



Differences between the true reproduction number and the apparent reproduction number of an epidemic time series

Oliver Eales^{a,b,c,d,*}, Steven Riley^{b,c,d}

^a Infectious Disease Dynamics Unit, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia

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

^c MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, United Kingdom

^d Abdul Latif Jameel Institute for Disease and Emergency Analytics, Imperial College London, London, United Kingdom

ARTICLE INFO

Keywords:

Reproduction number
Epidemics
SARS-CoV-2
Pandemics
COVID-19
Infection prevalence
Disease surveillance
Testing

ABSTRACT

The time-varying reproduction number $R(t)$ measures the number of new infections per infectious individual and is closely correlated with the time series of infection incidence by definition. The timings of actual infections are rarely known, and analysis of epidemics usually relies on time series data for other outcomes such as symptom onset. A common implicit assumption, when estimating $R(t)$ from an epidemic time series, is that $R(t)$ has the same relationship with these downstream outcomes as it does with the time series of incidence. However, this assumption is unlikely to be valid given that most epidemic time series are not perfect proxies of incidence. Rather they represent convolutions of incidence with uncertain delay distributions. Here we define the apparent time-varying reproduction number, $R_A(t)$, the reproduction number calculated from a downstream epidemic time series and demonstrate how differences between $R_A(t)$ and $R(t)$ depend on the convolution function. The mean of the convolution function sets a time offset between the two signals, whilst the variance of the convolution function introduces a relative distortion between them. We present the convolution functions of epidemic time series that were available during the SARS-CoV-2 pandemic. Infection prevalence, measured by random sampling studies, presents fewer biases than other epidemic time series. Here we show that additionally the mean and variance of its convolution function were similar to that obtained from traditional surveillance based on mass-testing and could be reduced using more frequent testing, or by using stricter thresholds for positivity. Infection prevalence studies continue to be a versatile tool for tracking the temporal trends of $R(t)$, and with additional refinements to their study protocol, will be of even greater utility during any future epidemics or pandemics.

1. Introduction

Infectious disease epidemics are a major threat to public health. Over the past two decades there has been major outbreaks of SARS (Anderson et al., 2004), MERS (Sharif-Yakan and Kanj, 2014), influenza (Gog et al., 2014), Ebola (Gire et al., 2014), dengue (MacCormack-Gelles et al., 2018) and most recently the emergence of SARS-CoV-2 (Li et al., 2021b), which resulted in the COVID-19 pandemic. Accurately quantifying the transmission dynamics of infection is crucial for informing decisions about changes to policy on public health interventions.

The reproduction number, the expected number of secondary infections resulting from a typical primary infection, is a key quantity

describing a pathogen's transmission dynamics. If public health interventions are to be implemented, $R(t)$ can inform on the magnitude of the interventions required to bring the epidemic under control (Anderson and May, 1992). Analysis of how $R(t)$ changes over time is also important for assessing the impact of interventions including vaccination and social distancing (Li et al., 2021a; Eales et al., 2022e).

Mathematically, $R(t)$ is linked to the time series of infection incidence (the rate of new infections) by a renewal process. If the infection incidence and the generation time (the time between a primary and secondary infection) distribution are known, then $R(t)$ can simply be calculated (Wallinga and Lipsitch, 2007). However, the timings of infections are rarely known, and so accurate time series of infection incidence rarely exist. Estimates of $R(t)$ must instead rely on

* Corresponding author at: Infectious Disease Dynamics Unit, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia.

E-mail addresses: oliver.eales@unimelb.edu.au (O. Eales), s.riley@imperial.ac.uk (S. Riley).

<https://doi.org/10.1016/j.epidem.2024.100742>

Received 10 July 2023; Received in revised form 21 December 2023; Accepted 11 January 2024

Available online 13 January 2024

1755-4365/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

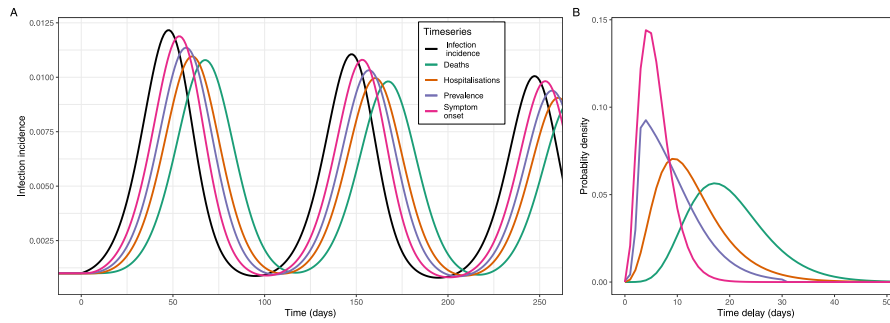


Fig. 1. An example of the relationship between infection incidence and other epidemic time series. (A) An example time series of infection incidence (proportion of population infected on a given day), $I(t)$ (black), and four other epidemic time series, $J(t)$ (see Section 2.2): deaths (green), hospitalisations (orange), infection prevalence (purple), and symptom onset (pink) (e.g. cases identified through symptomatic testing). (B) The probability distributions of the convolution functions linking infection incidence to the four epidemic time series. The convolution functions were chosen to match analytical estimates found for SARS-CoV-2 epidemic time series (see Section 6.3, Convolution functions for SARS-CoV-2 epidemic time series). Note that we have used the probability distributions of the convolution functions (which add up to one) so that the time series are all of comparable heights to infection incidence in panel (A). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

other epidemic time series (e.g., daily number of cases, hospitalisations or deaths) that are often treated as proxies for the infection incidence (Nash et al., 2022).

During an epidemic there are often many sources of time series data collected. These can include measurements of the frequency of downstream points in the natural history of infection such as onset of symptoms, hospitalisation, and death (UK Coronavirus, 2022). These can sometimes include measurements of infection prevalence (the proportion of the population testing positive for the pathogen) made using random sampling. This is a relatively new surveillance method with relatively few studies (all occurring during the SARS-CoV-2 pandemic) having collected time series data for infection prevalence (Elliott et al., 2022; Chadeau-Hyam et al., 2022). Infection prevalence time series are often less biased than other epidemic time series, which can be biased by changes in behaviour (health-care seeking behaviour, test-seeking behaviour, etc.) or in severity of the virus (e.g., due to vaccination). However, there are concerns that such data would be ill-suited for estimating $R(t)$ due to the presence of long-term shedders — individuals who continue to test positive for a long duration of time, following infection.

In general, an epidemic time series is linked to the time series of infection incidence by a convolution (Fig. 1); if the function for the convolution is known then it is possible to estimate the time series of incidence, and from that $R(t)$ (Huisman et al., 2022). However, in many instances the form of this function is not known accurately (or at all) and so $R(t)$ cannot be straightforwardly calculated. In these instances, it is common practice for the epidemic time series to be treated as a reasonable proxy for the infection incidence and estimates of $R(t)$ are obtained by assuming the same renewal process between downstream points in the natural history of infection, for example from symptom onset to symptom onset (Shim et al., 2021). This assumption is often made even when the convolution function is known (or could be estimated), despite the existence of computer packages that allow the inclusion of a convolution function when estimating $R(t)$ (Abbott et al., 2020; Lytras, 2020). Including examples of estimating $R(t)$ from case time series (as a proxy for infection incidence) for influenza (Ali et al., 2013; Thompson et al., 2019), Ebola (Thompson et al., 2019; International Ebola Response Team et al., 2016; Kraemer et al., 2020), MERS (Thompson et al., 2019; Park et al., 2016), SARS-CoV-2 (Dighe et al., 2020; Shim et al., 2021), Cholera (Zhao et al., 2019) and Zika (Perkins et al., 2019). It is important that we understand what biases are introduced when estimates of $R(t)$ are made under the assumption that an epidemic time series is a reasonable proxy for infection incidence so that future analyses can be properly interpreted.

There are other potential biases in epidemic time series which may influence estimates of $R(t)$. However, biases introduced by assuming

the epidemic time series to be a perfect proxy for infection incidence have often been overlooked. Here we investigate the form that these biases take and how they depend on the relationship between an epidemic time series and the infection incidence. We further investigate how estimates of $R(t)$ made during the SARS-CoV-2 pandemic may have been biased by the epidemic time series that were available (onset of symptoms, infection prevalence, hospitalisations and deaths considered), and how data could be better collected (using frequent testing and stricter viral thresholds to define test positivity) and used to minimise these biases.

2. Background

2.1. The relationship between incidence and $R(t)$

The reproduction number can be written as

$$R = \int \eta(\tau) d\tau, \quad (1)$$

where $\eta(\tau)$ describes the rate of secondary infections as a function of the time since a primary infection, τ . Normalising $\eta(\tau)$ gives the generation time,

$$g(\tau) = \frac{\eta(\tau)}{\int \eta(\tau) d\tau}. \quad (2)$$

the probability distribution describing the time between primary and secondary infections. The incidence at time t , $I(t)$ can then be written in terms of the incidence at times less than t :

$$\begin{aligned} I(t) &= \int_0^\infty I(t-\tau) \eta(\tau) d\tau \\ &= \int_0^\infty I(t-\tau) g(\tau) R(t) d\tau. \end{aligned} \quad (3)$$

Where $R(t)$ is the reproduction number at time t . Expressed in this way the reproduction number at time t can be calculated directly from the time-series of incidence (if the generation time function is known — which we will assume throughout) through the equation:

$$R(t) = \frac{I(t)}{\int_0^\infty I(t-\tau) g(\tau) d\tau}. \quad (4)$$

2.2. The relationship between epidemic time series and $R(t)$

A general epidemic time-series, $J(t)$, can be written in terms of the incidence at times less than t and a convolution function, $f(\zeta)$, that describes the relationship between the two time series:

$$J(t) = \int_0^\infty I(t-\zeta) f(\zeta) d\zeta. \quad (5)$$

The convolution function will take different forms depending on the epidemic time series. For example, for the time series of deaths due to an infectious disease, $f(\zeta)$ would describe the average rate of deaths as a function of the time since infection. Substituting the relationship in Eq. (3) into Eq. (5) we obtain:

$$J(t) = \int_0^\infty \int_0^\infty I(t - \zeta - \tau)g(\tau)R(t - \zeta)f(\zeta)d\zeta d\tau. \quad (6)$$

In general there is not a simple relationship between $J(t)$ and $R(t)$. To calculate $R(t)$ from $J(t)$ the convolution function, $f(\zeta)$, would need to be known (and of course $g(\tau)$). However, it is often assumed that the same relationship exists between $J(t)$ and $R(t)$, as the one between $I(t)$ and $R(t)$:

$$J(t) = \int_0^\infty J(t - \tau)g(\tau)R_A(t)d\tau. \quad (7)$$

Here $R_A(t)$ is the apparent reproduction number inferred from a general epidemic time series under this incorrect assumption, and is given by:

$$R_A(t) = \frac{J(t)}{\int_0^\infty J(t - \tau)g(\tau)d\tau}. \quad (8)$$

In general $R_A(t) \neq R(t)$, but under certain conditions relationships between the two will hold.

$R(t) = \text{Constant}$

If $R(t)$, has been constant for the period of time over which $f(\zeta)$ goes to 0 then Eq. (6) can be written as:

$$J(t) = \int_0^\infty J(t - \tau)g(\tau)\overline{R(t)}d\tau. \quad (9)$$

Where we have written $R(t)$ as $\overline{R(t)}$ to make it clear this equation is only valid when $R(t)$ is constant. In this situation the relationship between $J(t)$ and $R(t)$ is the same as the relationship between $I(t)$ and $R(t)$ and we have

$$R_A(t) = \overline{R(t)}. \quad (10)$$

Note that the period over which $f(\zeta)$ goes to zero will depend on the epidemic time series and pathogen in question (for SARS-CoV-2 example, see Fig. 1B). Hence Eq. (10) will be valid more often for epidemic time series in which $f(\zeta)$ goes to zero quickly ($R(t)$ does not have to be constant for as long a period for the equation to be valid)

$f(\tau) = \delta(\mu - \tau)$

If the convolution function, $f(\zeta)$, is the Dirac-delta function with mean μ , $\delta(\mu - \zeta)$, then $J(t)$ can be written as:

$$J(t) = I(t - \mu). \quad (11)$$

$R_A(t)$ can then be written as:

$$\begin{aligned} R_A(t) &= \frac{J(t)}{\int_0^\infty J(t - \tau)g(\tau)d\tau} \\ &= \frac{I(t - \mu)}{\int_0^\infty I(t - \mu - \tau)g(\tau)d\tau} \\ &= R(t - \mu). \end{aligned} \quad (12)$$

Note that it would be very unlikely for the convolution function linking infection incidence to an epidemic time series to be a Dirac-delta function; most convolution functions would be expected to have some variance. However, the less variance a convolution function has the closer an approximation it will be to a Dirac-delta function.

3. Results

3.1. The effect of convolution on estimates of $R(t)$

Infection incidence was simulated for two time series of $R(t)$: sinusoidal and square waves. A sinusoidal wave describes gradual changes

in $R(t)$; gradual changes in $R(t)$ would be expected over the course of an epidemic (due to the depletion of susceptibles). A square wave describes rapid changes in $R(t)$, which would be expected when some public health interventions (e.g., lockdowns, school closures) are implemented. Epidemic time series and their associated $R_A(t)$ were then calculated for different gamma distributed convolution functions (see Materials and Methods) (Fig. 2).

As we would expect, the greater the mean of the convolution function, the greater the time lag between $R_A(t)$ and $R(t)$. When $R(t)$ changed gradually (simulated as a sine function in time) (Fig. 2b) there was a delay between $R(t)$ and $R_A(t)$ reaching their respective maximums (Fig. 3c). This time delay was approximately equal to the mean of the convolution function. When $R(t)$ underwent a step decrease (Fig. 2a) there was a delay between $R(t)$ decreasing and $R_A(t)$ decreasing (with similar results for a step increase). Measuring the time delay between $R(t)$ decreasing (a step change) and $R_A(t)$ falling to under 95% of its maximum value (Fig. 3a) we observed a greater delay for convolution functions with greater means and lower standard deviations.

Trends in $R_A(t)$ (the shape of the curve over time) were also distorted relative to $R(t)$, with increases in the standard deviation of the convolution function resulting in greater levels of distortion. When $R(t)$ changed gradually (simulated as a sine function in time) the maximum value of $R_A(t)$ was lower than the maximum value of $R(t)$. The greater the standard deviation of the convolution function the greater the relative reduction in the maximum value of $R_A(t)$ (Fig. 3d). When $R(t)$ underwent a sharp decrease (a step change over a single day) $R_A(t)$ decreased at a slower rate (Fig. 2a). Measuring the time delay between $R_A(t)$ falling from 95% of its maximum value to 5% of its maximum value we observed a greater delay when the standard deviation of the convolution function was greater (Fig. 3b). Note that if the standard deviation of the convolution function was zero (Dirac-delta function) then trends in $R_A(t)$ would not be distorted and would only be lagged by the mean of the convolution function (see Section 2.2). However, this is unlikely to be the case for any convolution function linking infection incidence to an epidemic time series.

3.2. Case study: SARS-CoV-2 epidemic time series

We simulated four epidemic time series that would have been available during the SARS-CoV-2 pandemic: the daily number of deaths; the daily number of hospitalisations; the number of individuals reporting symptom onset; and the infection prevalence (see Materials and Methods). We simulated a representative scenario for trends in $R(t)$ (chosen to resemble trends in $R(t)$ for SARS-CoV-2 in England from March to December 2020) and calculated $R_A(t)$ for each simulated epidemic time series. For all epidemic time series there were clear differences between $R_A(t)$ and $R(t)$ (Fig. 4). During periods when $R(t)$ was constant, the value of $R_A(t)$ was the same (as expected). When $R(t)$ changed, $R_A(t)$ lagged $R(t)$ and had its shape distorted in a way that was dependent on the shape of the convolution function.

The values of $R_A(t)$ calculated from the simulated time series of symptom onset most closely resembled the underlying $R(t)$. The mean and standard deviation of the convolution function for symptom onset was lower than all other time series considered (Fig. 4b). This was expected for deaths and hospitalisations, for which the convolution functions were composed of the distributions of time from infection to symptom onset (incubation period) and time from symptom onset to outcome (death/hospitalisation) (see Methods). Interestingly, the peak probability density occurred at the same time delay for the convolution functions for infection prevalence and symptom onset. However, the mean and standard deviation of the convolution function for infection prevalence were greater (than those of the convolution function for symptom onset) due to a long tail in the probability distribution.

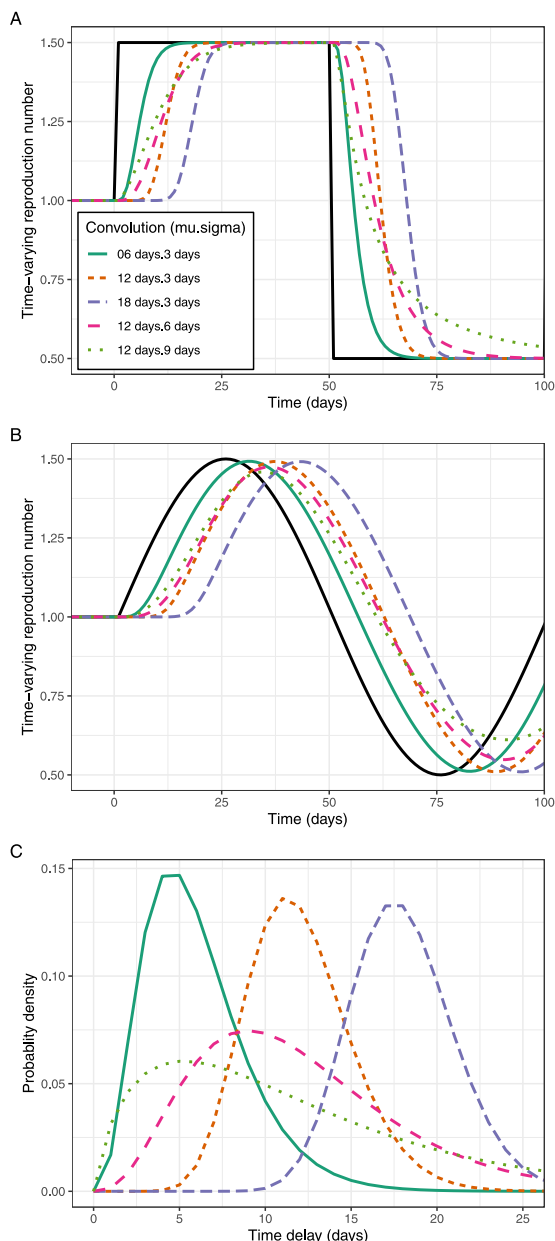


Fig. 2. Estimates of the time-varying reproduction number from epidemic time series with different gamma distributed convolution functions. (A, B) The actual reproduction number, $R(t)$ (black), and the apparent reproduction number, $R_A(t)$ (coloured), for epidemic time series with different gamma distributed convolution functions. (C) The gamma distributed convolution functions used in (A) and (B). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Improving inference using frequent testing

Increasing the frequency of random/asymptomatic testing reduced the mean and standard deviation of the convolution function (Fig. 5c). Random testing or asymptomatic testing for SARS-CoV-2 should, in the absence of repeat testing, have a convolution function describing the probability of testing positive following infection. In our simulations, increasing the frequency of testing decreased the mean convolution function, as infections were more likely to be identified earlier. Testing once every three days gave a convolution function with a lower mean than that of symptom onset. Testing every day resulted in a convolution function with mean and standard deviation of approximately 3 days and 1 day respectively. The lower mean and standard deviation

of the convolution function resulted in $R_A(t)$ estimates more closely resembling $R(t)$ as expected (Fig. 5a and b).

3.4. Improving inference from infection prevalence studies

Imposing a stricter viral threshold for positivity reduced the standard deviation of the convolution function (Fig. 6c). rt-PCR tests, which are often used to test for the presence of SARS-CoV-2, can also provide information on the Ct value (proxy for viral load) of a positive test. By only including individual tests with a Ct value lower (higher viral load) than a certain threshold, it becomes less likely that early-infections or long-term shedders are included within the epidemic time series. This led to substantial decreases in the standard deviation of the convolution function, and smaller decreases in the mean of the convolution function. Accordingly, this led to the shape of $R_A(t)$ more closely resembling that of $R(t)$, while the lag between the two signals stayed approximately constant (Fig. 6a and b). Imposing a stricter threshold reduced the number of individuals classified as positive and therefore the power of any analysis (Fig. 6d).

4. Discussion

During the SARS-CoV-2 pandemic, estimates of the time-varying reproduction number, $R(t)$, were crucial for informing public health interventions. In most counties there were numerous epidemic time series available for estimating $R(t)$. We have described the biases that will be present in estimates of ' $R(t)$ ' when it is assumed that these time series are reliable proxies for infection incidence (and not the convolution of infection incidence with some other function). This assumption is commonly made (Nash et al., 2022), despite methods existing to incorporate estimates of the convolution function into estimates of $R(t)$ (Abbott et al., 2020). We presented the convolution functions representing four different epidemic time series for SARS-CoV-2: date of symptom onset (symptom reporting studies), prevalence (rt-PCR testing of random samples or asymptomatic testing), hospitalisations and deaths. Additionally, the convolution function for case-numbers, obtained from mass testing (available in many countries worldwide), would likely be a linear combination of the convolution functions for prevalence (asymptomatic testing) and symptom onset (symptomatic testing) with an additional delay to test distribution (delay between symptom onset and positive test).

If the exact form of these convolution function was known, then estimates of $R(t)$ could easily be made from the corresponding epidemic time series (Huisman et al., 2022), but this is not always the case. During the early stages of a pandemic there is little epidemic information, and so convolution functions cannot be inferred. During later periods of a pandemic there can still be great uncertainty in estimates of the convolution function. There is also no guarantee that the convolution functions remain constant over time. For example, rates of symptomatic infection can change (effecting mass testing) (Eales et al., 2022b,c), severity can change (effecting hospitalisation and death time series) (Eales et al., 2023), and even the duration an individual remains test positive can change due to new variants (Eales et al., 2022d), or due to vaccination (Kissler et al., 2021). Thus, there are times when only the apparent reproduction number, $R_A(t)$, can be estimated, and so understanding the inherent biases in these estimates is crucial.

Changes in $R_A(t)$ lagged changes in $R(t)$, with the duration of the lag controlled by the mean of the convolution function. If timeliness in $R(t)$ estimates are required (e.g., for pandemic surveillance), then this time lag is a major limitation (though estimates could still be useful for retrospective analysis). The temporal trends in $R_A(t)$ were distorted relative to $R(t)$, with a greater degree of distortion when variance in the convolution function was greater; if there was no variance in the convolution function (standard deviation = 0), $R_A(t)$ would exhibit the

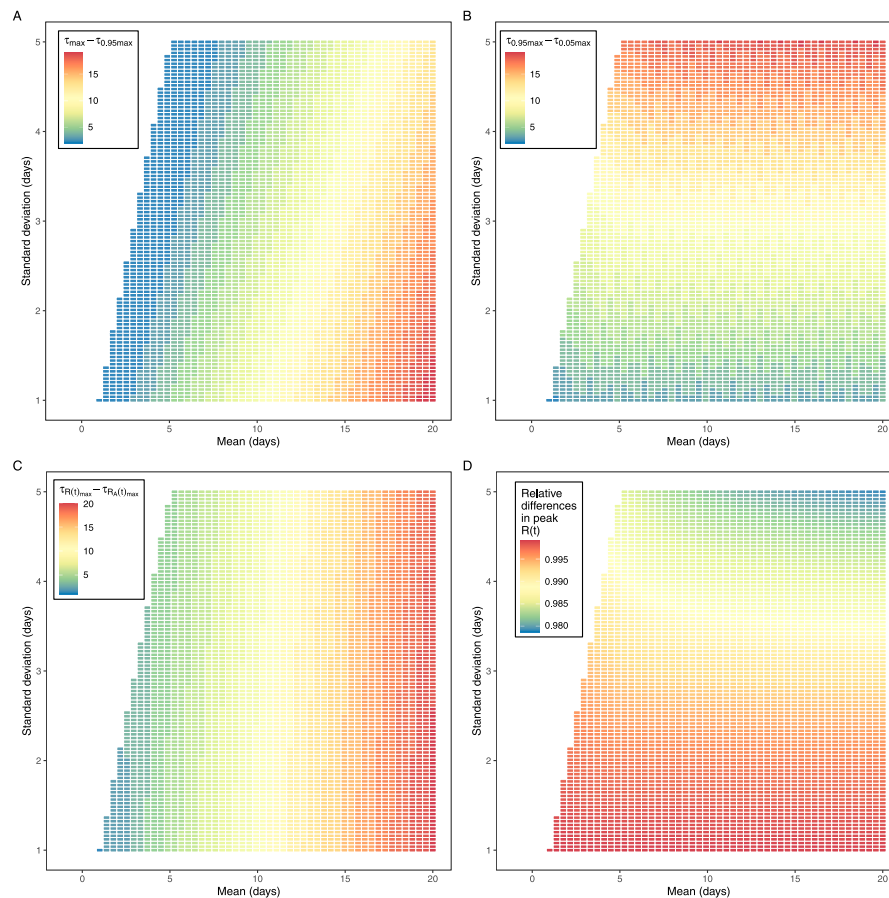


Fig. 3. Diagnostic differences between $R(t)$ and $R_A(t)$. Diagnostic differences between $R(t)$ and $R_A(t)$, for different gamma-distributed convolution functions. Only convolution functions with a mean value greater than their standard deviation are included. (A,B) Differences between $R_A(t)$ and $R(t)$, when $R(t)$ describes a step decrease (see Fig. 2a, day 50). (A) The number of days it takes for $R_A(t)$ to fall below 95% of its maximum value following the step change in $R(t)$. (B) The number of days between $R_A(t)$ falling below 95% of its maximum value and falling below 5% of its maximum value. (C,D) Differences between $R_A(t)$ and $R(t)$, when $R(t)$ describes a sine wave (see Fig. 2b). (C) The number of days between $R(t)$ reaching its maximum value and $R_A(t)$ reaching its maximum value. (D) A measure of the difference between the peak values of $R(t)$ and $R_A(t)$, defined as $1 - \frac{\max(R(t)) - \max(R_A(t))}{\max(R(t)) - \min(R(t))}$ it reflects the proportion of the maximum value that $R_A(t)$ reaches.

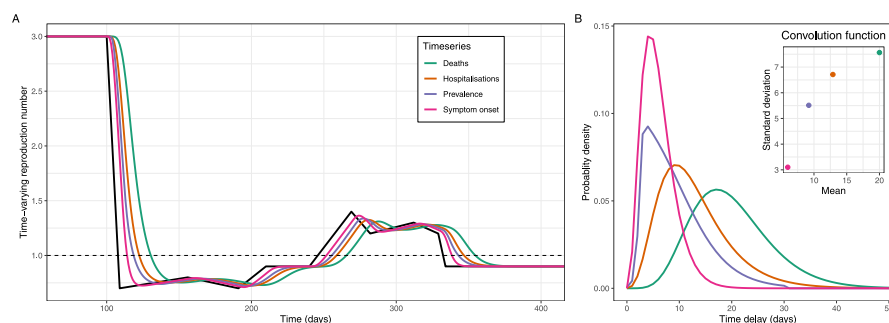


Fig. 4. Trends in $R(t)$ and $R_A(t)$ for representative epidemic time series during the SARS-CoV-2 pandemic. (A) The actual reproduction number, $R(t)$ (black), and the apparent reproduction number, $R_A(t)$ (coloured), for epidemic time series with different convolution functions. $R(t)$ was chosen to resemble a possible trajectory of $R(t)$ in England from May 2020 to December 2020 (B) The probability distribution of the convolution functions for deaths (Green), hospitalisations (Orange), prevalence (Purple), symptom onset (Pink). The mean and standard deviation of all convolution functions are presented in the inset graph. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

exact same temporal trends as $R(t)$ (lagged by the mean of the convolution function). Note that we have only considered the influence of first- and second-order moments (mean and variance) of the convolution function; changes in higher order moments (e.g. skewness) may also lead to greater levels of distortion. Distortion was most pronounced when $R(t)$ changed rapidly, and less pronounced when $R(t)$ changed more gradually. During many epidemics only gradual changes in $R(t)$

are observed (e.g., due to the changing proportion of susceptibility as individuals are infected and develop immunity) and such distortions would be minimal. However, in some instances $R(t)$ may change rapidly, such as when restrictions are implemented (Knock et al., 2021) – for example lock downs during the SARS-CoV-2 pandemic – or when schools are closed/opened (Li et al., 2021a). In these instances, there will be a high degree of distortion between $R(t)$ and $R_A(t)$.

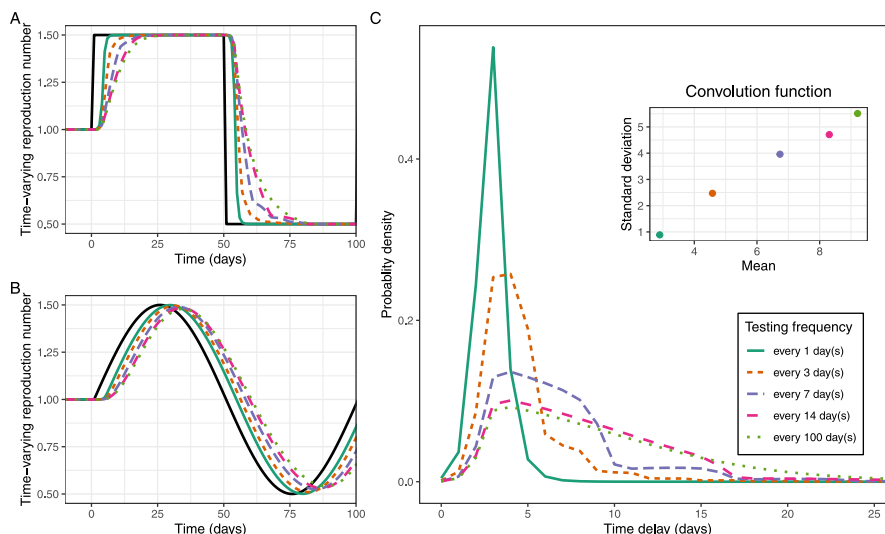


Fig. 5. The effect of increasing testing frequency on estimates of $R_A(t)$. (A,B) The actual reproduction number, $R(t)$ (black), and the apparent reproduction number, $R_A(t)$ (coloured), for epidemic time series with different convolution functions. (C) The probability distribution of convolution functions for different frequencies of rt-PCR testing. Testing frequency of ‘every 100 days’ reflects no repeat testing and has the same convolution function as for prevalence in Fig. 4. The mean and standard deviation of all convolution functions are presented in the inset graph. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

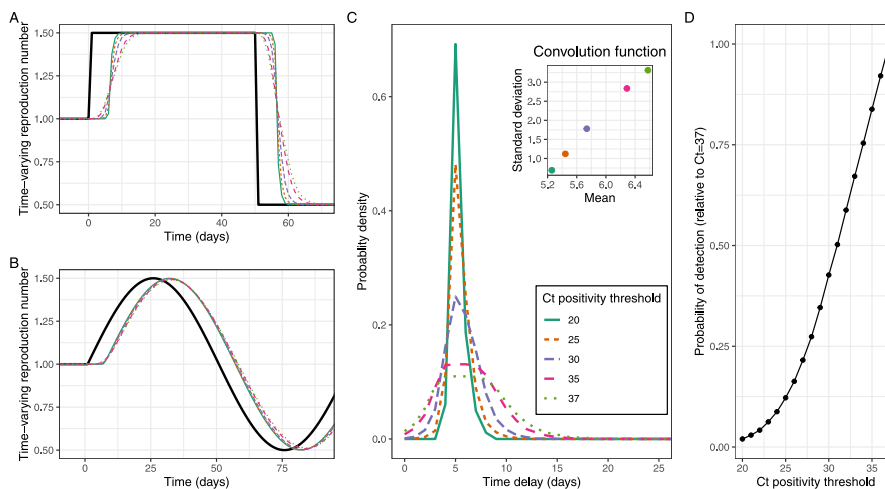


Fig. 6. The effect of imposing stricter Ct thresholds for positivity on estimates of $R_A(t)$. (A,B) The actual reproduction number, $R(t)$ (black), and the apparent reproduction number, $R_A(t)$ (coloured), for epidemic time series with different convolution functions. (C) The probability distribution of convolution functions for different Ct thresholds of rt-PCR positivity. Note that the convolution function was defined from a model of population Ct values rather than rt-PCR positivity, and so will not match the convolution function for prevalence in Fig. 4. The mean and standard deviation of all convolution functions are presented in the inset graph. (D) The proportion of individual tests classified as positive relative to the number defined as positive with a Ct threshold of 37. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The convolution function for SARS-CoV-2 symptom onset had the smallest mean and variance, of the four convolution functions considered, which would result in estimates of $R_A(t)$ that were the most similar to $R(t)$. However, there are often multiple circulating pathogens which may present similar symptom profiles. To estimate $R_A(t)$, it must be determined which individuals reporting symptoms are infected with the relevant pathogen. For this symptomatic testing the exact convolution function would rely on a distribution describing the time between symptom onset and testing (testing delay); this would likely increase the convolution function’s mean and variance. To improve estimates of $R_A(t)$ the testing delay distribution should be reduced as much as possible. Additionally, the date of symptom onset should be routinely collected; this would reduce the mean and variance of the convolution function (no testing delay component required) and would also allow $R(t)$ to be estimated (once the incubation period has been estimated) as the convolution function would be less likely to vary over time (testing delay can change over time).

Random sampling has been proposed as a method for tracking COVID-19 infections (Dean, 2022), without the biases present in standard pandemic surveillance methods such as mass testing (Ricca Peixoto et al., 2020). There has been some concern that such studies would be less well suited for estimating $R(t)$, due to the presence of long-term shedders — individuals who continue to test positive for a long duration of time, following infection. Random sampling studies measure the infection prevalence, the proportion of a population that test positive for the virus. We have demonstrated that the convolution function for prevalence of SARS-CoV-2 has a mean and variance only slightly greater than that of symptom onset. In fact, compared to mass-testing, which relies heavily on symptomatic testing, it is likely that the mean and variance of the convolution functions would be comparable, depending on the testing delay distribution.

The convolution function for symptomatic testing can at best match the incubation period of the pathogen. In contrast we have shown that there exist methods for improving the inference of $R_A(t)$ based

on prevalence studies. Imposing a stricter threshold for positivity can reduce variance of the convolution function (though larger samples may be required as more positive tests would be excluded). It is worth noting that the Ct value model we presented is not a perfect analogue to rt-PCR positivity and so results are informative, but not the exact outcome that would be obtained using a stricter threshold for positivity. Increasing the frequency of testing could also reduce both the mean and variation of the convolution function. We observed that for daily testing the convolution function's mean and variance was smaller than that of the incubation period of SARS-CoV-2. It is highly unlikely that a random sampling study, which is very expensive, could be designed to undertake such high rates of testing. However, the convolution function presented would also be valid for asymptomatic testing. Regular asymptomatic testing was performed by many individuals and businesses throughout the pandemic. Studies or surveillance systems set up around frequent asymptomatic testing could result in estimates of $R_A(t)$ that more closely resemble $R(t)$. In the latter stages of the pandemic Lateral Flow Viral Antigen detection devices (LFDs) were regularly used by some individuals. LFDs are less sensitive in general than rt-PCR testing, but they are still highly sensitive at detecting infections with high viral loads (Joint PHE Porton Down & University of Oxford SARS-CoV-2 test development and validation cell, 2020). This is at times a limitation as less infected individuals will be identified, but the resulting convolution function would likely have even less variation.

We have only considered the biases in $R(t)$ due to the convolution function. There are often many other sources of bias associated with epidemic time series data (Ricoca Peixoto et al., 2020). For example, changing testing rates can bias estimates of $R(t)$ obtained from cases identified through mass testing (Eales et al., 2022a). Additionally, estimates of $R(t)$ also rely on accurate estimates of the generation time distribution (Wallinga and Lipsitch, 2007). The generation time is not always well characterised (especially during the early period of a pandemic) and in some circumstances it can change over time (Abbott et al., 2022). When the generation time is not known (or poorly estimated), the growth rate is regularly used to quantify epidemic growth and decline as an alternative to $R(t)$ (Parag et al., 2022). However, though the growth rate is independent of generation time, estimating it from a general epidemic time series would introduce a similar set of biases as for $R(t)$ due to the convolution function.

5. Conclusion

There are clear biases when estimating $R(t)$ from an epidemic time series, under the assumption that the time series is a perfect proxy for infection incidence. By collecting data to inform on the convolution function linking the epidemic time series to incidence, estimates of $R(t)$ could be made without such biases. During periods in which the convolution function cannot be estimated, the apparent reproduction number can and should still be estimated. Despite its biases, it is highly useful for quantifying transmission dynamics and informing public health interventions. To minimise the inherent biases present, more careful consideration should be given into what epidemic data is used and how it can be better collected and manipulated. Additionally, studies in which it is the apparent reproduction number being estimated should state this explicitly, making it clear what possible biases may be present.

6. Materials and methods

6.1. Simulating epidemic time series

The time series of $R(t)$ are pre-specified. $R(t)$ is assumed to be 1 for a period of 100 days (days -99 to 0) prior to the simulation for ease of later integral calculations (integration's are all performed over the previous 100 days). The time series of $R(t)$ can then be defined in any way. We define two main time series of $R(t)$ following a square wave

and a sine wave, both with a period of 100 days, a maximum value of 1.5 and a minimum value of 0.5. We also define a time series of $R(t)$ that follows approximately the same trend as the estimated $R(t)$ in England from March 2020 to December 2020 (Knock et al., 2021). With a time series of $R(t)$, the time series for infection incidence can be calculated through Eq. (3):

$$I(t) = \int_0^\infty I(t-\tau)g(\tau)R(t)d\tau. \quad (13)$$

The generation time is assumed to be gamma distributed with a mean of 6.36 days and a standard deviation of 4.20 days (Bi et al., 2020). This was chosen to match the generation time of SARS-CoV-2, but as we assume perfect knowledge of the generation time all results are independent of the choice of the generation time distribution used. When calculating the value of $I(t)$ we perform the integral over the previous 100 days. This ensures the integral has fallen to approximately 0. Days -99 to 0 are assumed to have $I(t) = 1$.

General epidemic time series, $J(t)$ are calculated directly from $I(t)$ through Eq. (5):

$$J(t) = \int_0^\infty I(t-\tau)f(\tau)d\tau. \quad (14)$$

The integration is again performed over the previous 100 days, ensuring the integral has fallen to approximately 0. $f(\tau)$ is a convolution function which we define over a period of only 100 days.

6.2. Calculating the reproduction number

We calculate the apparent reproduction number, $R_A(t)$, directly from the general epidemic time series $J(t)$ through Eq. (8):

$$R_A(t) = \frac{J(t)}{\int_0^\infty J(t-\zeta)g(\zeta)d\zeta}. \quad (15)$$

The integration is again performed over the previous 100 days, ensuring the integral has fallen to approximately 0. As before $g(\zeta)$ is the generation time distribution which is known exactly when performing the calculation.

6.3. Convolution functions

We define many convolution functions, $f(\tau)$, for converting $I(t)$ into a general epidemic time-series. All convolution functions are defined only for values of τ between 0 and 100.

Gamma-distributed convolution functions

The gamma-distributed convolution functions are defined using the gamma distribution. A gamma distribution defined with scale parameter, k , and shape parameter, θ has mean, $k\theta$ and standard deviation, $k^{1/2}\theta$. We then define our gamma-distributed convolution functions with mean, μ and standard deviation, σ using the inverse relationships: $k = \mu^2/\sigma^2$ and $\theta = \sigma^2/\mu$.

Convolution functions for SARS-CoV-2 epidemic time series

The convolution function for symptom onset is defined by the incubation period of SARS-CoV-2, which we assume is gamma distributed with mean 6.0 days and standard deviation 3.1 days (Linton et al., 2020).

The convolution functions for deaths and hospitalisations are calculated using the equation:

$$f(\tau) = \int_\tau^0 f_1(\tau-\zeta)f_2(\zeta)d\zeta. \quad (16)$$

In this equation $f_1(\tau)$ describes the incubation period of SARS-CoV-2 (same gamma distribution as above) and $f_2(\zeta)$ describes the distribution of the time between symptom onset and death/hospitalisation. For deaths we assume $f_2(\zeta)$ is gamma distributed with mean 15.0 days and standard deviation 6.9 days (Linton et al., 2020). For hospitalisations

we assume $f_2(\zeta)$ is gamma distributed with mean 7.8 days and standard deviation 6.0 days (Hawryluk et al., 2020). The parameters used for all distributions are only meant to be informative, there are many other potential values that could have been selected from the literature.

The convolution function for prevalence is defined using the probability of testing positive as a function of time since infection. We use the median of the modelled probability of testing positive (as a function of time since infection) estimated in Hellewell et al. (2021).

Convolution function for frequent testing

The convolution function for frequent testing is calculated using the distribution describing the probability of testing positive as a function of time since infection (the convolution function for prevalence). When there is frequent testing the convolution function describes the day in which an individual first tests positive only. Writing the probability of testing positive τ days after infection as $P(\tau)$, we can calculate the probability of first testing positive τ days after infection as:

$$P(\text{first positive}, \tau) = P(\tau) \prod_{n=1}^{N-1} 1 - P(\tau - n \times \gamma), \quad (17)$$

where N is the total number of tests performed since infection and γ is the number of days between tests. The product in the above equation is calculating the probability of testing negative in all previous tests.

Convolution function for different Ct value thresholds

The convolution function for different Ct value thresholds of positivity was calculated by simulating the trajectory of Ct values over time since infection for 1000 individuals (see below). The proportion of individuals with a Ct value less than a threshold Ct value was then calculated as a function of time since infection. We had to assume the time of infection for all individuals. We assumed that all individuals reached peak viral load (minimum Ct value) exactly 5 days after infection.

Individual-level trajectories of Ct values were simulated using the results and data of Hay et al. (2022) (Ct values represent ORF1ab). In the paper an individual's Ct value trajectories can be quantified using three parameters: the minimum Ct value reached, the time taken for Ct value to decrease from the limit of detection (Ct value = 40) to its minimum value (linear decrease assumed), and the time taken for Ct value to increase from its minimum value to the limit of detection (linear increase assumed). These three parameters for an individual are assumed to be drawn from three independent population level parameter distributions (Normal distributions with mean and standard deviation).

In Hay et al. 2022 the population level parameter distributions are fitted to different groups of individual level data. However, the values of the standard deviations of these parameter distributions are not reported. We used the readily available code and data supporting Hay et al. (2022) to fit their model to a subset of their data. We included the subset of individuals with an Omicron infection, who had not received a booster vaccination, and who had a low antibody titre against the ancestral strain of SARS-CoV-2 (< 250 AU/ml). This subset was chosen as it reflected the individuals with likely the least exposure to previous infections or vaccinations. The inferred convolution function would thus be the most comparable to the convolution of prevalence defined previously; the previous convolution function relied on a model that was fit to data collected early in the SARS-CoV-2 pandemic before Omicron emerged or vaccinations began. However, as the data in Hay et al. 2022 were collected during a period in which most infections were Omicron and most individuals had received at least 1 dose of vaccine, the models are not directly comparable.

The model fit to the data included the full posteriors for the parameters describing the population level parameter distributions (means and standard deviations of Normal distributions). We extracted the mean value of these parameters. Using the population level parameter distributions, that the mean values of these parameters define, we randomly selected 1000 individual level parameters and used them to simulate 1000 Ct value trajectories. The convolution function for different Ct value thresholds of positivity was then calculated as described above.

Code availability

All code to perform the analysis in this paper is available at <https://github.com/Eales96/apparent-reproduction-number>

CRedit authorship contribution statement

Oliver Eales: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Steven Riley:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

References

- Abbott, S., Hellewell, J., Hickson, J., Munday, J., Gostic, K., Ellis, P., Sherratt, K., Gibbs, H., Thompson, R., Meakin, S., Bosse, N., Mee, P., Funk, S., 2020. EpiNow2: Estimate real-time case counts and time-varying epidemiological parameters. <http://dx.doi.org/10.5281/zenodo.3957489>.
- 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. *bioRxiv*.
- Ali, S.T., Kadi, A.S., Ferguson, N.M., 2013. Transmission dynamics of the 2009 influenza A (H1N1) pandemic in India: the impact of holiday-related school closure. *Epidemics* 5 (4), 157–163.
- Anderson, R.M., Fraser, C., Ghani, A.C., Donnelly, C.A., Riley, S., Ferguson, N.M., Leung, G.M., Lam, T.H., Hedley, A.J., 2004. Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359 (1447), 1091–1105.
- Anderson, R.M., May, R.M., 1992. *Infectious Diseases of Humans: Dynamics and Control*. OUP Oxford.
- Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S.A., Zhang, T., Gao, W., Cheng, C., Tang, X., Wu, X., Wu, Y., Sun, B., Huang, S., Sun, Y., Zhang, J., Ma, T., Lessler, J., Feng, T., 2020. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* 20 (8), 911–919.
- Chadeau-Hyam, M., Eales, O., Bodinier, B., Wang, H., Haw, D., Whitaker, M., Elliott, J., Walters, C.E., Jonnerby, J., Atchison, C., Diggle, P.J., Page, A.J., Ashby, D., Barclay, W., Taylor, G., Cooke, G., Ward, H., Darzi, A., Donnelly, C.A., Elliott, P., 2022. Breakthrough SARS-CoV-2 infections in double and triple vaccinated adults and single dose vaccine effectiveness among children in Autumn 2021 in England: REACT-1 study. *EclinicalMedicine* 48, 101419.
- Dean, N., 2022. Tracking COVID-19 infections: time for change. *Nature* 602 (7896), 185.
- Dighe, A., Cattarino, L., Cuomo-Dannenburg, G., Skarp, J., Imai, N., Bhatia, S., Gaythorpe, K.A.M., Ainslie, K.E.C., Baguelin, M., Bhatt, S., Boonyasiri, A., Brazeau, N.F., Cooper, L.V., Coupland, H., Cucunubá, Z., Dorigatti, I., Eales, O.D., van Elsland, S.L., FitzJohn, R.G., Green, W.D., Haw, D.J., Hinsley, W., Knock, E., Laydon, D.J., Mellan, T., Mishra, S., Nedjati-Gilani, G., Nouvellet, P., Pons-Salort, M., Thompson, H.A., Unwin, H.J.T., Verity, R., Vollmer, M.A.C., Walters, C.E., Watson, O.J., Whittaker, C., Whittles, L.K., Ghani, A.C., Donnelly, C.A., Ferguson, N.M., Riley, S., 2020. Response to COVID-19 in South Korea and implications for lifting stringent interventions. *BMC Med.* 18 (1), 321.
- Eales, O., Ainslie, K.E.C., Walters, C.E., Wang, H., Atchison, C., Ashby, D., Donnelly, C.A., Cooke, G., Barclay, W., Ward, H., Darzi, A., Elliott, P., Riley, S., 2022a. Appropriately smoothing prevalence data to inform estimates of growth rate and reproduction number. *Epidemics* 40, 100604.
- Eales, O., Haw, D., Wang, H., Atchison, C., Ashby, D., Cooke, G.S., Barclay, W., Ward, H., Darzi, A., Donnelly, C.A., Chadeau-Hyam, M., Elliott, P., Riley, S., 2023. Dynamics of SARS-CoV-2 infection hospitalisation and infection fatality ratios over 23 months in England. *PLoS Biol.* 21 (5), e3002118.
- Eales, O., de Oliveira Martins, L., Page, A.J., Wang, H., Bodinier, B., Tang, D., Haw, D., Jonnerby, J., Atchison, C., Ashby, D., Barclay, W., Taylor, G., Cooke, G., Ward, H., Darzi, A., Riley, S., Elliott, P., Donnelly, C.A., Chadeau-Hyam, M., 2022b. Dynamics of competing SARS-CoV-2 variants during the Omicron epidemic in England. *Nature Commun.* 13 (1), 1–11.

- Eales, O., Page, A.J., de Oliveira Martins, L., Wang, H., Bodinier, B., Haw, D., Jonnerby, J., Atchison, C., Robson, S.C., Connor, T.R., Loman, N.J., Golubchik, T., Nunez, R.T.M., Bonsall, D., Rambaut, A., Snell, L.B., Livett, R., Ludden, C., Corden, S., Nastouli, E., Nebbia, G., Johnston, I., Lythgoe, K., Torok, M.E., Goodfellow, I.G., Prieto, J.A., Saeed, K., Jackson, D.K., Houlihan, C., Frampton, D., Hamilton, W.L., Witney, A.A., Bucca, G., Pope, C.F., Moore, C., Thomson, E.C., Harrison, E.M., Smith, C.P., Rogan, F., Beckwith, S.M., Murray, A., Singleton, D., Eastick, K., Sheridan, L.A., Randall, P., Jackson, L.M., Ariani, C.V., Gonçalves, S., Fairley, D.J., Loose, M.W., Watkins, J., Moses, S., Nicholls, S., Bull, M., Amato, R., Smith, D.L., Aanensen, D.M., Barrett, J.C., Aggarwal, D., Shepherd, J.G., Curran, M.D., Parmar, S., Parker, M.D., Williams, C., Glaysher, S., Underwood, A.P., Bashton, M., Pacchiarini, N., Loveson, K.F., Byott, M., Carabelli, A.M., Templeton, K.E., de Silva, T.I., Wang, D., Langford, C.F., Sillitoe, J., Gunson, R.N., Cottrell, S., O'Grady, J., Kwiatkowski, D., Lillie, P.J., Cortes, N., Moore, N., Thomas, C., Burns, P.J., Mahungu, T.W., Liggett, S., Beckett, A.H., Holden, M.T.G., Levett, L.J., Osman, H., Hassan-Ibrahim, M.O., Simpson, D.A., Chand, M., Gupta, R.K., Darby, A.C., Paterson, S., Pybus, O.G., Volz, E.M., de Angelis, D., Robertson, D.L., Martincorena, I., Aigrain, L., Bassett, A.R., Wong, N., Taha, Y., Erkiert, M.J., Chapman, M.H.S., Dewar, R., McHugh, M.P., Mookerjee, S., Aplin, S., Harvey, M., Sass, T., Umpleby, H., Wheeler, H., McKenna, J.P., Warne, B., Taylor, J.F., Chaudhry, Y., Izuagbe, R., Jahun, A.S., Young, G.R., McMurray, C., McCann, C.M., Nelson, A., Elliott, S., Lowe, H., Price, A., Crown, M.R., Rey, S., Roy, S., Temperton, B., Shaaban, S., Heskeith, A.R., Laing, K.G., Monahan, I.M., Heaney, J., Pelosi, E., Silveira, S., Wilson-Davies, E., Fryer, H., Adams, H., du Plessis, L., Johnson, R., Harvey, W.T., Hughes, J., Orton, R.J., Spurgin, L.G., Bourgeois, Y., Ruis, C., O'Toole, Á., Gourtovaia, M., Sanderson, T., Fraser, C., Edgeworth, J., Breuer, J., Michell, S.L., Todd, J.A., John, M., Buck, D., Gajee, K., Kay, G.L., Peacock, S.J., Heyburn, D., Kitchman, K., McNally, A., Pritchard, D.T., Derivevic, S., Muir, P., Robinson, E., Vipond, B.B., Ramadan, N.A., Jeanes, C., Weldon, D., Catalan, J., Jones, N., da Silva Filipe, A., Williams, C., Fuchs, M., Miskelly, J., Jeffries, A.R., Oliver, K., Park, N.R., Ash, A., Koshy, C., Barrow, M., Buchan, S.L., Mantzouratou, A., Clark, G., Holmes, C.W., Campbell, S., Davis, T., Tan, N.K., Brown, J.R., Harris, K.A., Kidd, S.P., Grant, P.R., Xu-McCrace, L., Cox, A., Madona, P., Pond, M., Randall, P.A., Withell, K.T., Williams, C., Graham, C., Denton-Smith, R., Swindells, E., Turnbull, R., Sloan, T.J., Bosworth, A., Hutchings, S., Pymont, H.M., Casey, A., Ratcliffe, L., Jones, C.R., Knight, B.A., Haque, T., Hart, J., Irish-Tavares, D., Witele, E., Mower, C., Watson, L.K., Collins, J., Eltringham, G., Crudginton, D., Macklin, B., Iturriza-Gomara, M., Lucaci, A.O., McClure, P.C., Carlie, M., Holmes, N., Moore, C., Storey, N., Rooke, S., Yebra, G., Craine, N., Perry, M., Alikhan, N.-F., Bridgett, S., Cook, K.F., Fearn, C., Goudarzi, S., Lyons, R.A., Williams, T., Haldenby, S.T., Durham, J., Leonard, S., Davies, R.M., Batra, R., Blane, B., Spyer, M.J., Smith, P., Yavus, M., Williams, R.J., Mahanama, A.I.K., Samaraweera, B., Giris, S.T., Hansford, S.E., Green, A., Beaver, C., Bellis, K.L., Dorman, M.J., Kay, S., Prestwood, L., Rajatileka, S., Quick, J., Poplawski, R., Reynolds, N., Mack, A., Morris, A., Whalley, T., Patel, B., Georgana, I., Hosmillo, M., Pinckert, M.L., Stockton, J., Henderson, J.H., Hollis, A., Stanley, W., Yew, W.C., Myers, R., Thornton, A., Adams, A., Annett, T., Asad, H., Birchley, A., Coombes, J., Evans, J.M., Fina, L., Gatica-Wilcox, B., Gilbert, L., Graham, L., Hey, J., Hilvers, E., Jones, S., Jones, H., Kumziene-Summerhayes, S., McKerr, C., The COVID-19 Genomics UK (COG-UK) Consortium, 2022c. SARS-CoV-2 lineage dynamics in England from September to November 2021: high diversity of Delta sub-lineages and increased transmissibility of AY.4.2. *BMC Infect. Dis.* 22 (1), 647.
- Eales, O., Walters, C.E., Wang, H., Haw, D., Ainslie, K.E.C., Atchison, C.J., Page, A.J., Prosolek, S., Trotter, A.J., Le Viet, T., Alikhan, N.-F., Jackson, L.M., Ludden, C., Ashby, D., Donnelly, C.A., Cooke, G., Barclay, W., Ward, H., Darzi, A., Elliott, P., Riley, S., COVID-19 Genomics UK Consortium, 2022d. Characterising the persistence of RT-PCR positivity and incidence in a community survey of SARS-CoV-2. *Wellcome Open Res.* 7, 102.
- Eales, O., Wang, H., Haw, D., Ainslie, K.E.C., Walters, C.E., Atchison, C., Cooke, G., Barclay, W., Ward, H., Darzi, A., Ashby, D., Donnelly, C.A., Elliott, P., Riley, S., 2022e. Trends in SARS-CoV-2 infection prevalence during England's roadmap out of lockdown, January to July 2021. *PLoS Comput. Biol.* 18 (11), e1010724.
- Elliott, P., Eales, O., Bodinier, B., Tang, D., Wang, H., Jonnerby, J., Haw, D., Elliott, J., Whitaker, M., Walters, C.E., Atchison, C., Diggle, P.J., Page, A.J., Trotter, A.J., Ashby, D., Walters, W., Taylor, G., Ward, H., Darzi, A., Cooke, G.S., Chadeau-Hyam, M., Donnelly, C.A., 2022. Dynamics of a national Omicron SARS-CoV-2 epidemic during January 2022 in England. *Nature Commun.* 13 (1), 4500.
- Gire, S.K., Goba, A., Andersen, K.G., Sealfon, R.S.G., Park, D.J., Kanneh, L., Jalloh, S., Momoh, M., Fullah, M., Dudas, G., Wohl, S., Moses, L.M., Yozwiak, N.L., Winicki, S., Matranga, C.B., Malboeuf, C.M., Qu, J., Gladden, A.D., Schaffner, S.F., Yang, X., Jiang, P.-P., Nekoui, M., Colubri, A., Coomber, M.R., Fonnies, M., Moigboi, A., Gbakie, M., Kamara, F.K., Tucker, V., Konuwa, E., Saffa, S., Sellu, J., Jalloh, A.A., Kovoma, A., Koninga, J., Mustapha, I., Kargbo, K., Foday, M., Yillah, M., Kanneh, F., Robert, W., Massally, J.L.B., Chapman, S.B., Boicchio, J., Murphy, C., Nusbaum, C., Young, S., Birren, B.W., Grant, D.S., Scheffelin, J.F., Lander, E.S., Hapipi, C., Gevao, S.M., Gnirke, A., Rambaut, A., Garry, R.F., Khan, S.H., Sabeti, P.C., 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345 (6202), 1369–1372.
- Gog, J.R., Ballesteros, S., Viboud, C., Simonsen, L., Bjornstad, O.N., Shaman, J., Chao, D.L., Khan, F., Grenfell, B.T., 2014. Spatial transmission of 2009 pandemic influenza in the US. *PLoS Comput. Biol.* 10 (6), e1003635.
- Hawryluk, I., Mellan, T.A., Hoeltgebaum, H., Mishra, S., Schnekenberg, R.P., Whitaker, C., Zhu, H., Gandy, A., Donnelly, C.A., Flaxman, S., Bhatt, S., 2020. Inference of COVID-19 epidemiological distributions from Brazilian hospital data. *J. R. Soc. Interface* 17 (172), 20200596.
- Hay, J.A., Kissler, S.M., Fauver, J.R., Mack, C., Tai, C.G., Samant, R.M., Connolly, S., Anderson, D.J., Khullar, G., MacKay, M., Patel, M., Kelly, S., Manhertz, A., Eiter, I., Salgado, D., Baker, T., Howard, B., Dudley, J.T., Mason, C.E., Nair, M., Huang, Y., DiFiori, J., Ho, D.D., Grubaugh, N.D., Grad, Y.H., 2022. Quantifying the impact of immune history and variant on SARS-CoV-2 viral kinetics and infection rebound: A retrospective cohort study. *Elife* 11.
- Hellewell, J., Russell, T.W., SAFER Investigators and Field Study Team, Crick COVID-19 Consortium, CMMID COVID-19 working group, Beale, R., Kelly, G., Houlihan, C., Nastouli, E., Kucharski, A.J., 2021. Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. *BMC Med.* 19 (1), 106.
- Huisman, J.S., Scire, J., Angst, D.C., Li, J., Neher, R.A., Maathuis, M.H., Bonhoeffer, S., Stadler, T., 2022. Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2. *Elife* 11.
- International Ebola Response Team, Agua-Agum, J., Ariyaratna, A., Aylward, B., Bawo, L., Bilivogui, P., Blake, I.M., Brennan, R.J., Cawthorne, A., Cleary, E., Clement, P., Conteh, R., Cori, A., Dafee, F., Dahl, B., Dangou, J.-M., Diallo, B., Donnelly, C.A., Dorigatti, I., Dye, C., Eckmanns, T., Fallah, M., Ferguson, N.M., Fiebig, L., Fraser, C., Garske, T., Gonzalez, L., Hamblin, E., Hamid, N., Hersey, S., Hinsley, W., Jambei, A., Jombart, T., Kargbo, D., Keita, S., Kinzer, M., George, F.K., Godefrey, B., Gutierrez, G., Kannangarage, N., Mills, H.L., Moller, T., Meijers, S., Mohamed, Y., Morgan, O., Nedjati-Gilani, G., Newton, E., Nouvellet, P., Nyenswah, T., Perea, W., Perkins, D., Riley, S., Rodier, G., Rondy, M., Sagrado, M., Savulescu, C., Schafer, I.J., Schumacher, D., Seyler, T., Shah, A., Van Kerkhove, M.D., Wesseh, C.S., Yoti, Z., 2016. Exposure patterns driving ebola transmission in West Africa: A retrospective observational study. *PLoS Med.* 13 (11), e1002170.
- Joint PHE Porton Down & University of Oxford SARS-CoV-2 test development and validation cell, 2020. Preliminary Report: Rapid Evaluation of Lateral Flow Viral Antigen Detection Devices (LFDs) for Mass Community Testing.
- Kissler, S.M., Fauver, J.R., Mack, C., Tai, C.G., Breban, M.I., Watkins, A.E., Samant, R.M., Anderson, D.J., Metti, J., Khullar, G., Baits, R., MacKay, M., Salgado, D., Baker, T., Dudley, J.T., Mason, C.E., Ho, D.D., Grubaugh, N.D., Grad, Y.H., 2021. Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated persons. *N. Engl. J. Med.* 385 (26), 2489–2491.
- Knock, E.S., Whittles, L.K., Lees, J.A., Perez-Guzman, P.N., Verity, R., FitzJohn, R.G., Gaythorpe, K.A.M., Imai, N., Hinsley, W., Okell, L.C., Rosello, A., Kantas, N., Walters, C.E., Bhatia, S., Watson, O.J., Whittaker, C., Cattarino, L., Boonyasiri, A., Djaafara, B.A., Fraser, K., Fu, H., Wang, H., Xi, X., Donnelly, C.A., Jauneikaite, E., Laydon, D.J., White, P.J., Ghani, A.C., Ferguson, N.M., Cori, A., Baguelin, M., 2021. Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Sci. Transl. Med.* 13 (602).
- Kraemer, M.U.G., Pigott, D.M., Hill, S.C., Vanderslott, S., Reiner, Jr., R.C., Stasse, S., Brownstein, J.S., Gutierrez, B., Dennig, F., Hay, S.I., Wint, G.R.W., Pybus, O.G., Castro, M.C., Vinck, P., Pham, P.N., Nilles, E.J., Cauchemez, S., 2020. Dynamics of conflict during the Ebola outbreak in the Democratic Republic of the Congo 2018–2019. *BMC Med.* 18 (1), 113.
- Li, Y., Campbell, H., Kulkarni, D., Harpur, A., Nundy, M., Wang, X., Nair, H., Usher Network for COVID-19 Evidence Reviews (UNCOVER) group, 2021a. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect. Dis.* 21 (2), 193–202.
- Li, J., Lai, S., Gao, G.F., Shi, W., 2021b. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* 600 (7889), 408–418.
- Linton, N.M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A.R., Jung, S.-M., Yuan, B., Kinoshita, R., Nishiura, H., 2020. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *J. Clin. Med. Res.* 9 (2).
- Lytras, T., 2020. bayESTim: Estimate epidemic effective reproduction number in a Bayesian framework.
- MacCormack-Gelles, B., Lima Neto, A.S., Sousa, G.S., Nascimento, O.J., Machado, M.M.T., Wilson, M.E., Castro, M.C., 2018. Epidemiological characteristics and determinants of dengue transmission during epidemic and non-epidemic years in Fortaleza, Brazil: 2011–2015. *PLoS Negl. Trop. Dis.* 12 (12), e0006990.
- Nash, R.K., Nouvellet, P., Cori, A., 2022. Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges. *PLOS Digit. Health* 1 (6), e0000052.
- Parag, K.V., Thompson, R.N., Donnelly, C.A., 2022. Are epidemic growth rates more informative than reproduction numbers? *J. R. Stat. Soc. Ser. A Stat. Soc.*
- Park, S.H., Kim, W.J., Yoo, J.-H., Choi, J.-H., 2016. Epidemiologic parameters of the Middle East respiratory syndrome outbreak in Korea, 2015. *Infect. Chemother.* 48 (2), 108–117.

- Perkins, T.A., Rodriguez-Barraquer, I., Manore, C., Siraj, A.S., España, G., Barker, C.M., Johansson, M.A., Reiner, Jr., R.C., 2019. Heterogeneous local dynamics revealed by classification analysis of spatially disaggregated time series data. *Epidemics* 29, 100357.
- Ricoca Peixoto, V., Nunes, C., Abrantes, A., 2020. Epidemic surveillance of Covid-19: Considering uncertainty and under-ascertainment. *Port. J. Public Health* 38 (1), 23–29.
- Sharif-Yakan, A., Kanj, S.S., 2014. Emergence of MERS-CoV in the Middle East: origins, transmission, treatment, and perspectives. *PLoS Pathog.* 10 (12), e1004457.
- Shim, E., Tariq, A., Chowell, G., 2021. Spatial variability in reproduction number and doubling time across two waves of the COVID-19 pandemic in South Korea, February to July, 2020. *Int. J. Infect. Dis.* 102, 1–9.
- Thompson, R.N., Stockwin, J.E., van Gaalen, R.D., Polonsky, J.A., Kamvar, Z.N., Demarsh, P.A., Dahlqvist, E., Li, S., Miguel, E., Jombart, T., Lessler, J., Cauchemez, S., Cori, A., 2019. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* 29, 100356.
2022. Official UK coronavirus dashboard. Accessed: 2022-10-20. <https://coronavirus.data.gov.uk/>.
- Wallinga, J., Lipsitch, M., 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* 274 (1609), 599–604.
- Zhao, S., Musa, S.S., Qin, J., He, D., 2019. Associations between public awareness, local precipitation, and cholera in Yemen in 2017. *Am. J. Trop. Med. Hyg.* 101 (3), 521–524.