epsilon$  allows us to quantify the expected precision in  $\epsilon$  estimates. Since the coefficient of variation (CV) of a gamma distributed random variable

is  $1/\sqrt{\text{shape}}$ , the CV of the marginal posterior distribution of  $\epsilon$  is

$$\begin{aligned} CV & \left( \epsilon | \mathbf{I}_0^{L,V}, \dots, \mathbf{I}_T^{L,V}, \mathbf{R}_1^L, \dots, \mathbf{R}_T^L, \mathbf{w}^V \right) \\ & = \frac{1}{\sqrt{a + \sum_{v=0}^1 I_t^{l,v}}} \\ & \leq \frac{1}{\sqrt{\sum_{v=0}^1 I_t^{l,v}}} \end{aligned}$$

That is, the CV of  $\epsilon$  scales inversely with the square root of the total incidence across all locations in the time-window used for estimation.

**Monte Carlo Markov Chain (MCMC) inference.** The analytical formulation of the marginal posterior distributions for  $R_t^l$  and  $\epsilon$  allow us to use a multi-stage Gibbs sampler for the MCMC inference.

To initialise  $R_t^l$ , we use EpiEstim to estimate a single reproduction number for the reference variant over the entire time period of observations, and using incidence aggregated across all locations. The posterior mean is then used as the initial value for  $R_t^l$ . We independently use the same approach to estimate a single reproduction number for the new variant;  $\epsilon$  is then initialised to the median of the ratio of the reproduction numbers for the new variant and the reference.

We first sample from the marginal distribution of  $R_t^l$ , conditional on  $\epsilon$ , and then we sample from the marginal distribution of  $\epsilon$ , conditional on the newly sampled value of  $R_t^l$ . We repeat this procedure for a fixed number of iterations or until convergence is achieved. Convergence is assessed using Gelman-Rubin convergence diagnostic [29] using 1.1 as a cut-off value.

**Implementation.** The inference method is implemented in a new function “estimate.advantage” of the development version of the R package EpiEstim available at <https://github.com/mrc-ide/EpiEstim>.

**Choosing a time-period for estimation of  $\epsilon$ .** Users can set the time period over which estimation will be carried out. We recommend that the estimation is started after at least one generation of cases has been observed. The default starting point in the software is set to the first day of non-zero incidence across all locations plus the 95<sup>th</sup> percentile of the serial interval distribution.

**Classification of a variant.** We used the posterior distribution of the effective transmission advantage to classify a new variant (in relation to the reference variant) as:

- ‘More transmissible’ if the 2.5<sup>th</sup> quantile of the posterior distribution was greater than 1;
- ‘Less transmissible’ if the 97.5<sup>th</sup> quantile of the posterior distribution was less than 1; and,
- ‘Unclear’ if the 95% CrI contained 1.



We note that here we used the 2.5 and 97.5 posterior percentiles for variant classification, which provides an easy metric to summarise across simulations. However our EpiEstim implementation of the approach provides the entire posterior distribution of the transmission advantage, and therefore the user could use different thresholds for classification, balancing the desired sensitivity and specificity. The user could also quantify the posterior probability that a novel variant is more transmissible.

**Method validation.** We assessed the validity of our method using a large simulation study, where we considered several scenarios with different values for the transmissibility of each variant, allowing for superspreading, under-reporting or time-varying transmission advantages, as well as differences in natural history between variants (Suppl Secs. 5 and 5.1). We measured method performance on the simulations using several metrics (detailed in Suppl Sec. 5.1). First, we measured the bias as the difference between the mean posterior estimate of the effective transmission advantage and its true value. Second, we measured uncertainty in the posterior distribution and considered the coverage probability, which measures whether uncertainty in estimates is adequate. Finally, we considered the ability of the model to adequately classify the variant as “more transmissible”, “less transmissible” than the reference or “unclear”.

**Literature review.** On 21st February 2023, we searched for all studies published from 2020 onward which estimated the transmission advantage of SARS-CoV-2 variants. The following search terms were entered into Web of Science: AB=((Transmission OR Transmissibility) & (Variant OR VOC) & (estimat\*) & (SARS-CoV-2 OR COVID)). In total, 336 studies were identified in the search and uploaded to Covidence for screening. 244 studies were excluded after title and abstract screening, and an additional 31 studies were excluded during full text screening. For completeness, 5 additional papers that were found prior to the search were also included. Of these 66 (61 + 5) studies, 53 explicitly provided one of more estimates of the transmission advantage of one variant over another. For these, we extracted the value and type of transmission advantage estimated, as well as the method used and its availability (or not) as packaged software. See section 6 of the Supplementary Material and the Supplementary Database for further details.

## Results

### Transmission advantage of SARS-CoV-2 variants

We used MV-EpiEstim to retrospectively estimate the transmission advantage of SARS-CoV-2 variants using data from England and France. The Alpha variant originated in late summer to early Autumn 2020 in England (before vaccination was initiated), where it became dominant in early 2021 (Fig. 1A). The Delta variant, first detected in India, emerged in England around March 2021 and accounted for most SARS-CoV-2 cases in all regions by late Spring 2021 (Fig. S1). England never experienced substantial transmission of the Beta and

Gamma variants, first detected in South Africa and Brazil respectively [3].

In France, the Alpha variant emerged in early 2021, rapidly dominating cases in metropolitan France and the French West Indies [30, 31]. The Beta and Gamma variants were also circulating from January 2021 in most regions, and accounted for the majority of cases in French Guyana and la Réunion from Spring 2021 [32] (Fig. S2). The Delta variant emerged in France in early June 2021 after the period covered in this study.

We considered daily variant-specific incidence data from 7 National Health Service (NHS) regions in England between 1<sup>st</sup> September 2020 and 14<sup>th</sup> March 2021 (Fig. S1), and from 18 ADM2 regions in France between 18<sup>th</sup> February and 30<sup>th</sup> May 2021 (Figs. S2 and S6).

For simplicity, we refer to all lineages of SARS-CoV-2 other than the VOCs that were circulating at the time as ‘wildtype’.  $R_t$  estimates obtained independently for the wildtype and for Alpha indicated that Alpha was more transmissible (Fig. 1B). However, the magnitude of the transmission advantage (naively estimated as the ratio between the two  $R_t$ s, see Suppl Sec. 3 for details) varied over time and across regions. Pooling these non-parametric estimates over time and regions yielded a highly uncertain and non-significant transmission advantage of 1.41 (95% Credible Interval (CrI) 0.86-2.01) for Alpha compared to the wildtype across all times and regions in England.

MV-EpiEstim allows further exploring how the effective transmission advantage, which we denote as  $\epsilon$ , varies over time and space, by estimating  $\epsilon$  over various temporal and spatial units, within which it is assumed constant.

We first assumed that  $\epsilon$  was constant across all regions but potentially varied over time. Using weekly data aggregated across regions,  $\epsilon$  estimates from MV-EpiEstim showed a strong temporal variation with the central estimate initially increasing from 1.01 (95% CrI 0.89-1.15) in October to 1.72 (95% CrI 1.67-1.78) in December, and then declining again to 1.24 (95% CrI 0.95-1.62) in March 2021. We found a similar trend over time when  $\epsilon$  was estimated independently for each region (Fig. S5). We also used MV-EpiEstim to estimate  $\epsilon$  separately for each NHS region assuming that it remains constant over time. This highlighted minor regional differences with  $\epsilon$  estimated across the whole time period ranging from 1.36 (95% CrI 1.33-1.39) in the South-East to 1.54 (95% CrI 1.50-1.58) in the Midlands (Fig. 1D, Suppl Tab. S1).

The consistent temporal variability in  $\epsilon$  estimates when using data from all or individual regions suggests an underlying change in transmission dynamics. Across the entire time period and all regions we found strong evidence that Alpha was more transmissible than the wildtype, with an overall central estimate of  $\epsilon$  at 1.45 (95% CrI 1.43-1.46) when ignoring differences in time and space (see also Suppl Tab. S1). However these results mask substantial temporal heterogeneity and small levels of spatial heterogeneity (Fig. S3).

We estimated a similar, albeit slightly lower, overall effective transmission advantage for Alpha using data from the 18 ADM2 regions in France (central  $\epsilon$  estimate assuming no temporal or spatial heterogeneity at 1.29 (95% CrI 1.29-1.31), see Fig. S6). As for Alpha in England, these estimates masked changes in the transmission advantage over time, declining from 1.46 (95% CrI 1.44-

1.48) in March to 0.90 (95% CrI 0.86-0.95) in May 2021 (Fig. S7), and to a lesser extent between regions (from 1.21 (95% CrI 1.20-1.23) in Île-de-France to 1.41 (95% CrI 1.37-1.46) in Bourgogne-Franche-Comté, excluding regions where Beta/Gamma were dominant, Fig. S6 and Suppl Tab. S2).

Following the same approach, and using data from France, we demonstrated that the Beta and Gamma variants (combined) are also more transmissible than the wildtype, with  $\epsilon$  estimated to be 1.25 (95% CrI 1.25-1.27). However behind this overall central estimate we identified a decline in  $\epsilon$  over time, and heterogeneities between regions Fig. S8 and Suppl Tab. S3). Finally, using data from England, we estimated that Delta is 1.77 (95% CrI 1.69-1.85) times more transmissible than Alpha (Fig. S10 and Suppl Tab. S4). The spatial and temporal trends in  $\epsilon$  estimates for Delta, while present, were less marked (Fig. S11).

### Method validation

The method performed well across most scenarios considered, with a small bias (defined as the difference between the mean posterior estimate and the true  $\epsilon$  value, Fig. 2). The coverage probability, which measures whether uncertainty in estimates is adequate (Suppl Sec. 5.1), was also good across most scenarios (Fig. S12 to Fig. S31).

MV-EpiEstim was able to accurately estimate the transmission advantage when variants were known to differ in their natural history (characterised by the serial interval distribution, i.e., the delay between onset of symptoms in a case and their infector, Fig. 2c and e, Suppl Secs. 5.4 and 5.6). We also explored a scenario typical of real-time outbreak analysis where the natural history of the new variant is different, but in the absence of information, is assumed to be the same as that of the reference (Suppl Secs. 5.5 and 5.7). Misspecifying the mean serial interval led to substantial bias (median bias ranged from -1.8 to 16.7) and poor coverage, especially when the transmission advantage was moderate (more than 1.5) and the mean serial interval of the new variant was much shorter (0.5 times) than that of the reference (Fig. 2d). Misspecifying the coefficient of variation of the serial interval had little impact on the quality of the estimates (range of median bias: -0.4 to 1.0), unless the transmission advantage was very high (more than 2, Fig. 2f).

Even in the presence of substantial superspreading (i.e., equivalent to that of SARS-CoV-1, Fig. 2b) or poor case-reporting (i.e., up to 80% cases not reported, Fig. S26), neither of which is explicitly accounted for by MV-EpiEstim, the transmission advantage remained unbiased (range of median bias with overdispersion parameter 0.1 (-0.3, 0.1); range of median bias with probability of reporting 0.2 (-0.5, 0.0)). However, coverage tended to be low in the presence of high superspreading or high underreporting, indicating that the credible intervals were too narrow in these scenarios (Suppl Secs. 5.8 and 5.9).

In all scenarios, using more days of data reduced both the bias and the uncertainty (defined as the posterior standard deviation) in the estimated effective transmission advantage (Fig. S14 and Suppl Secs. 5.3 to 5.7).

We used the uncertainty in the estimates of  $\epsilon$  (i.e., the width of the 95% CrI) to determine if the effective transmission advantage was significant and classify the variant as more or less transmissible than the reference (see Methods). Crucially, in many scenarios including some where the bias was substantial, MV-EpiEstim was able to correctly characterise a variant as being more transmissible than the reference. For instance, when the mean serial interval of the new variant was shorter but misspecified, the variant was correctly classified as more transmissible since  $\epsilon$  was over-estimated (Fig. S18, scenario type low). Conversely, when the mean serial interval of the new variant was longer but was misspecified, classification performance was generally poor and correct classification was only feasible with sufficient days of data and a large transmission advantage (Fig. S18, scenario type high).

We also tested the performance of our method in a scenario in which the transmission advantage is changing over time (Fig. S30). We show that, when applied to short time windows, our method is able to detect the changes in transmission advantage, but it can give misleading estimates if applied to a longer time window where the transmission advantage varies (Fig. S31).

More results demonstrating the performance of the method when using fewer days of data, two locations, time-varying  $R_t$  and accounting for underreporting are shown in Suppl Sec. 5.

## Literature review results

105 estimates for the transmission advantage of SARS-CoV-2 variants were found across 53 studies. Of the literature which provided estimates for the transmission advantage in the reproduction number  $R$ , the estimated advantage for Alpha was in the range of 1.35-1.75, with associated uncertainty estimates (95% CrIs/CIs) ranging from 1.02-2.30. Similarly, the range of the central estimates of the transmission advantage for the Delta variant in the literature was 1.5-2.4.

Out of the 53 studies, only 5 provided packaged code, with a single package requiring only incidence data for the estimation. All literature review results, including extracted estimates of the transmission advantages and hyperlinks to the available code and R packages can be found in the Supplementary Database.

## Discussion

In this study we present a novel method, MV-EpiEstim, to estimate the transmission advantage of a new variant of a pathogen over a reference variant. MV-EpiEstim builds on the EpiEstim method [26, 27]. As such, MV-EpiEstim offers the same functionalities as EpiEstim. Because it is based on analytical formulations of the marginal posterior densities, the run time of a typical analysis using MV-EpiEstim is less than a few minutes on a standard laptop. MV-EpiEstim is implemented as a new function in the R package EpiEstim [33].

To illustrate the use of MV-EpiEstim, we retrospectively estimated the effective transmission advantage of Alpha, Delta and Beta/Gamma combined over the wildtype variants in England and France. Our analyses showed substantial changes in the effective transmission advantage of Alpha over time, particularly in England, where our analysis covers the full period from early emergence to dominance. The consistency of these temporal trends across regions provides stronger evidence of underlying changes in transmission trends. Volz et al. found a broadly similar temporal trend in the transmission advantage in the UK [6], which Kraemer et al. suggested may be in part explained by spatial patterns of spread [34].

Our analysis also identified temporal changes in the transmission advantage of Beta and Gamma over the wildtype, and Delta over Alpha. Such temporal changes may be due to a combination of factors. First, the detection of a VOC can trigger interventions (such as increased testing and contact tracing) targeted at sub-populations in which the VOC is circulating (e.g. travellers). This can lead to a lower rate of spread of the VOC early on, before its spread is generalised, and therefore an apparent increase in the transmission advantage over time. Changes in population immunity over time, such as increasing immunity to Alpha and waning immunity to previous variants can also contribute to the observed temporal trends in its transmission advantage. Future work could explore extending our approach to incorporate explanatory variables such as the proportion of population susceptible to specific variants over time and across different locations.

The fast run-time and reliance solely on variant-specific incidence time-series and serial interval distributions make it easy to explore several hypotheses about spatio-temporal trends. We recommend that future users of MV-EpiEstim run multiple sensitivity analyses exploring the spatio-temporal heterogeneities in effective transmission advantage. Consistent signals of a significant transmission advantage independently estimated across these analyses can help raise early warnings about emergence of VOCs.

Our method works well across a range of simulated scenarios, designed to mimic a variety of real-time epidemic contexts, including in the presence of superspreading and when the natural history of the new variant is imperfectly characterised. As expected, the performance deteriorated with larger errors and lower coverage probability in scenarios with high superspreading or under-reporting, and with large misspecification of natural history. Of the packaged tools identified in our literature review, one did explicitly account for overdispersion, but did not demonstrate method performance on simulated data [35]. Moreover, other than a couple of studies, which did not provide a packaged tool [36, 37], all other approaches to estimate the transmission advantage suffer from a similar identifiability issue between a change in generation time versus a change in transmission (see Supplementary Database for details).

In the absence of precise information on natural history, MV-EpiEstim's fast run time offers the possibility of exploring various assumptions and in turn estimate a range of plausible transmission advantages. Our method is robust to moderate levels of under-reporting and temporal changes in reporting if these

affect both the reference and the variant equally.

Importantly, we show that our method can accurately characterise a variant as being ‘more’ or ‘less’ transmissible than a reference variant across many scenarios, including some where the performance at estimating the transmission advantage was only modest. This simple but robust characterisation could be as important as estimating the exact value of the transmission advantage, especially in informing public health response during the early phase of a new variant emerging. For example, in scenarios where the emerging variant has a shorter serial interval, as was the case for Omicron for example [38], our method will generally be able to detect an increase in transmission, albeit with an overestimated transmission advantage. Where the new variant has a longer serial interval though, our approach will have poor ability to detect the increased transmissibility. However, again, these caveats are common to many approaches.

Classification performance depends on multiple factors including the characteristics of the reference and new variants and the amount of data available for estimation. Across all scenarios, the probability of correctly classifying a more transmissible variant (i.e., true positive rate) increases with increasing baseline transmissibility, higher transmission advantage, and as more data are used for estimation. Conversely, in scenarios with high levels of superspreading or large misspecification of the natural history of the variant, the sensitivity of the classification is reduced (Figs. S17 and S23) but improves when either more data are used (Figs. S18 and S24) or with increasing transmission advantage. It is worth noting that in all scenarios, including scenarios where method sensitivity is low, the probability of misclassifying a variant as “more transmissible” when it is not (i.e., false positive rate) remains low (Suppl Tab. S6) i.e., the method specificity is high. Critically, we note that in such cases, the variant is classified as “unclear”, with very low probability of incorrectly classifying it as less transmissible. Further, both the true and false positive rate of classification also depend on the threshold (quantile of the posterior distribution of  $\epsilon$ ) used for classifying a variant as more transmissible, which can easily be modified by the user depending on the relative costs of false negatives and false positives.

We emphasise that our method estimates the *effective* transmission advantage, which will often reflect a combination of several factors such as a true increase in underlying transmissibility and the ability of a new variant to escape immunity. Disentangling these effects is particularly challenging in the context of changing population immunity e.g., due to vaccination roll-out, and may require additional data (see Supplementary Database for details). However, regardless of its drivers, early identification of a transmission advantage is a critical first step to a timely response.

We note that in the extreme case where estimates are made independently at all time steps and locations, our method reduces to what we have called the “non-parametric” approach as it provides a non-parametric estimate of the transmission advantage at each time and location as the ratio between the reproduction numbers for each variant. While such an approach can help in initial exploration, the assumption of independence across time and space can lead to

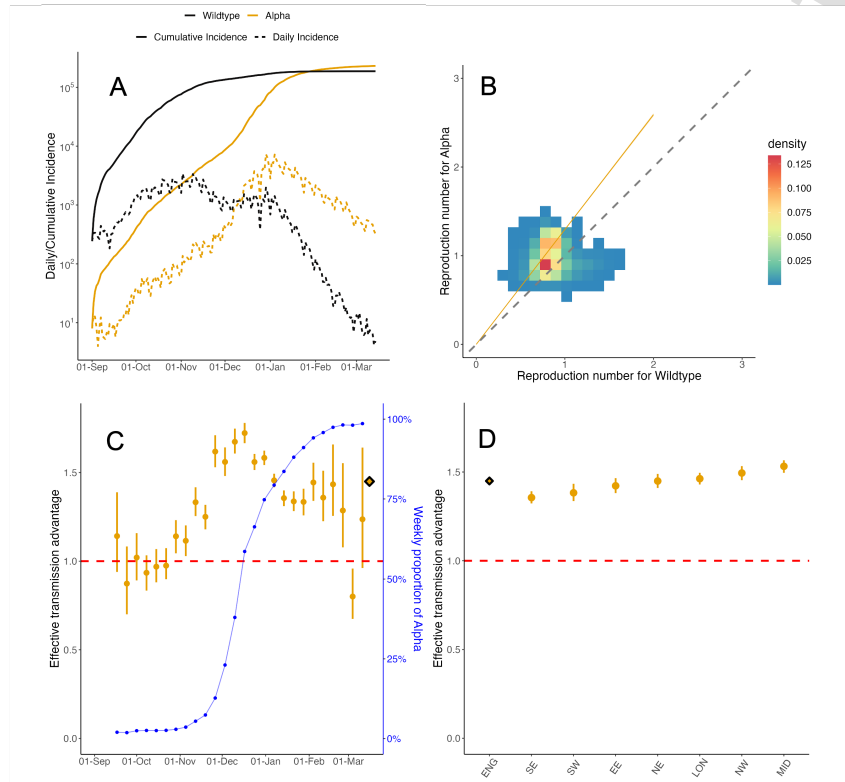
highly uncertain estimates. MV-EpiEstim allows combining information across time and/or locations, assuming that the effective transmission advantage is constant across these. This allows reducing the uncertainty in the estimates. Temporal or spatial heterogeneity in the transmission advantage (e.g. reflecting heterogeneity in population immunity) can also be characterised by applying the method separately by location or time period, which is easy to do in our software.

Our estimated transmission advantage of the SARS-CoV-2 Alpha variant (over the wildtype) is consistent with those from other analyses identified in our literature review, with estimates in the range of 1.02 to 2.30 (See Supplementary Database). Similarly, our estimates of the transmission advantage of the Delta variant over the Alpha variant are broadly consistent with the literature, with central estimates of the advantage in the range of 1.5 to 2.4. In addition, our results highlight temporal changes in the transmission advantage which were overlooked by some of these studies. The agreement of our findings with those from other studies employing a diversity of modelling approaches including renewal equations, semi-mechanistic models, and phylodynamic models suggests that MV-EpiEstim can be a useful tool for early characterisation of new variants. Importantly, where specific bio-markers are sufficient to distinguish variants (e.g. S-gene), MV-EpiEstim does not need any whole-genome sequencing data. Therefore, it could be used in near real-time, relying only on routinely collected incidence data and not necessarily suffering from potential delays in the sequencing pipeline.

Given the continued transmission of SARS-CoV-2 and low vaccination coverage globally [39], new variants are likely to continue emerging. Our tool can be used to monitor their transmissibility and rapidly identify variants of concern. Our estimates of the transmission advantage of Delta have been used to inform UK national policy in real-time.

Applications of our work are not limited to SARS-CoV-2; our generic method could easily be used to monitor other pathogens with multiple co-circulating strains such as influenza or *streptococcus pneumoniae*.

## Figures



**Figure 1: Effective transmission advantage of Alpha over wildtype in England** (A) The daily reported incidence of cases of the wildtype (black) and Alpha (yellow) in England from September 2020 to March 2021. (B) The effective reproduction number  $R_t$  estimated independently for the wildtype (x-axis) and Alpha (y-axis) on sliding weekly windows. The colour of the cells indicates the density of the draws from the respective posterior distributions of  $R_t$ . The dashed diagonal line indicates the  $x = y$  threshold. Coloured cells lying above the diagonal line suggest that Alpha is more transmissible. The yellow line denotes the median effective transmission advantage estimated using MV-EpiEstim, assuming no temporal or spatial heterogeneity. 95% CrI were so narrow that they could not be distinguished from the line. (C) Effective transmission advantage estimated using MV-EpiEstim using data in the week ending on the date specified on the x-axis (yellow) and the entire time series (diamond). The dark blue circles and the vertical bars denote respectively the mean and 95% binomial confidence interval of the proportion of incidence of Alpha (right y-axis) in the week of estimation. Because of the high incidence of both wildtype and Alpha in the study period, the 95% CI are small and hence difficult to distinguish. (D) Effective transmission advantage estimated using MV-EpiEstim for all NHS England regions together (diamond) and separately (solid circles), using data from 1<sup>st</sup> September 2020 to 14<sup>th</sup> March 2021. The NHS England regions are - East of England<sub>1</sub>(EE), London (LON), Midlands (MID), North-East (NE), North-West (NW), South-East (SE), South-West (SW). In panels (C) and (D), the solid yellow circles denote the median estimate, the vertical lines indicate the 95% CrI, and the red dashed line denotes the  $\epsilon = 1$  threshold.



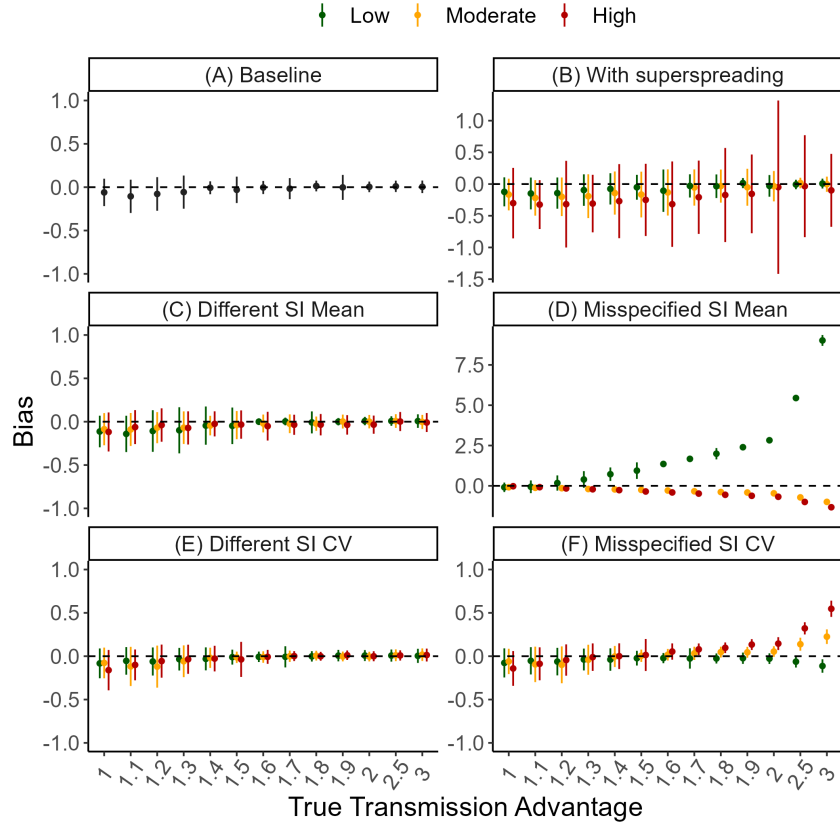


Figure 2: **Method performance on simulated data.** We assessed the performance of MV-EpiEstim on a range of scenarios. In each panel, the x-axis shows the true value of the effective transmission advantage (on categorical scale). The y-axis shows the bias i.e., the difference between the posterior mean estimate of the transmission advantage and the true value. The solid dots represent the mean bias (across 100 simulations) and the vertical bars show the standard deviation (SD) of the bias. Each panel corresponds to a different simulation scenario. In all scenarios, the  $R_t$  for the reference variant was 1.1 and the  $R_t$  for the new variant was  $\epsilon$  times the reference  $R_t$  (see Suppl Sec. 5 for details). (A) In the baseline scenario, we assumed no superspreading and the same natural history for both variants. (B) As (A), but with low (overdispersion parameter  $\kappa = 1$ ), moderate ( $\kappa = 0.5$ ) and high ( $\kappa = 0.1$ ) levels of superspreading. (C) As (A), but the mean serial interval of the new variant is 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference and is correctly specified during estimation. (D) As (C), but the mean serial interval of both variants are assumed to be the same during estimation. (E) As (A), but the coefficient of variation (CV, ratio of standard deviation to mean) of the serial interval distribution of the new variant is 0.5 (low), 1.5 (moderate) or 2 (high) times that of the reference and is correctly specified during estimation. (F) As (E), but the CV of the serial interval distribution of the reference and new variants are assumed to be the same during estimation. Note that the y-axis range is different for panel D. Results using  $R_t = 1.6$  for the reference and using fewer days of data are presented in Suppl Secs. 5.3 to 5.9. Results using time-varying reference  $R_t$  in one or two locations are shown in Suppl Secs. 5.10 and 5.11.

## References

- [1] *Prime Minister's statement on coronavirus (COVID-19): 19 December 2020*. Dec. 2020. URL: <https://www.gov.uk/government/speeches/prime-ministers-statement-on-coronavirus-covid-19-19-december-2020> (visited on 11/08/2021).
- [2] *PM statement at coronavirus press conference: 14 June 2021*. June 2021. URL: <https://www.gov.uk/government/speeches/pm-statement-at-coronavirus-press-conference-14-june-2021> (visited on 10/25/2021).
- [3] *Tracking SARS-CoV-2 Variants*. 2021. URL: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> (visited on 10/25/2021).
- [4] R. Sonabend et al. "Non-Pharmaceutical Interventions, Vaccination, and the SARS-CoV-2 Delta Variant in England: A Mathematical Modelling Study". In: *The Lancet* (Oct. 2021). DOI: 10.1016/S0140-6736(21)02276-5.
- [5] A. Rambaut et al. *Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations - SARS-CoV-2 Coronavirus / nCoV-2019 Genomic Epidemiology*. Dec. 2020. URL: <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (visited on 07/12/2022).
- [6] E. Volz et al. "Assessing transmissibility of SARS-CoV-2 lineage B. 1.1.7 in England". In: *Nature* 593.7858 (2021), pp. 266–269. DOI: 10.1038/s41586-021-03470-x.
- [7] N. G. Davies et al. "Estimated Transmissibility and Impact of SARS-CoV-2 Lineage B.1.1.7 in England". In: *Science* 372.6538 (Apr. 2021). DOI: 10.1126/science.abg3055.
- [8] K. Leung et al. "Early Transmissibility Assessment of the N501Y Mutant Strains of SARS-CoV-2 in the United Kingdom, October to November 2020". In: *Eurosurveillance* 26.1 (Jan. 2021). DOI: 10.2807/1560-7917.ES.2020.26.1.2002106.
- [9] W. Yang and J. Shaman. "Development of a Model-Inference System for Estimating Epidemiological Characteristics of SARS-CoV-2 Variants of Concern". In: *Nature Communications* 12.1 (Dec. 2021), p. 5573. DOI: 10.1038/s41467-021-25913-9.
- [10] S. Zhao et al. "The Co-Circulating Transmission Dynamics of SARS-CoV-2 Alpha and Eta Variants in Nigeria: A Retrospective Modeling Study of COVID-19". In: *Journal of Global Health* 11 (Dec. 2021), p. 05028. DOI: 10.7189/jogh.11.05028.

- [11] S. Zhao et al. “Quantifying the Transmission Advantage Associated with N501Y Substitution of SARS-CoV-2 in the UK: An Early Data-Driven Analysis”. In: *Journal of Travel Medicine* 28.2 (Jan. 2021), taab011. DOI: 10.1093/jtm/taab011.
- [12] M. Chand et al. *Investigation of novel SARS-COV-2 variant: Variant of Concern 20212/01*. 2021. URL: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/959438/Technical\\_Briefing\\_VOC\\_SH\\_NJL2\\_SH2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf) (visited on 09/01/2021).
- [13] M. S. Graham et al. “Changes in Symptomatology, Reinfection, and Transmissibility Associated with the SARS-CoV-2 Variant B.1.1.7: An Ecological Study”. In: *The Lancet Public Health* 6.5 (May 2021), e335–e345. DOI: 10.1016/S2468-2667(21)00055-4.
- [14] C. Piantham et al. *Estimating the Elevated Transmissibility of the B.1.1.7 Strain over Previously Circulating Strains in England Using GISAID Sequence Frequencies*. medRxiv. Mar. 2021. DOI: 10.1101/2021.03.17.21253775.
- [15] *SARS-CoV-2 variants of concern and variants under investigation in England: Technical Briefing 15*. 2021. URL: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/993879/Variants\\_of\\_Concern\\_VOC\\_Technical\\_Briefing\\_15.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993879/Variants_of_Concern_VOC_Technical_Briefing_15.pdf) (visited on 09/01/2021).
- [16] N. M. Ferguson. *B.1.617.2 Transmission in England: Risk Factors and Transmission Advantage*. June 2021. URL: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/993159/S1270\\_IMPERIAL\\_B.1.617.2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993159/S1270_IMPERIAL_B.1.617.2.pdf) (visited on 07/11/2022).
- [17] S. Alizon et al. “Rapid Spread of the SARS-CoV-2 Delta Variant in Some French Regions, June 2021”. In: *Eurosurveillance* 26.28 (July 2021). DOI: 10.2807/1560-7917.ES.2021.26.28.2100573.
- [18] F. Campbell et al. “Increased Transmissibility and Global Spread of SARS-CoV-2 Variants of Concern as at June 2021”. In: *Eurosurveillance* 26.24 (June 2021). DOI: 10.2807/1560-7917.ES.2021.26.24.2100509.
- [19] M. J. Keeling. *Estimating the Transmission Advantage for B.1.617.2*. 2021. URL: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/993156/S1269\\_WARWICKTransmission\\_Advantage.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993156/S1269_WARWICKTransmission_Advantage.pdf) (visited on 09/15/2021).
- [20] W. Yang et al. “Epidemiological Characteristics of the B.1.526 SARS-CoV-2 Variant”. In: *Science Advances* 8.4 (Jan. 2022), eabm0300. DOI: 10.1126/sciadv.abm0300.
- [21] R. M. Coutinho et al. “Model-Based Estimation of Transmissibility and Reinfection of SARS-CoV-2 P.1 Variant”. In: *Communications Medicine* 1.1 (Dec. 2021), p. 48. DOI: 10.1038/s43856-021-00048-6.

- [22] N. R. Faria et al. “Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil”. In: *Science* 372.6544 (May 2021), pp. 815–821. DOI: 10.1126/science.abh2644.
- [23] C. H. van Dorp et al. “Estimating the strength of selection for new SARS-CoV-2 variants”. In: *Nature Communications* 12.1 (Dec. 14, 2021), p. 7239. DOI: 10.1038/s41467-021-27369-3.
- [24] C. H. van Dorp et al. “Global estimates of the fitness advantage of SARS-CoV-2 variant Omicron”. In: *Virus Evolution* 8.2 (Oct. 20, 2022), veac089. DOI: 10.1093/ve/veac089.
- [25] M. D. Figgins and T. Bedford. *SARS-CoV-2 variant dynamics across US states show consistent differences in effective reproduction numbers*. medRxiv. Dec. 11, 2021. DOI: 10.1101/2021.12.09.21267544.
- [26] A. Cori et al. “A New Framework and Software to Estimate Time-Varying Reproduction Numbers during Epidemics”. In: *American Journal of Epidemiology* 178.9 (2013), pp. 1505–1512. DOI: 10.1093/aje/kwt133.
- [27] R. Thompson et al. “Improved inference of time-varying reproduction numbers during infectious disease outbreaks”. In: *Epidemics* 29 (2019), p. 100356.
- [28] A. Cori. *EpiEstim*. Dec. 2021. URL: <https://github.com/mrc-ide/EpiEstim> (visited on 11/16/2021).
- [29] A. Gelman and D. B. Rubin. “Inference from Iterative Simulation Using Multiple Sequences”. In: *Statistical Science* 7.4 (Nov. 1992). DOI: 10.1214/ss/1177011136.
- [30] *COVID-19 Point épidémiologique hebdomadaire du 28 janvier 2021*. URL: <https://www.santepubliquefrance.fr/content/download/315275/2903017> (visited on 10/25/2021).
- [31] *COVID-19 Point épidémiologique hebdomadaire du 4 mars 2021*. URL: <https://www.santepubliquefrance.fr/content/download/324805/2944195> (visited on 10/25/2021).
- [32] *Tableau synthétique des résultats par vague d’enquêtes*. URL: <https://www.santepubliquefrance.fr/content/download/368440/3132368> (visited on 10/25/2021).
- [33] *EpiEstim package - RDocumentation*. 2021. URL: <https://www.rdocumentation.org/packages/EpiEstim/versions/2.2-4> (visited on 09/19/2021).
- [34] M. U. G. Kraemer et al. “Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence”. In: *Science* 373.6557 (2021), pp. 889–895. DOI: 10.1126/science.abj0113.

- [35] R. Hinch et al. “Estimating SARS-CoV-2 variant fitness and the impact of interventions in England using statistical and geo-spatial agent-based models”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380.2233 (Aug. 2022). Publisher: Royal Society, p. 20210304. DOI: 10.1098/rsta.2021.0304. URL: <https://royalsocietypublishing.org/doi/10.1098/rsta.2021.0304> (visited on 04/05/2023).
- [36] F. Blanquart et al. “Selection for infectivity profiles in slow and fast epidemics, and the rise of SARS-CoV-2 variants”. In: *eLife* 11 (May 2022). Ed. by B. S. Cooper and M. P. Davenport. Publisher: eLife Sciences Publications, Ltd, e75791. ISSN: 2050-084X. DOI: 10.7554/eLife.75791. URL: <https://doi.org/10.7554/eLife.75791> (visited on 04/05/2023).
- [37] K. Ito et al. “Estimating relative generation times and reproduction numbers of Omicron BA.1 and BA.2 with respect to Delta variant in Denmark”. en. In: *Mathematical Biosciences and Engineering* 19.9 (2022). Cc.license.type: cc.by Number: mbe-19-09-418 Primary\_atype: Mathematical Biosciences and Engineering Subject\_term: Research article Subject\_term\_id: Research article, pp. 9005–9017. ISSN: 1551-0018. DOI: 10.3934/mbe.2022418. URL: <http://www.aimspress.com/article/doi/10.3934/mbe.2022418> (visited on 04/05/2023).
- [38] J. A. Backer et al. “Shorter serial intervals in SARS-CoV-2 cases with Omicron BA.1 variant compared with Delta variant, the Netherlands, 13 to 26 December 2021”. en. In: *Eurosurveillance* 27.6 (Feb. 2022). Publisher: European Centre for Disease Prevention and Control, p. 2200042. ISSN: 1560-7917. DOI: 10.2807/1560-7917.ES.2022.27.6.2200042. URL: <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2022.27.6.2200042> (visited on 04/05/2023).
- [39] E. Mathieu et al. “A Global Database of COVID-19 Vaccinations”. In: *Nature Human Behaviour* 5.7 (July 2021), pp. 947–953. DOI: 10.1038/s41562-021-01122-8.

## Acknowledgements

The use of pillar-2 PCR testing data was made possible thanks to PHE colleagues, and we extend our thanks to N Gent for facilitation and insights into these data. We also thank Edward S Knock for his inputs on the data for England. This study is partially funded by the National Institute for Health Research (NIHR) Health and Care Protection Research Unit in Modelling and Health Economics, a partnership between Public Health England, Imperial College London and LSHTM (grant code NIHR200908); the authors acknowledge funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), which is jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2

programme supported by the European Union. JW acknowledges research funding from the Wellcome Trust (grant 102169/Z/13/Z). SB acknowledges funding from the Wellcome Trust (grant 219415). RKN acknowledges funding from the Medical Research Council Doctoral Training Partnership (MR/N014103/1). AC was supported by the Academy of Medical Sciences Springboard scheme, funded by the AMS, Wellcome Trust, BEIS, the British Heart Foundation and Diabetes UK [REF:SBF005\1044] Disclaimer: The views expressed are those of the author(s) and not necessarily those of the NIHR, Public Health England or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

- Our method estimates the effective transmission advantage ( $\epsilon$ ) of pathogen variants
- We show the method performs well across a range of simulated scenarios
- Our  $\epsilon$  estimates for SARS-CoV-2 variants are similar to other studies
- We provide an easy-to-use implementation of our method in the R package EpiEstim

Journal Pre-proof

SB: Conceptualization; Software; Validation; Formal analysis; Writing - Original Draft; Writing - Review & Editing.

JW: Conceptualization; Software; Validation; Formal analysis; Writing - Review and Editing.

RKN: Conceptualization; Software; Validation; Formal analysis; Writing - Review and Editing.

PN: Conceptualization; Methodology; Software; Formal analysis; Writing - Review and Editing; Supervision

AC: Conceptualization; Methodology; Software; Formal analysis; Writing - Original Draft; Writing - Review and Editing; Supervision

Journal Pre-proof



**Conflicts of interest**

AC has received payment from Pfizer for teaching of mathematical modelling of infectious diseases.

*Journal Pre-proof*